



HAL
open science

Analyser les opinions politiques sur Internet. Enjeux théoriques et défis méthodologiques.

Julien Boyadjian

► **To cite this version:**

Julien Boyadjian. Analyser les opinions politiques sur Internet. Enjeux théoriques et défis méthodologiques.. Science politique. Université de Montpellier 1, 2014. Français. NNT: . tel-01643768

HAL Id: tel-01643768

<https://hal.science/tel-01643768>

Submitted on 21 Nov 2017

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

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

THÈSE

Pour obtenir le grade de
Docteur

Délivré par l'Université Montpellier 1

**Préparée au sein de l'école doctorale
Droit et Science politique
Et du Centre d'Études Politiques de l'Europe Latine**

Spécialité : Science politique

Présentée par Julien BOYADJIAN

**Analyser les opinions politiques sur
internet**

Enjeux théoriques et défis méthodologiques

Soutenue le 23 octobre 2014 devant le jury composé de

M. Dominique CARDON, Professeur associé, Université Paris-Est

Membre invité

M. Jean-Yves DORMAGEN, Professeur, Université Montpellier 1

Directeur de thèse

M. Daniel GAXIE, Professeur, Université Paris 1

Rapporteur

Mme Fabienne GREFFET, Maître de conférences, Université de Lorraine

Membre

M. Patrick LEHINGUE, Professeur, Université de Picardie

Rapporteur

M. Éric SAVARESE, Professeur, Université Montpellier 1

Président

« La Faculté n'entend donner aucune approbation ni improbation aux opinions émises dans cette thèse ; ces opinions doivent être considérées comme propres à leur auteur »

Remerciements

Ces quatre années de thèse n'ont pas un été un exercice solitaire, bien au contraire. Je tiens ici à remercier l'ensemble des personnes qui ont contribué à la réalisation de ce travail.

Mes premiers remerciements vont à mon directeur de thèse, Jean-Yves Dormagen, pour la confiance qu'il m'a accordée, son soutien sans faille et ses conseils toujours stimulants. Je tiens également à remercier la Région Languedoc-Roussillon pour le soutien financier qu'elle m'a apporté et qui m'a permis de réaliser ce travail dans des conditions matérielles enviables. Je remercie également Malik Larabi, Florence Paulin et Xavier Marc de m'avoir fait découvrir, lors de mon stage de Master, ces nouveaux outils numériques. Les réflexions de Malik ont été particulièrement utiles à l'élaboration de mon projet de thèse.

Je tiens également à remercier Julien Velcin et l'ensemble des chercheurs, issus d'horizons disciplinaires variés, qui ont participé au projet ANR Imagiweb, dans lequel s'inscrit cette thèse. Cette interdisciplinarité a été une réelle source d'enrichissement et une vraie aide à la réalisation de ma recherche.

Mes plus vifs remerciements vont également à la société AMI Software, sans qui ma thèse n'aurait clairement pas été la même. Je tiens tout particulièrement à remercier Éric Fourboul, Leila Khouas et Aurélien Lauf pour leur aide, leur disponibilité et leur gentillesse. Merci également à Emmanuelle Rimbault, sans qui le panel « audience » n'aurait pu voir le jour.

Et bien sûr comment ne pas remercier le CEPTEL, mon laboratoire d'accueil, et l'ensemble des personnels, administratifs et enseignants, que j'ai eu le bonheur de côtoyer ces quatre dernières années. Je tiens tout particulièrement à remercier Alexandre Dézé, Sylvia Goncalves, Jacques Fontaine, Maguelone Sie, Ludovic Sposito et Aurélia Troupel. Je remercie également l'équipe des doctorants, d'aujourd'hui ou d'hier, avec qui j'ai passé des moments inoubliables : Julien Audemard, Anne-Laure Beaussier, Pauline Benjebbar, Batoul Benyamina, Sylvain Bertschy, Fanny Debil, Nicolas Ferran, Ghislain Gabalda, David Gouard, Ulrike Lepont, Sami Mahbouli, Ludovic Mariotti, Lamine Savane, Sahar Saeidnia et Alice Simon. Des remerciements tout particuliers pour finir à Laurie Patouillard qui m'a aidé dans cette dernière ligne droite et Marie Neihouser avec qui j'ai vécu cette expérience doctorale du début à la fin, et qui, en plus d'avoir été une relectrice hors-pair, a toujours été là pour me soutenir dans les moments de doute.

SOMMAIRE

Remerciements	5
Sigles et abréviations.....	8
INTRODUCTION GÉNÉRALE : LES CONDITIONS MÉTHODOLOGIQUES D'UNE ANALYSE SOCIOLOGIQUE DES OPINIONS NUMÉRIQUES.....	13
PREMIÈRE PARTIE : LES AUTOPUBLICATIONS NUMÉRIQUES COMME SOURCE DE RENOUVELLEMENT THÉORIQUE ET MÉTHODOLOGIQUE DE L'ANALYSE DES OPINIONS POLITIQUES	48
CHAPITRE 1 - LES USAGES COMMERCIAUX ET SCIENTIFIQUES DES AUTOPUBLICATIONS NUMÉRIQUES	50
CHAPITRE 2 - PANÉLISER TWITTER, UNE MÉTHODE HYBRIDE DE MESURE DES OPINIONS	130
SECONDE PARTIE : TWITTER, UN OBSERVATOIRE DU MARCHÉ DES OPINIONS POLITIQUES	207
CHAPITRE 3 – TWITTER, UN NOUVEL OBSERVATOIRE	209
DU « PAYS LÉGAL »	209
CHAPITRE 4 – LES INÉGALITÉS DE PRODUCTION DE TWEETS POLITIQUES ..	279
CHAPITRE 5 – LES LOGIQUES SOCIALES ET POLITIQUES DE CONSOMMATION DES BIENS POLITIQUES SUR TWITTER.....	331
CONCLUSION : LES APPORTS ET LIMITES DE LA MÉTHODE DE PANÉLISATION DES RÉSEAUX SOCIAUX	381
Table des matières	416
Bibliographie.....	423
Table des illustrations.....	458
Annexes	463

Sigles et abréviations

AMI : *Automatic Meaning Interpreter*

ANR : Agence Nationale de la Recherche

API : *Application Programming Interface*

ANOVA : *Analysis of variance*

BEP : Brevet d'Études Professionnelles

BTS : Brevet de Technicien Supérieur

CAP : Certificat d'Aptitude Professionnelle

CEE : Centre d'Études Européennes

CEVIPOF : Centre d'Études de la Vie Politique Française

PCS : Professions et Catégories Socioprofessionnelles

DEA : Diplôme d'Études Approfondies

DESS : Diplôme d'Études Supérieures Spécialisées

DEUG : Diplôme d'Études Universitaires Générales

DLR : Debout La République

EELV : Europe Écologie-Les Verts

FN : Front National

INSEE : Institut National de la Statistique et des Études Économiques

IUT : Institut Universitaire de Technologie

Modem : Mouvement démocrate

NPA : Nouveau Parti Anticapitaliste

PEF : Panel Électoral Français

PS : Parti Socialiste

r : coefficient de Pearson

RERS : Repères et Références Statistiques sur les enseignements, la formation et la recherche

STS : Section de Technicien Supérieur

UMP : Union pour un Mouvement Populaire

URL : *Uniform Resource Locator*

P (ρ) : Rho de Spearman (coefficient)

χ^2 : Khi-deux

V : V de Cramer

À mes parents,

INTRODUCTION GÉNÉRALE : LES CONDITIONS MÉTHODOLOGIQUES D'UNE ANALYSE SOCIOLOGIQUE DES OPINIONS NUMÉRIQUES

Le 9 février 2011 paraît sur le site internet de TF1 et dans le journal gratuit *Métro* « Le Twittoscope TF1 News : le baromètre de l'opinion publique sur Twitter »¹. Ce nouveau « baromètre » est une étude réalisée chaque mois par TNS Sofres, l'un des plus importants instituts de sondage français, en collaboration avec SemioCast, une entreprise créée en 2009, spécialisée dans l'étude des réseaux sociaux sur internet. Bien qu'il soit qualifié de « baromètre de l'opinion publique », le « Twittoscope » n'est pas un sondage par questionnaire administré à un échantillon représentatif de la population française en âge de voter, comme le sont pourtant, depuis maintenant près de quatre-vingts ans, les enquêtes ordinairement désignées sous ce terme². Ce nouveau « baromètre » ne tire pas en effet ses résultats d'un questionnaire à choix fermés mais d'une analyse informatique semi-automatisée d'un corpus exhaustif de messages (auto-)publiés par des agents sociaux sur le réseau – ou site de *micro-blogging*³ – Twitter. Ce site internet, comme de nombreux autres (tels que Facebook, Google+ ou encore les plateformes de blogs), offre la possibilité aux internautes qui y sont inscrits de publier facilement et sans intermédiaire des messages sur les sujets ou préoccupations de leur choix, dont, entre autres, la politique. Le « Twittoscope » comptabilise ainsi chaque mois le nombre de tweets⁴ citant le nom des principales personnalités politiques, ainsi que la « tonalité » de ces messages : les tweets comprenant l'occurrence « Sarkozy » sont-ils « positifs », « neutres » ou « négatifs » envers l'homme politique ?

Ces messages auto-publiés par les internautes sur les blogs et les réseaux sociaux constituent donc, à l'instar des réponses de sondage, un nouveau matériau empirique dont l'analyse peut permettre de « sonder » autrement l'« opinion publique ». Tel est en tout cas le constat que semblent désormais partager certaines sociétés d'études opérant dans le champ de production de l'opinion. Durant la campagne électorale française du printemps 2012, la

¹ Études accessibles en ligne : <http://lci.tf1.fr/politique/twittoscope/>

² Pour reprendre les mots de Loïc Blondiaux, les sondages d'opinion ont depuis leur création acquis un « quasi-monopole de l'énonciation de l'opinion publique », cf. Loïc Blondiaux, *La Fabrique de l'opinion. Une histoire sociale des sondages*, Paris, Le Seuil, 1998, p. 9.

³ Concernant les termes techniques relatifs à internet et aux nouveaux outils d'analyse des données numériques, nous invitons le lecteur à consulter notre glossaire, p. 464.

⁴ On désigne sous ce terme les messages d'une taille maximale de 140 caractères publiés sur ce réseau social (cf. glossaire).

majorité des grands sites internet d'information ont proposé à leur tour, en partenariat avec des instituts d'études et des entreprises informatiques, de nouveaux « baromètres d'opinion » sur internet⁵. Ces études se sont également diffusées dans le champ politique. Ainsi, en 2012, le Service d'Information du Gouvernement (SIG), dont l'une des principales missions consiste à « analyser l'évolution de l'opinion publique »⁶, s'est doté d'un logiciel et d'analystes dédiés à la surveillance de Twitter⁷. Internet et les réseaux sociaux seraient donc devenus pour ces agents un autre observatoire possible de l'« opinion publique ».

Une tentative d'importation d'outils de *web-tracking* en sociologie politique

Paradoxalement, la science politique et la sociologie politique n'ont accordé que très peu d'intérêt à ce nouveau matériau empirique – les autopublications numériques – et à leurs outils de collecte et de mesure – les logiciels informatiques de *web-tracking*, que nous définirons par la suite. Pour ne donner qu'un exemple, une des principales revues de la discipline, la *Revue française de science politique*, n'a publié à ce jour qu'un seul article analysant des autopublications numériques⁸. Cet état de fait est paradoxal pour au moins deux raisons : d'une part car les opinions politiques constituent précisément un objet canonique de ces disciplines et d'autre part parce que les différentes méthodes des sciences sociales, comme les sondages⁹ ou les *focus groups*¹⁰, ont toujours circulé entre le champ académique et le secteur professionnel des études de marché.

Il n'est alors pas anodin de préciser que c'est lors d'un stage réalisé au sein d'une entreprise d'études d'opinion et de marché – l'institut Somme Toute – qu'est né le questionnement qui a motivé et jalonné la réalisation de cette recherche doctorale. C'est en effet lors de ce stage que nous avons été familiarisé à ce nouveau type d'enquête, aux potentialités heuristiques de ce nouveau matériau empirique et aux logiciels de *web-tracking*.

⁵ On peut citer entre autres le « Politivox » de liberation.fr, « Le Scan » du figaro.fr, ou encore le « Twittrack2012 » d'atlantico.fr.

⁶ Verbatim recueilli sur le site de l'institution : <http://www.gouvernement.fr/gouvernement/service-d-information-du-gouvernement-sig>.

⁷ Jeanne Dussueil, « Comment le gouvernement écoute ce qui se dit sur Twitter grâce la technologie d'une start up », *Challenges*, 2012. Article en ligne : <http://www.challenges.fr/internet/20121123.CHA3487/comment-le-gouvernement-ecoute-ce-qui-se-dit-sur-twitter-grace-la-technologie-d-une-start-up.html> (accès le 12/12/2012).

⁸ Sylvain Paraisie et Jean-Philippe Cointet, « La presse en ligne au service de la démocratie locale. Une analyse morphologique de forums politiques », *Revue Française de Science Politique*, 2012/1 (Vol. 62), p. 45-70.

⁹ Loïc Blondiaux, *La Fabrique de l'opinion*, op. cit.

¹⁰ Sophie Duchesne et Florence Haegel, *L'entretien collectif. L'enquête et ses méthodes*, Paris, Armand Colin, 2008.

L'objectif de cette recherche peut donc se résumer à une tentative d'importation en science politique d'outils conçus et mobilisés par des agents issus des secteurs professionnels des études de marché, du marketing, de la communication et de la veille stratégique. Tout notre propos consistera alors à démontrer au lecteur la plus-value heuristique de ce nouveau matériau et de ses nouveaux outils de mesure et d'analyse, tout en explicitant les conditions nécessaires à la réussite de cette « importation » en sociologie politique.

Des perspectives de recherche stimulantes pour la sociologie politique

Ce nouveau matériau empirique – les autopublications numériques – et leurs outils de collecte, de mesure et d'analyse – les logiciels de *web-tracking* – offrent à la sociologie politique des perspectives méthodologiques innovantes et stimulantes.

Des données empiriques générées en dehors de toute interaction de recherche

Les productions écrites auto-publiées par les agents sociaux sur internet constituent un véritable matériau d'analyse du monde social. À l'instar des réponses de sondage, des verbatims recueillis en entretien ou des observations ethnographiques, ces messages peuvent en effet renseigner le sociologue ou le politiste sur ce que les agents sociaux disent, pensent, font ou envisagent de faire. Ainsi, chaque jour, des millions d'internautes commentent sur les blogs, les forums ou les réseaux sociaux les derniers produits qu'ils ont consommés, leurs intérêts pour le sport, le cinéma ou la politique, évoquent leurs activités quotidiennes, etc. En 2011, plus d'un tiers des Français âgés de 16 à 74 ans déclaraient ainsi publier du contenu sur les réseaux sociaux¹¹.

Cependant, à la différence des autres types de données empiriques évoqués, les autopublications numériques ne sont pas produites en situation de laboratoire. Les internautes ne publient pas des messages sur les réseaux sociaux à la demande d'un chercheur ou d'un « sondeur ». Ainsi, les individus dont les tweets ont été analysés par le « Twittoscope » ignoraient être « sondés » par un « baromètre ». La production de leurs messages n'a donc pas été altérée par le dispositif d'enquête mis en place par TNS Sofres.

¹¹ Eurostat, « Statistiques sur la société de l'information », DEPS, 2013.

Pour les sciences sociales, cette différence est tout simplement fondamentale, car tout leur « malheur »¹² est précisément d'avoir à faire à « *des êtres qui peuvent s'approprier le discours des scientifiques et modifier leurs comportements en conséquence* »¹³. N'étant pas produites en situation de laboratoire, les autopublications numériques présentent au chercheur la garantie de ne pas être des artefacts générés par la situation d'enquête. Ne se sachant pas « observés », les internautes ne peuvent alors adapter leur discours en fonction du chercheur ou du sondeur qui étudient leurs messages. Il s'agit là de l'un des principaux apports heuristiques de ce nouveau matériau d'analyse du monde social. L'analyse de ces données numériques peut en effet se prévaloir de ne pas reproduire les biais inhérents aux enquêtes déclaratives, en premier lieu desquels le risque d'imposition de problématique.

Les avantages méthodologiques des logiciels de web-tracking

De plus, parallèlement à la diffusion de cette nouvelle pratique sociale – l'autopublication numérique – se sont développés dans le secteur de la veille économique puis des études marketing des outils permettant de collecter, d'archiver et d'analyser ces données numériques : les logiciels de *web-crawling* et de *web-tracking*. Pour reprendre la définition proposée par Dominique Cardon *et alii*, un *crawler* « *est un robot d'indexation qui explore automatiquement le web pour collecter certains types de ressources* »¹⁴, dont, par exemple, les messages publiés par les internautes sur les forums, les blogs ou encore les réseaux sociaux. Un logiciel de *crawling* permet ainsi de rechercher, collecter puis archiver l'ensemble des tweets contenant une occurrence donnée (par exemple « Sarkozy »). Une fois paramétré par l'utilisateur, le logiciel collectera automatiquement, c'est-à-dire sans intervention humaine, l'ensemble des tweets correspondant aux paramètres de la recherche. Les logiciels de *web-tracking* permettent quant à eux de suivre sur la longue durée l'activité de publication d'un site internet ou d'un compte d'utilisateur sur un réseau social. Une fois l'adresse internet du compte saisie dans le logiciel, le *tracker* collectera et archivera l'exhaustivité des messages publiés par ce compte. L'enquêteur – ou le chercheur – pourra

¹² Pierre Bourdieu, Jean-Claude Chamboredon et Jean-Claude Passeron, *Le métier de sociologue. Préalables épistémologiques*, Paris, Éditions de l'EHESS, 1968.

¹³ Michel Grossetti, « Plus largement, les sciences sociales... », *SociologieS* [En ligne], mis en ligne le 27 janvier 2012, consulté le 14 octobre 2013. URL : <http://sociologies.revues.org/3788>

¹⁴ Dominique Cardon, Guilhem Fouetillou, Clémence Lerondeau et Christophe Prieur, « Esquisse de géographie de la blogosphère politique (2007-2009) », in Fabienne Greffet (dir.), *Continuerlalutte.com*, Paris, Presses de la Fondation Nationale des Sciences politiques, 2011, p. 75.

ainsi précisément connaître la fréquence de publication de ce compte, ainsi que le contenu et l'heure de publication de chacun de ses messages.

Dans le cadre de cette recherche doctorale, nous avons mobilisé le logiciel *Ami Opinion Tracker*¹⁵, développé par la société *Ami Software*. Initialement destiné au secteur de la veille et de l'intelligence économique, le logiciel s'est par la suite repositionné à destination des secteurs des études et de la communication. Notre collaboration avec *Ami Software* s'est déroulée durant cette phase de repositionnement. Le logiciel a alors subi de nombreuses transformations, du fait notamment de nos recommandations.

Notre partenariat avec la société Ami Software est né d'un projet commun, Netscope. À l'instar du Bureau of Applied Social Research (B.A.S.R.), il s'agissait de produire des recherches de sciences sociales appliquées. Le projet Netscope réunissait des sociologues du politique (dont principalement Jean-Yves Dormagen et Alexandre Dézé, du département de science politique de l'Université Montpellier 1), des « sondeurs » professionnels (Malik Larabi et Florence Paulin, de l'institut Somme Toute, où nous avons réalisé notre stage de Master) et des informaticiens de la société Ami Software. Dans le cadre de ce projet, les responsables d'Ami Software nous avaient alors donné un accès gratuit à leur logiciel (l'acquisition de la licence est de l'ordre de 26 000 euros) afin que nous lui apportions une « expertise » de sciences sociales, qu'ils estimaient susceptible d'intéresser les clients « traditionnels » des études d'opinion (dont notamment les professionnels de la politique) qu'ils cherchaient alors à « conquérir ». Dans le cadre du projet Netscope ont notamment été réalisées durant 22 semaines, du 5 décembre 2011 au 4 mai 2012, des études d'opinion sur internet pour le pôle numérique de campagne de François Hollande, alors candidat à la Présidence de la République. Cette collaboration avec la société Ami Software s'est par la suite prolongée dans le cadre d'un projet cette fois purement académique, le projet ANR Imagiweb, dans lequel s'inscrit cette recherche doctorale.

Comme la majorité des logiciels de *web-tracking*, *AMI Opinion Tracker* répond à trois principales fonctions : collecter de façon automatisée les messages publiés sur internet, archiver et classifier ces messages, les analyser à l'aide d'outils lexicométriques et

¹⁵ Du nom de la technologie brevetée *Automatic Meaning Interpreter* (AMI).

cartographiques. Dans le cadre du projet Netscope, nous avons ainsi « mis sous surveillance » 146 personnalités et partis politiques sur Twitter. Le logiciel nous a permis de collecter en temps réel l'ensemble des tweets contenant l'une des 146 occurrences définies (« Hollande », « Sarkozy », « TVA », « UMP », « République », etc.) et de les archiver durant plusieurs années. Ainsi, depuis 2010, sous réserve que le logiciel ait fonctionné de manière optimale, tous les tweets publiés sur le réseau social contenant l'une de ces occurrences doivent avoir été collectés et archivés par notre logiciel. En mai 2014, *Ami Opinion Tracker* avait ainsi collecté environ cent millions de messages (101 917 707 pour être précis).

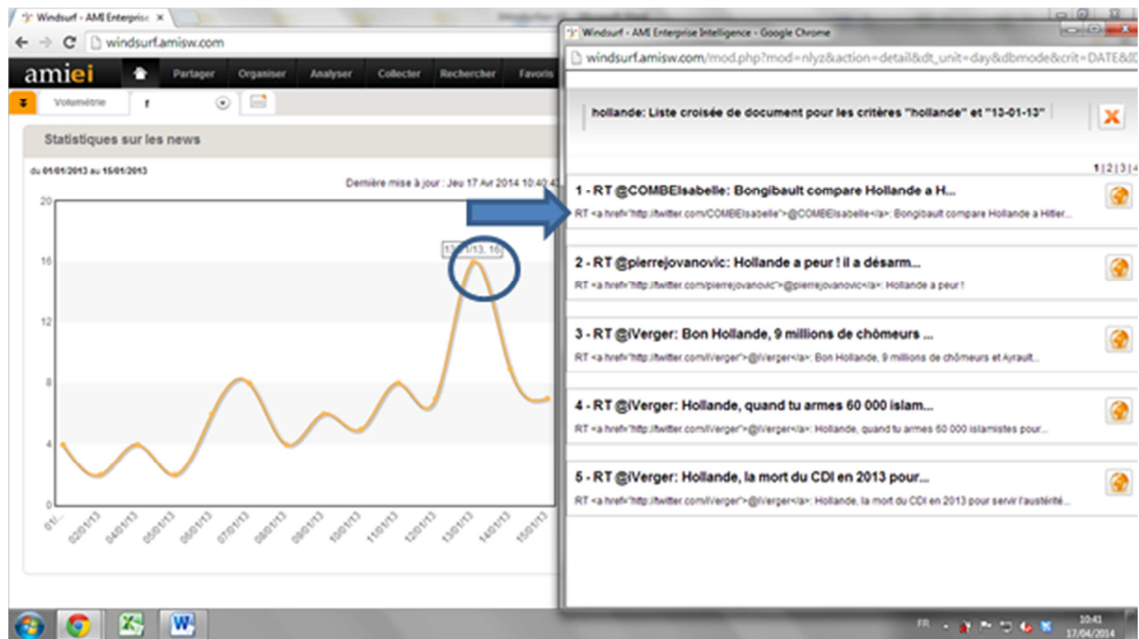
Ces logiciels de *web-tracking* offrent selon nous trois principaux avantages méthodologiques. Le premier est d'autoriser un traitement quantitatif de données qualitatives et, inversement, un traitement qualitatif de données quantitatives. Les logiciels de *web-tracking* collectent chaque jour des millions de messages publiés sur internet, c'est-à-dire des millions de verbatims. En automatisant la collecte et l'archivage de ces verbatims et en proposant des outils lexicométriques pour les analyser, ces logiciels autorisent un traitement quantitatif de ces données textuelles, qualitatives. Ainsi, à partir d'un corpus de messages collectés par le logiciel, un politiste ou un sociologue pourra choisir par exemple d'étudier les modes de production des opinions politiques exprimées dans ces messages, en analysant le degré de technicité des termes politiques mobilisés. Le chercheur prendra connaissance des différents modes de production des opinions publiées sur internet, ainsi que les termes les plus couramment associés à chacun d'eux. Mais ce n'est pas tout. En ayant recours à des outils de lexicométrie, il sera alors possible dans un second temps de mesurer le poids relatif de chaque mode de production. Après avoir analysé et isolé différents cas d'étude (logique de saturation), le chercheur pourra ensuite mesurer leurs poids respectifs (logique de représentativité).

De plus, les logiciels de *web-tracking* autorisent un traitement qualitatif de données quantitatives. En paramétrant le logiciel, le politiste pourra par exemple connaître le nombre de messages politiques publiés par une population d'internautes sur une séquence de temps donnée. Il constatera alors sans doute une distribution du nombre de messages publiés très inégale dans le temps. Cette information, de nature quantitative, est intéressante en soi : elle permet de démontrer l'existence d'effets conjoncturels sur la prise de parole politique¹⁶. Mais ces données chiffrées, quantitatives, peuvent également être représentées sous forme de données textuelles, qualitatives. À chaque donnée chiffrée, à chaque pourcentage, correspond

¹⁶ Cet exemple illustratif est tiré de l'un de nos résultats de recherche. Cf. chapitre 4.

en effet une donnée textuelle, un verbatim. En un simple clic, le chercheur peut alors connaître le contenu des messages correspondant à un « pic » de production constaté sur le graphique délivré par le logiciel (cf. capture d'écran 0.1). Les logiciels de *web-tracking* permettent donc de contextualiser et de qualifier les données quantitatives recueillies.

Illustration 0.1. Représentation sous forme textuelle de données quantitatives



Source : *Ami Opinion Tracker*, 2014

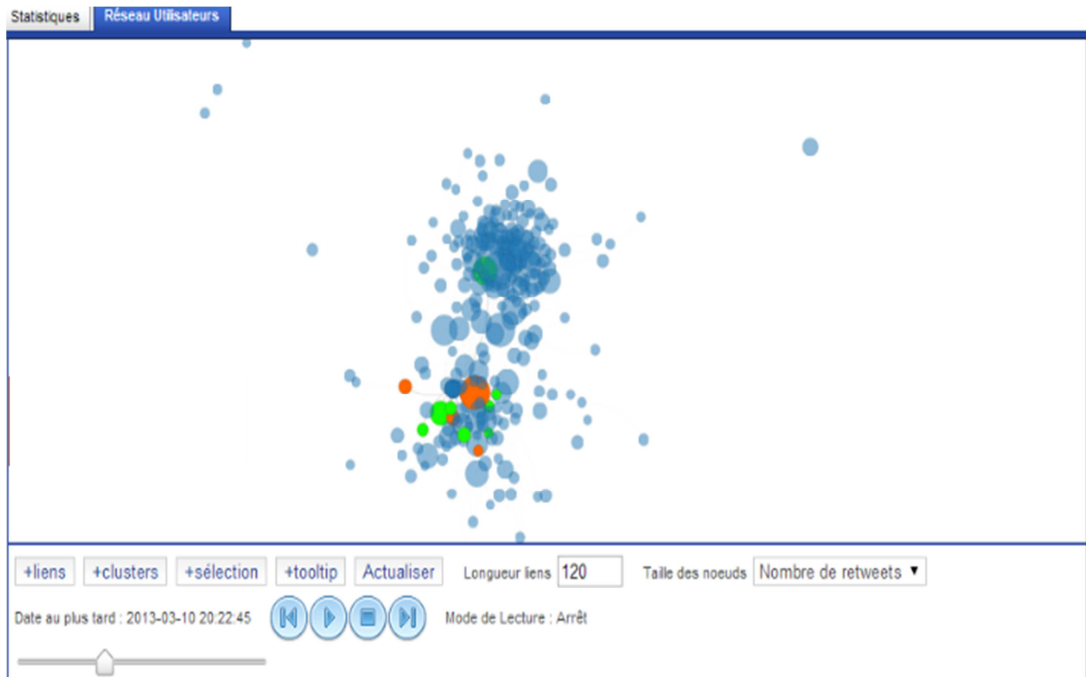
Un deuxième avantage méthodologique qu'un sociologue ou un politiste peut tirer de ces logiciels de *web-tracking* est de pouvoir analyser *a posteriori* des réactions « à chaud » à des « événements » ou des faits passés. Lorsqu'un politiste souhaite par exemple reconstituer la trajectoire de vote de certains enquêtés, il n'a souvent d'autre choix que de faire appel à leur mémoire, à leurs souvenirs, plus ou moins lointains. Or l'usage des propos rétrospectifs en entretien pose de nombreux problèmes épistémologiques : mensonges, dissimulations, défaillances de la mémoire, phénomènes de reconstitution *a posteriori*, etc.¹⁷. Les logiciels de *web-tracking* permettent quant à eux, nous l'avons dit, de collecter l'exhaustivité des messages publiés par un internaute durant plusieurs années. Ces messages sont le plus souvent

¹⁷ Gilles Pinson, « Peut-on vraiment se passer de l'entretien en sociologie de l'action publique ? », *Revue Française de Science politique*, 2007/5 (Vol. 57), p. 555-597.

des réactions « à chaud » à des *stimuli* ou à des « événements » : débat télévisé d'entre-deux-tours, victoire d'un candidat, adoption d'un projet de loi, etc. Toutes ces réactions, exprimées sous forme de messages écrits, sont donc archivées par le logiciel. Ces verbatims peuvent être alors analysés plusieurs mois voire plusieurs années après qu'ils aient été publiés et collectés. Le politiste peut donc analyser les réactions, les discours et les représentations des internautes sur des faits ou des événements passés au moment-même où se sont déroulés ces faits, ou du moins au moment où les internautes ont produit un discours sur ces faits ; et non, comme c'est le cas avec les enquêtes par questionnaire, en faisant appel à leur mémoire *a posteriori*, plusieurs jours, plusieurs mois voire plusieurs années après.

Enfin, un troisième avantage méthodologique qu'un politiste ou un sociologue peut tirer de ces logiciels de *web-tracking* est de pouvoir représenter graphiquement les dynamiques de diffusion et de circulation des messages publiés sur internet. Sur les réseaux sociaux numériques, chaque utilisateur est « lié » à d'autres utilisateurs, le plus souvent en fonction de liens affinitaires électifs. Ainsi, sur Twitter, chaque nouvel inscrit est invité à s'« abonner » à des comptes d'autres utilisateurs. En s'abonnant à un compte *x*, l'utilisateur *y* recevra l'ensemble des tweets publiés par ce compte *x* dans sa *time-line*, un espace d'édition où défilent en temps réel les messages publiés par l'ensemble des comptes auxquels il est abonné. Mais les titulaires de ces comptes peuvent également s'abonner en retour au compte de l'utilisateur *y*. Chaque utilisateur est donc défini par un nombre délimité d'*abonnés* et d'*abonnements*. De plus, chaque utilisateur a la possibilité de *retweeter* – c'est-à-dire de reproduire à l'identique sur son propre compte – un message publié par un autre utilisateur. Le message publié originellement par l'utilisateur *z* sera alors reproduit à l'identique sur le compte de l'utilisateur *w*. Ces trois types de données – abonnements, abonnés, et retweets – peuvent être représentés graphiquement à l'aide d'outils cartographiques (*cf.* capture d'écran 0.2). La cartographie de ces réseaux d'utilisateurs peut ainsi renseigner le politiste sur les logiques sociales et politiques de consommation et de circulation des messages politiques. Les individus s'abonnent-ils à des individus qui leur ressemblent politiquement et socialement ? Un message politique publié par un agent positionné à droite circule-t-il dans des groupes politiquement ancrés à gauche ?

Illustration 0.2. Cartographie d'un ensemble d'inscrits au réseau social Twitter



Source : *Ami Opinion Tracker*, 2014

Les autopublications numériques et les logiciels de *web-tracking* offrent donc à la sociologie politique des pistes de recherche stimulantes. Cependant, s'il ne s'arme pas d'un cadre théorique et d'un dispositif méthodologique adaptés, le politiste peut se trouver rapidement démuné face à cette masse de données numériques.

Les limites des logiciels de web-tracking

Les logiciels de *web-tracking* permettent de collecter et d'archiver quotidiennement des centaines de milliers voire des millions de données numériques. Le chercheur peut alors rapidement se trouver démuné face à cette masse d'informations. Trois principales limites peuvent être associées aux analyses de *web-tracking* : agréger de manière indifférenciée des informations de nature très différente, privilégier une analyse centrée sur les flux de messages (les autopublications numériques) plutôt que sur les individus (les producteurs des autopublications) et surtout ne pas parvenir à situer ces internautes dans l'espace social « réel ».

Une agrégation de données hétéroclites

Il existe une très grande variété d'espaces d'autopublication numérique : blogs, forums électroniques, sites d'information ou sites de commerce en ligne, plateformes de partage de contenu (Youtube, Dailymotion, Instagram, Deezer, etc.), réseaux sociaux (Facebook, Twitter, Google+, etc.), etc.

Le logiciel *Ami Opinion Tracker* permet d'indexer chacun de ces espaces. En quelques secondes, le chercheur peut ainsi connaître le nombre de messages politiques publiés durant une journée sur l'ensemble du « Web 2.0 » (on désigne d'ordinaire sous ce terme les sites internet permettant aux individus de publier par eux-mêmes des contenus numériques).

Séduit par le vieux rêve positiviste de pouvoir étudier, depuis sa tour d'ivoire – ou sa tour d'ordinateur en l'occurrence – la totalité du monde social – du moins tel qu'il se donne à voir sur internet – le chercheur peut être alors tenté d'indexer le plus grand nombre de sites internet possible. Cependant, cette quête d'exhaustivité nous paraît vaine. Chaque espace d'autopublication numérique est défini par ses propres règles d'utilisation, ses propres contraintes d'énonciation et ses propres formats d'expression et de publication. Le profil sociologique des utilisateurs de ces espaces numériques et leurs pratiques sociales diffèrent également d'un espace à l'autre. En réalité, chaque espace d'autopublication peut être appréhendé comme un terrain d'enquête en soi. Comprendre les logiques qui structurent les pratiques sociales sur un espace d'autopublication donné est un projet de recherche à part entière et à temps plein, serait-on tenté de dire.

Choisir d'étudier *scientifiquement* plusieurs espaces d'autopublication numérique – c'est-à-dire plusieurs terrains d'étude – nécessite alors d'adopter une méthode comparatiste. Or, les logiciels de *web-tracking*, en agrégeant de manière relativement indifférenciée des données numériques aussi différentes que des billets de blog de plusieurs pages (dont certains peuvent s'apparenter à des articles de journaux dits de « réflexion ») ou de très brefs messages publiés sur Facebook (dont beaucoup prennent la forme d'échanges de pair à pair) invitent le chercheur à comparer ce qui n'est pas comparable, ou plutôt, à ne pas comparer (au sens scientifique du terme) des terrains qui devraient l'être.

Une analyse centrée sur les flux de messages plutôt que sur les individus

Une autre limite des analyses de *web-tracking* est de raisonner en termes de flux de messages et non en termes d'individus¹⁸. L'unité de mesure du logiciel *Ami Opinion Tracker* est ainsi le message (un tweet, un billet de blog, un commentaire sur un site d'information, etc.) et non le producteur du message, l'individu. Si l'utilisateur du logiciel peut très rapidement connaître le nombre de messages politiques publiés sur une période de temps donnée, il ne peut en revanche connaître le nombre d'utilisateurs uniques, le nombre d'individus qui ont publié ces messages. Ainsi, le « Twittoscope » évoqué précédemment mesure le nombre de tweets mensuel consacrés à chaque personnalité politique étudiée par le « baromètre », mais ne dévoile pas le nombre de comptes uniques qui se dissimule derrière ces messages.

De plus, nous le verrons, le niveau de production de messages politiques est très inégal d'un individu à l'autre. Le nombre moyen *individuel* de messages est donc une variable tout à fait centrale à prendre en compte dans l'analyse des logiques sociales de production des opinions politiques sur internet. Cette variable pourrait permettre entre autre d'identifier et d'isoler certains internautes militants, pour qui l'activité de publication de messages politiques s'apparente à une véritable pratique militante. Cette variable du nombre de messages publiés par individu est pourtant absente des analyses de *web-tracking*.

Une impossible caractérisation socio-démographique des producteurs d'autopublications numériques

Mais la principale limite des analyses de *web-tracking*, qui constitue un véritable obstacle à leur importation dans le champ de la sociologie politique, réside dans leur incapacité à qualifier socialement et politiquement les producteurs des messages publiés sur internet.

Comme le notent Dominique Boullier et Audrey Lohard, « *ce qui est alors étudié* [par les analyses de *web-tracking*] *ne fait plus référence à des émetteurs bien identifiés dont les*

¹⁸ Précisons que si la majorité des comptes d'utilisateurs sur Twitter sont des comptes personnels, gérés par des agents sociaux « en leur nom propre », d'autres comptes, notamment d'entreprise, de média ou d'organisation politique sont administrés par des agents « professionnels ». Certains comptes sont également alimentés artificiellement par des robots informatiques.

chercheurs posséderaient les caractéristiques sociodémographiques classiques et considérées comme définitives de la personne »¹⁹. En effet, les messages publiés par les individus sur internet ne sont pas systématiquement rattachés à des auteurs dont on connaît les caractéristiques socio-démographiques. Et si sur les réseaux sociaux, tels que Facebook ou Twitter, les individus indiquent parfois certaines informations concernant leur sexe, leur âge, leur profession ou leur lieu d'habitation, ces données sont parcellaires, fragmentées, et ne permettent pas au sociologue d'opérer une qualification sociologique *systématique* des auteurs des messages qu'il étudie.

Les autopublications numériques présentent l'intérêt, nous l'avons dit, de ne pas être des productions générées par une interaction avec un chercheur. Pourtant, ce sont bien ces interactions qui permettent au chercheur – mais aussi au sondeur – de pouvoir situer socialement et politiquement les enquêtés, en les *interrogeant* directement sur leurs caractéristiques sociodémographiques et politiques. Cette logique de questionnement étant absente des analyses de *web-tracking*, le sociologue ne peut alors être en mesure de situer socialement et politiquement les producteurs des messages qu'il étudie.

Pour certains sociologues et politistes, cette situation ne serait pas problématique. Ainsi, pour Dominique Boullier et Audrey Lohard, ces informations socio-démographiques n'étant plus accessibles sur internet, les chercheurs devraient en faire alors leur « *deuil* » et « *changer de paradigme* » :

« Les contraintes et les limites de l'informatique sont déjà difficiles à accepter sans aucun doute pour les SHS mais celles du web sont différentes et tout aussi difficiles à accepter. [...] Ce qui est alors étudié ne fait plus référence à des émetteurs bien identifiés dont les chercheurs posséderaient les caractéristiques sociodémographiques classiques et considérées comme définitives de la personne (en tant que sujet statistique tout au moins). Ici les énoncés sont nivelés et c'est leur agrégation qui fait sens. Ce qui ne devrait pas effrayer toute une partie des sciences sociales qui n'a pas de culte particulier des individus et qui est prête à considérer que ces flux d'opinions sont des objets sociologiques tout à fait légitimes. Mais il faut alors changer de paradigme et ne plus chercher des explications par des causes qui seraient logées dans des propriétés socio-démographiques considérées comme des attributs d'un autre statut que toutes celles auxquelles on peut accéder sur le web [...] Il

¹⁹ Dominique Boullier et Audrey Lohard, *Opinion mining et Sentiment analysis : Méthodes et outils*, Marseille, OpenEdition Press, 2012, p. 189.

faut donc faire son deuil et pour longtemps des repères socio-démographiques sur le web actuel et pour cela changer de paradigme d'« explication » dans les sciences humaines et sociales, tellement fondées sur une théorie trop implicite de l'individu comme individu et non comme sociation »²⁰.

À l'inverse de ces deux auteurs, nous proposons de ne pas faire « notre deuil » des références socio-démographiques. Pour autant, nous pensons que ces données numériques et leurs outils de mesure et d'analyse – les logiciels de *web-tracking* – peuvent apporter, nous l'avons dit, une réelle plus-value heuristique à l'analyse sociologique des opinions politiques. Comment alors tirer profit de ces données sans être contraints d'abandonner la référence socio-démographique ? Cette question en appelle une autre : à partir de quels cadres théoriques et de quels dispositifs méthodologiques le sociologue ou le politiste peut-il analyser ces données numériques ? Il s'agit là de l'interrogation fondatrice de ce travail.

À partir de quel cadre théorique et selon quelle méthodologie analyser les données collectées par les logiciels de *web-tracking* ?

Afin d'être en mesure de proposer une analyse sociologique des autopublications numériques – c'est-à-dire une analyse centrée sur les propriétés sociales et politiques des auteurs de ces messages – il convient de mobiliser un cadre théorique et une méthodologie ajustés aux propriétés spécifiques de ce nouveau matériau empirique. Les logiciels de *web-tracking* ne constituent pas en soi une méthode, mais simplement un outil, à mettre au service d'un dispositif méthodologique. Un dispositif qui doit être lui-même ajusté à des questions de recherche traitées à partir d'un cadre théorique opérant.

Si ce nouveau type de données empiriques a très peu été analysé par la sociologie politique, d'autres disciplines universitaires ont fait au contraire de l'analyse des autopublications numériques un axe de recherche majeur, comme notamment les sciences de l'information et de la communication ou encore les sciences du marketing. La première question que nous nous sommes posée est alors de savoir si l'on pouvait trouver dans cette littérature « étrangère » à la sociologie politique une méthodologie et un cadre théorique ajustés à notre projet.

²⁰ *Ibidem.*

Le cadre délibératif habermassien et les logiques d'influence interpersonnelle de Lazarsfeld comme principaux horizons théoriques des recherches sur les autopublications numériques

Si elles ont été peu analysées par la sociologie politique, les autopublications numériques ont été en revanche au centre d'une abondante littérature dans d'autres disciplines universitaires. Nous proposerons une synthèse plus complète de ces travaux dans le premier chapitre de cette thèse, consacré notamment aux usages scientifiques des autopublications numériques. Nous évoquerons alors simplement ici deux des principaux cadres théoriques mobilisés dans certains pans de cette littérature.

Parmi ces nombreux travaux, on peut en effet distinguer plus particulièrement deux approches. Une première, principalement issue des recherches en sciences de l'information et de la communication, s'intéresse aux discussions politiques en ligne et questionne la « qualité de leur délibération » à partir d'une grille de lecture inspirée des travaux de Jürgen Habermas. Une seconde s'intéresse aux phénomènes de « viralité » des données numériques et mobilise les théories de Paul Lazarsfeld sur l'influence interpersonnelle et le leadership d'opinion.

Le cadre théorique habermassien de la délibération

En sciences de l'information et de la communication notamment, de nombreuses recherches ont entrepris d'analyser les discussions politiques en ligne en questionnant la qualité de « leur délibération », à partir d'une grille théorique inspirée des travaux de Jürgen Habermas sur la situation idéale de prise de parole²¹.

Comme l'a résumé Julien Talpin, « *la théorie délibérative défend qu'il est souhaitable que les décisions collectives soient arrêtées après une évaluation des avantages et des inconvénients de telle ou telle proposition, plutôt que suite à une simple agrégation de votes ou par le jeu du rapport de force entre intérêts contradictoires. Afin de parvenir à la meilleure décision collective cette évaluation doit se faire de façon argumentée, guidée par la seule "force du meilleur argument". Les propositions avancées doivent donc viser le bien commun et non l'intérêt privé des locuteurs. En effet, si la discussion dans l'espace public est valorisée, n'importe quel type de discours n'est pas en mesure de faire émerger les vertus*

²¹ Jürgen Habermas, *Droit et démocratie. Entre faits et normes*, Paris, Gallimard, 1997.

escomptées d'une "bonne délibération" »²². Afin d'évaluer la qualité d'une « bonne délibération », Jürgen Habermas a établi « un certain nombre de critères : (1) l'inclusion dans la délibération ; (2) l'égalité des participants à la discussion ; (3) le degré de justification des arguments ; (4) le respect et l'écoute des autres points de vue ; (5) la capacité à prendre la position de ses interlocuteurs ; (6) la sincérité des arguments énoncés. »²³.

Cette grille de lecture *théorique*, issue d'une réflexion de philosophie politique, a alors donné lieu à des évaluations *empiriques* en sciences sociales. De nombreux travaux ont en effet entrepris d'évaluer empiriquement des dispositifs participatifs « hors-ligne »²⁴ et « en ligne »²⁵. C'est à cette seconde catégorie de dispositifs – les discussions politiques sur internet – que nous allons ici nous intéresser.

Ces recherches consistent pour la plupart à coder « manuellement » – c'est-à-dire de façon non-automatisée – des discussions politiques sur internet à partir de grilles d'évaluation multi-critérielles, inspirées pour la plupart de la typologie proposée par Habermas. Il s'agit alors de savoir si la discussion analysée répond aux principaux critères d'une « bonne délibération ». Comme le remarque Julien Talpin, « d'un point de vue méthodologique, [ces recherches] requièrent de construire des catégories, puis de coder des séquences de discussion, en leur attribuant des points selon qu'elles atteignent plus ou moins les standards attendus »²⁶. En synthèse de ces travaux, Fabienne Greffet et Stéphanie Wojcik notent que « confronté au déroulement concret des discussions en ligne, le cadre d'analyse habermassien se heurte à des difficultés théoriques et méthodologiques. Ainsi, les critères d'évaluation de la qualité de la délibération connaissent, selon les auteurs, divers degrés de raffinement »²⁷, les deux auteures citant différents exemples de typologie, comme celle mobilisée par Lincoln Dahlberg²⁸, qui « identifie six indices de délibération – la rationalité, la

²² Julien Talpin, « Qualité de la délibération », in Ilaria Casillo avec Rémi Barbier, Loïc Blondiaux, Francis Chateauraynaud, Jean-Michel Fourniau, Rémi Lefebvre, Catherine Neveu et Denis Salles (dir.), *Dictionnaire critique et interdisciplinaire de la participation*, Paris, GIS Démocratie et Participation, 2013, ISSN : 2268-5863. URL : <http://www.participation-et-democratie.fr/node/1346>.

²³ *Ibidem*.

²⁴ Voir par exemple Loïc Blondiaux, « La démocratie par le bas. Prise de parole et délibération dans les conseils de quartier du vingtième arrondissement de Paris », *Hermès*, n°26-27, 2010, p. 323-338 ; ou encore Jon Elster, « Argumenter et négocier dans deux assemblées constituantes », *Revue française de science politique*, vol. 44, n°2, 1994, p. 187-256.

²⁵ Fabienne Greffet et Stéphanie Wojcik, « Parler politique en ligne. Une revue des travaux français et anglo-saxons », *Réseaux*, 2008/4 (n° 150), p. 19-50.

²⁶ *Ibidem*.

²⁷ *Ibidem*.

²⁸ Lincoln Dahlberg, « Computer-Mediated Communication and the Public Sphere: a critical analysis », *Journal of Computer-Mediated Communication*, vol. 7, n°1, 2001 cité dans Fabienne Greffet et Stéphanie Wojcik, « Parler politique en ligne. Une revue des travaux français et anglo-saxons », art. cit.

réflexivité, l'inversion des rôles (capacité à se mettre à la place de l'autre pour comprendre son point de vue), la sincérité, l'inclusion et l'équité discursive, l'autonomie par rapport au pouvoir de l'État et au pouvoir économique » ou encore celle de Jennifer Stromer-Galley²⁹ qui « synthétise, quant à elle, huit éléments repérables empiriquement qui contribueraient à une délibération en ligne de qualité : l'expression d'opinions raisonnées ; la référence à des sources extérieures ; l'absence d'attaques ad hominem ; la possibilité d'avoir des opinions ou des perspectives opposées ; une égalité de participation ; la cohérence ; la réflexivité ; la narrativité »³⁰.

Malgré ses ajustements, ce cadre analytique ne nous paraît pas répondre aux exigences d'une analyse sociologique des opinions numériques qui serait centrée sur leurs producteurs, et ce pour plusieurs raisons. La première est liée au caractère clairement normatif de ces approches. Si dans une approche de philosophie politique, ce caractère *normatif* n'est par définition pas problématique, il est en revanche plus discutable, dans une perspective scientifique, de chercher à confronter systématiquement la réalité empirique observée à un modèle de délibération idéal qui n'a aucune existence réelle. À ce titre, de nombreuses critiques ont été adressées aux travaux mesurant ainsi la « qualité » des délibérations en ligne. Julien Talpin nous en donne un résumé :

« Non seulement [les critères d'évaluation de la qualité des délibérations observées] s'avèrent très largement procéduraux, mais ils apparaissent également excluants. En dépit du principe d'inclusion, purement formel, la centralité offerte à l'argumentation laisse de côté des formes moins structurées de prise de parole et marginalise l'expression des émotions dans l'espace public (Young, 1996³¹). Sachant que ceux qui maîtrisent le mieux l'argumentation sont également les individus les mieux dotés en ressources culturelles, la qualité de la délibération se traduirait par l'exclusion des classes populaires et des minorités. Ces critiques ont conduit à défendre une conception élargie et post-habermassienne de la délibération, s'intéressant à toutes les formes d'expression des citoyens ordinaires dans

²⁹ Jenifer Stromer-Galley, « Decoding deliberation online », *Paper presented at the Second Conference on Online Deliberation*, Palo Alto, California, Mai 2005 cité dans Fabienne Greffet et Stéphanie Wojcik, « Parler politique en ligne. Une revue des travaux français et anglo-saxons », art. cit.

³⁰ Fabienne Greffet et Stéphanie Wojcik, « Parler politique en ligne. Une revue des travaux français et anglo-saxons », art. cit.

³¹ Marion Young, « Communication and the Other: Beyond Deliberative Democracy », in Seyla Benhabib (dir.), *Democracy and Difference. Contesting the Boundaries of the Political*, Princeton, Princeton University Press, 1996, p. 120-135.

l'espace public, du témoignage à l'expression de la colère, des discussions ordinaires à la rhétorique »³².

Cependant, même élargie à des formes d'expression moins « légitimes », la notion de délibération, prise dans son acception habermassienne ou « *post-habermassienne* », nous semble trop réductrice. Les messages publiés sur les réseaux sociaux, les blogs ou les forums ne s'insèrent pas nécessairement dans des « discussions ». Nous le verrons, en particulier sur Twitter, beaucoup de messages publiés ne donnent lieu à aucun échange collectif. Ces productions ne peuvent donc pas être analysées au prisme d'une supposée « intention délibérative » de leur auteur. Surtout, ces approches tendent selon nous à accorder une attention trop grande à la valeur intrinsèque des messages (« que disent les internautes ? » et « que veulent dire les internautes ? »), sans tenir compte des logiques sociales qui président à leur production (« pourquoi les internautes publient ces messages ? »), logiques dont n'ont pas nécessairement conscience les agents sociaux eux-mêmes. Pour ces raisons, le cadre théorique délibératif ne nous semble donc pas adapté à notre projet d'une analyse sociologique des opinions numériques.

Un deuxième pan de la littérature académique, très différent des travaux que nous venons d'évoquer, appréhende les autopublications numériques non pas comme des contributions à une « délibération » mais comme des productions dont on questionne les logiques de « viralité ». Ces travaux mobilisent alors un cadre théorique plus proche de la sociologie politique, celui de l'influence interpersonnelle de Paul Lazarsfeld.

Les phénomènes de « viralité » et les théories de l'influence interpersonnelle

Les théories sur l'influence interpersonnelle et les leaders d'opinion, développées dans les années 1940 par Paul Lazarsfeld et les chercheurs du *B.A.S.R.*³³ sont en effet très régulièrement mobilisées par toute une partie de la littérature sur les données numériques. Ces travaux proviennent de milieux académiques variés, dont principalement les sciences du marketing – et notamment le « *word-of-mouth marketing* », une sous-branche de cette

³² Julien Talpin, « Qualité de la délibération », art. cit.

³³ Voir notamment Elihu Katz et Paul Lazarsfeld, *Influence personnelle*, Paris, Arman Colin, 2008 ou pour une synthèse de ces travaux : Michel Grumbach et Nicolas Herpin, « À propos de quelques travaux de Lazarsfeld et de son école », *Enquête* [En ligne], 4 | 1988, mis en ligne le 27 juin 2013, consulté le 15 octobre 2013. URL : <http://enquete.revues.org/65>.

discipline – et les analyses statistiques et mathématiques des réseaux, et leurs applications aux réseaux sociaux « numériques », désignées par Duncan Watts sous l'appellation de « *the new science of networks* »³⁴.

Dans leur majorité, ces travaux retiennent de la théorie de la communication à double étage – « *two-step flow of communication* »³⁵ – l'idée selon laquelle certains individus, considérés comme des « leaders d'opinion », seraient en mesure d'influencer les décisions (de vote, d'achat, de consommation) d'autres individus de façon bien plus tangible que les médias de masse. Dès lors, dans une perspective de marketing, l'enjeu consiste pour les firmes à cibler leur campagne de communication sur ces « leaders d'opinion », en supposant qu'ils « relaieront » par la suite les messages communiqués à leur entourage. Il serait alors plus efficace – et moins coûteux – de cibler quelques individus dont on sait qu'ils relaieront activement les messages communiqués plutôt que de viser un public de masse.

Pour les tenants du « marketing viral »³⁶, cette logique revêt sur internet une réelle dimension opérationnelle. La quatrième de couverture de la traduction française de *Personal Influence* note ainsi que « *le buzz sur Internet ou bien encore le marketing viral ont remises au goût du jour* »³⁷ les conclusions de l'ouvrage de Lazarsfeld et Katz. En effet, à la différence de la télévision ou la radio, il est possible pour les annonceurs de cibler sur internet les publicités en fonction des caractéristiques individuelles des internautes, du moins lorsque celles-ci sont facilement accessibles, comme le sexe ou le lieu d'habitation (qui peut être identifié automatiquement à partir des données de géolocalisation). Par exemple, une publicité peut être envoyée – et donc potentiellement vue – uniquement à des hommes habitant l'Agglomération parisienne. Afin d'optimiser cette stratégie de ciblage, les tenants du marketing viral entendent donc identifier sur internet et les réseaux sociaux les « leaders d'opinion », mesurer leur « influence », et analyser les logiques de diffusion des messages, les logiques de « viralité ».

Avec les outils de *web-tracking* et de cartographie des réseaux socionumériques, l'« influence » pourrait être alors mesurée sans avoir recours à des données déclaratives. Ces outils seraient donc une alternative aux enquêtes par questionnaire mobilisées par les

³⁴ Duncan Watts, « The “new” science of networks », *Annual Review of Sociology*, n° 30, 2004, p. 243-70 et pour une synthèse de ces approches, cf. Kévin Mellet, « Aux sources du marketing viral », *Réseaux*, 2009/5-6 (n° 157-158), p. 267-292.

³⁵ Elihu Katz et Paul Lazarsfeld, *Influence personnelle*, op. cit.

³⁶ Pour reprendre l'expression mobilisée par Kévin Mellet dans son article de synthèse.

³⁷ Elihu Katz et Paul Lazarsfeld, *Influence personnelle*, op. cit.

chercheurs du *B.A.S.R.* dans leurs différents travaux³⁸ et dont Elihu Katz a lui-même reconnu qu'elles étaient en réalité relativement inadaptées à mesurer l'influence³⁹.

Ainsi, pour certains auteurs⁴⁰, les données délivrées automatiquement par Twitter, telles que les nombres d'abonnés et d'abonnements d'un compte, ou encore le nombre de retweets d'un message, pourraient être appréhendés comme des indices permettant de mesurer l'influence des individus inscrits sur le réseau social. L'enjeu pour le chercheur est alors de déterminer, parmi ces différents indices, lesquels mesurent au mieux l'influence. Ainsi, pour ne donner qu'un exemple – nous reviendrons plus amplement sur cette littérature dans le premier chapitre de cette thèse – pour Bakshy *et alii* le nombre d'abonnés serait un indicateur pertinent pour prédire l'influence d'un individu sur *Twitter* mais, pour être réellement opérant, devrait être associé à un second indicateur, plus complexe, calculé à partir du nombre de retweets moyen des messages publiés par cet individu⁴¹.

Plusieurs critiques peuvent être adressées à ces recherches. La première est d'oublier le caractère profondément localisé de l'influence, du moins telle qu'elle est appréhendée dans les travaux de Katz et Lazarsfeld, dans lesquels elle ne s'exerce qu'au sein de groupes primaires, composés d'un petit nombre d'individus, qui développent entre eux des contacts très fréquents (dont notamment la famille nucléaire et les amis proches). Dans ce modèle, un « leader d'opinion » n'est en capacité d'influencer qu'une petite poignée de personnes (quatre tout au plus), et le plus souvent uniquement dans un seul domaine (le cinéma, la mode, ou les affaires publiques, pour reprendre les domaines étudiés dans *Personal Influence*). Cette dimension localisée de l'influence interpersonnelle est bien souvent oubliée des travaux de marketing viral⁴². Sur les réseaux sociaux numériques, les individus peuvent être « amis » ou « abonnés » à plusieurs centaines d'autres individus et certains messages peuvent être retweetés plusieurs milliers de fois. En considérant le nombre d'abonnés comme un indicateur d'influence, certains auteurs conçoivent alors les leaders d'opinion sur internet comme des individus en capacité d'influencer plusieurs centaines voire plusieurs milliers d'individus. De

³⁸ Paul Lazarsfeld, Bernard Berelson et Hazel Gaudet, *The people's choice: How the voter makes up his mind in a presidential campaign*, New York: Duell, Sloan, and Pearce 1944, p. 50 ; Bernard Berelson, Paul Lazarsfeld et William McPhee, *Voting. A study of Opinion Formation in a Presidential Campaign*, Chicago, The University of Chicago Press, 1954.

³⁹ Elihu Katz, « The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis Author », *The Public Opinion Quarterly*, Vol. 21, n°1, 1957, p. 61-78.

⁴⁰ Voir par exemple Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto et Krishna Gummadi, « Measuring User Influence on Twitter: the Million Follower Fallacy », *ICWSM'10: Proceedings of international AAAI Conference on Weblogs and Social*, mai 2010.

⁴¹ Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis*, *op. cit.*

⁴² Kévin Mellet, « Aux sources du marketing viral », *art. cit.*

plus, ces recherches n'interrogent pas les significations que les agents sociaux confèrent à ces pratiques que le marketing viral appréhende comme étant des « marqueurs » d'influence. S'abonner au compte Twitter d'un individu ou retweeter l'un de ses messages signifie-t-il que l'on est « influencé » par cet individu ? Suffit-il de constater qu'un individu x retweete un message publié par un individu y pour déduire une relation d'influence ? Ces questions ne sont pas abordées par ces recherches, qui adoptent une approche exclusivement quantitative et macrosociologique de l'influence.

Les théories d'Habermas et de Lazarsfeld constituent donc deux des principaux cadres théoriques mobilisés par la littérature consacrée aux données numériques. Cependant, aucun des deux ne nous paraît totalement ajusté à une analyse centrée sur les propriétés sociologiques individuelles des producteurs d'opinions numériques. À l'inverse, bien que très rarement mobilisées par cette littérature, les théories de Pierre Bourdieu sur les opinions politiques constituent selon nous un cadre théorique particulièrement ajusté à une analyse sociologique des opinions numériques.

Les théories de Pierre Bourdieu : un cadre opérant d'analyse des opinions numériques

Les théories de Pierre Bourdieu relatives à la compétence politique et aux modes de production des opinions politiques (exposées notamment dans le huitième chapitre de *La Distinction*⁴³) sont très peu mobilisées dans les travaux académiques consacrés à l'analyse des autopublications numériques. Elles constituent pourtant selon nous un cadre théorique opérant pour analyser et rendre intelligible ce nouveau type de données.

Cependant, mobiliser ces théories afin d'analyser les opinions numériques nécessite d'opérer un double ajustement : *désarticuler* la question des opinions politiques de celle des sondages d'une part et *intégrer* une dimension *conjoncturelle* à l'analyse *dispositionnelle* des modes de production des opinions d'autre part.

⁴³ Pierre Bourdieu, *La Distinction. Critique sociale du jugement*, Paris, Les Éditions de Minuit, 1979.

Désarticuler la question des opinions de celle des sondages

C'est principalement à travers le débat né dans les années 1970 en France autour du concept d'opinion publique (l'opinion au singulier) et de son « nouvel » instrument de mesure – les sondages d'opinion – que Pierre Bourdieu analysa les opinions politiques (au pluriel)⁴⁴. Le sociologue a ainsi démontré que toutes les réponses aux questions de sondage n'étaient pas des « opinions » et que les considérer comme telles et leur attribuer la même valeur faciale était un pur artefact dont la principale fonction est de dissimuler « *que l'état de l'opinion à un moment donné du temps est un système de forces, de tensions* »⁴⁵. Le sociologue a donc défini les opinions politiques avant tout *par la négative*, en dévoilant *ce qu'elles n'étaient pas* nécessairement : des réponses à des questions de sondages.

Parmi les différentes catégories de réponses de sondages, Pierre Bourdieu distingue alors les « opinions politiquement constituées », définies assez sommairement comme « *quelque chose qui peut se formuler en discours avec une certaine prétention à la cohérence* »⁴⁶ et dont la production requiert un certain niveau de compétence politique, et l'expression de dispositions éthiques sur la politique, générées par l'éthos de classe, « *c'est-à-dire un système de valeurs implicites que les gens ont intériorisées depuis l'enfance et à partir duquel ils engendrent des réponses à des problèmes extrêmement différents* »⁴⁷. Seule la première catégorie de réponses – les « opinions politiquement constituées » – constitue donc de « véritables » opinions.

Cependant, ces discours politiquement constitués et l'expression de ces dispositions éthiques sur la politique sont parfois désignés par le sociologue sous le terme générique d'« opinions », entretenant ainsi une certaine confusion définitionnelle de la notion. Ainsi, quand Pierre Bourdieu détaille les trois modes de production des opinions politiques, il inclut dans sa typologie les dispositions éthiques sur la politique, générées par l'éthos de classe⁴⁸. Il nous semble alors nécessaire d'opérer une distinction entre les opinions *au sens large* et les opinions *au sens restreint* de discours politiques prétendant à la cohérence.

⁴⁴ Pierre Bourdieu, « L'opinion publique n'existe pas », *Les Temps modernes*, n°318, janvier 1973, p. 1292-1309.

⁴⁵ *Ibidem*.

⁴⁶ *Ibidem*.

⁴⁷ *Ibidem*.

⁴⁸ Pierre Bourdieu, *La Distinction*, *op. cit.*

Nous qualifierons alors par la suite les opinions (au sens étroit) d'« opinions politiquement constituées » et définirons les opinions (au sens large) simplement comme *des discours* (« éthiques » ou « politiques ») portant sur des objets politiques. Par « objets politiques », nous entendons des objets qui se réfèrent explicitement au champ politique (partis, hommes politiques, institutions, etc.). Nous adopterons donc une définition volontairement restreinte de la politique, ce notamment afin de comparer nos données aux résultats avancés par des recherches consacrées aux inégalités de compétence politique⁴⁹. Nous qualifierons enfin d'« opinion numérique » la formulation écrite d'une opinion (au sens large) sur un espace d'autopublication numérique (blogs, réseaux socionumériques, etc.).

Précisons que suivant notre logique, une *opinion numérique* peut être donc tout à la fois une opinion politiquement constituée ou au contraire une opinion générée par l'éthos de classe. Postuler que l'on puisse identifier sur les espaces d'autopublication numérique des opinions qui soient générées par l'éthos de classe revient à clarifier un second point relativement confus dans les écrits de Pierre Bourdieu, relatif à la réductibilité de ces discours éthiques à des artefacts générés par la situation d'enquête. En effet, alors que dans « L'opinion publique n'existe pas »⁵⁰, les opinions générées par l'éthos de classe sont appréhendées comme des artefacts provoqués par la situation d'imposition de problématique, dans *La Distinction*⁵¹, l'expression de ces dispositions semblent pouvoir être formulées en dehors de la situation d'enquête, par les agents sociaux eux-mêmes.

Nous prendrons ici le parti théorique de ne pas réduire l'expression de ces dispositions éthiques à des artefacts générés par la situation d'enquête. De ce fait, nous dissociions la question des opinions (au pluriel) de leur instrument traditionnel de mesure (les sondages). Si les opinions ne sont pas toutes nécessairement des produits artificiels générés par la situation d'enquête, on peut donc les analyser indépendamment de leur instrument de mesure, les sondages. Autrement dit, affirmer que les agents sociaux peuvent produire des discours politiques générés par l'éthos de classe dans d'autres situations que celle de l'enquête par questionnaire revient à conférer à l'objet (les opinions) une certaine autonomie par rapport à l'instrument (les sondages). Rappelons de plus ici que la spécificité-même des autopublications numériques est de ne pas être des données générées par une interaction de recherche.

⁴⁹ Dans la tradition des recherches ouvertes en France notamment par Daniel Gaxie avec le cens caché, cf. Daniel Gaxie, Daniel Gaxie, *Le cens caché. Inégalités culturelles et ségrégation politique*, Paris, Éditions du Seuil, 1978.

⁵⁰ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

⁵¹ Pierre Bourdieu, *La Distinction*, op. cit.

Dès lors, la question qui se pose au chercheur analysant les opinions numériques est de savoir quels sont les facteurs qui favorisent la production de ces opinions – qu’elles soient « éthiques » ou « politiquement constituées » – si ce n’est pas la situation d’enquête ? Pourquoi un individu exprime-t-il une opinion politique s’il n’est pas contraint de le faire par un chercheur ou un sondeur ? Cette question est en réalité un point-aveugle des théories de Pierre Bourdieu. Afin de pouvoir appréhender totalement la réalité de ce nouveau matériau empirique (les opinions numériques), il nous semble alors nécessaire d’associer aux théories *dispositionnelles* des modes de production des opinions politiques une dimension de nature davantage *conjoncturelle*.

Concilier dimension conjoncturelle et dispositionnelle de production des opinions politiques

L’analyse des modes de production des opinions politiques proposée par Pierre Bourdieu dans *La Distinction* est avant tout dispositionnelle. Ce sont en effet les propriétés sociales des agents sociaux, leur système de dispositions durablement incorporées – leur habitus – qui déterminent leur capacité à produire une opinion politiquement constituée ou au contraire une opinion générée par l’ethos de classe. La question posée n’est pas alors de savoir *quand et dans quel contexte micro ou macro sociologique* les agents sociaux produisent des opinions politiques mais *comment et selon quelles logiques* ils les produisent.

Si Pierre Bourdieu ne fait donc pas explicitement référence à des facteurs de nature conjoncturelle dans son analyse des modes de production des opinions politiques, on retrouve néanmoins en filigrane de ses écrits certains éléments se référant à des aspects extra-dispositionnels⁵².

Ainsi, dans « L’opinion publique n’existe pas »⁵³, Bourdieu fait référence à des « situations de crise », c’est-à-dire des séquences temporellement situées, des situations donc *conjoncturelles*, où les agents sociaux, mêmes les moins compétents politiquement, sont en situation de (re)produire des opinions politiquement constituées :

« Dans les situations où se constitue l’opinion, en particulier les situations de crise, les gens sont devant des opinions constituées, des opinions soutenues par des groupes, en

⁵² Daniel Gaxie, « Retour sur les modes de production des opinions politiques », in Philippe Coulangeon et Julien Duval (dir.), *Trente ans après La Distinction de Pierre Bourdieu*, Paris, La Découverte, 2013.

⁵³ Pierre Bourdieu, « L’opinion publique n’existe pas », art. cit.

sorte que choisir entre des opinions, c'est très évidemment choisir entre des groupes. Tel est le principe de l'effet de politisation que produit la crise : il faut choisir entre des groupes qui se définissent politiquement et définir de plus en plus de prises de position en fonction de principes explicitement politiques »⁵⁴

Bourdieu va même plus loin en précisant qu'en « temps de crise », des agents sociaux qui, d'ordinaire, ne sont pas en capacité de produire par eux-mêmes des opinions politiquement constituées, pourront choisir en « termes de compétence politique » :

« Toutefois, le fait, attesté par les non-réponses, que les dispositions de certaines catégories n'accèdent pas au statut d'opinion, c'est-à-dire de discours constitué prétendant à la cohérence, prétendant à être entendu, à s'imposer, etc., ne doit pas faire conclure que, dans des situations de crise, les gens qui n'avaient aucune opinion choisiront au hasard : si le problème est politiquement constitué pour eux (problèmes de salaire, de cadence de travail pour les ouvriers), ils choisiront en termes de compétence politique ; s'il s'agit d'un problème qui n'est pas constitué politiquement pour eux (répressivité dans les rapports à l'intérieur de l'entreprise) ou s'il est en voie de constitution, ils seront guidés par le système de dispositions profondément inconscient qui oriente leurs choix dans les domaines les plus différents, depuis l'esthétique ou le sport jusqu'aux préférences économiques. L'enquête d'opinion traditionnelle ignore à la fois les groupes de pression et les dispositions virtuelles qui peuvent ne pas s'exprimer sous forme de discours explicite. C'est pourquoi elle est incapable d'engendrer la moindre prévision raisonnable sur ce qui se passerait en situation de crise ».⁵⁵

Et le sociologue de conclure que l'opinion publique est en réalité « un système de forces, de tension et qu'il n'est rien de plus inadéquat pour représenter l'état de l'opinion qu'un pourcentage »⁵⁶. Autrement dit, l'opinion n'est pas un objet statique (un « pourcentage ») mais au contraire un système de force, par essence dynamique, en perpétuelle « tension ». Dès lors, analyser l'opinion nécessite d'étudier la conjoncture dans laquelle prennent forme ces mouvements d'opinion et distinguer les temps de crise des temps de « non-crise » politique.

Dans *La Distinction*, Bourdieu ne fait plus référence à ce « système de forces » mais à un « marché de l'opinion » où se rencontrent une « offre » et une « demande ». Pour le

⁵⁴ *Ibidem.*

⁵⁵ *Ibidem.*

⁵⁶ *Ibidem.*

sociologue, « le fait de produire une réponse à un questionnaire sur la politique, comme le fait de voter ou, à un autre niveau de participation, de lire un journal d'opinion ou d'adhérer à un parti, est un cas particulier de rencontre entre une offre et une demande : d'un côté, le champ de production idéologique, univers relativement autonome, où s'élaborent, dans la concurrence et le conflit, les instruments de pensée du monde social objectivement disponibles à un moment donné du temps et où se définit du même coup le champ du pensable politiquement, ou, si l'on veut, la problématique légitime ; de l'autre, des agents sociaux occupant des positions différentes dans le champ des rapports de classes et définis par une compétence politique spécifique plus ou moins grande [...] »⁵⁷ .

Le cadre analytique du marché proposé par Bourdieu nous semble particulièrement opérant pour analyser les logiques de production et de consommation des opinions politiques, et notamment, nous le démontrerons, sur internet. Le marché de l'opinion, comme n'importe quel marché, peut donc donner lieu à une analyse à la fois *structurelle* et *conjoncturelle*. L'analyse *structurelle* mettra en lumière les logiques sociales de production et de consommation des opinions par les agents sociaux, en fonction de leurs dispositions durables, de leur habitus. L'analyse *conjoncturelle* s'intéressera quant à elle aux évolutions temporelles du niveau de l'offre et de la demande d'opinions politiques, en situation « normale » et en « situation de crise », c'est-à-dire de forte intensité politique.

Nous proposons donc en résumé d'adopter une analyse à la fois dispositionnelle et conjoncturelle de la production des opinions politiques sur internet, à travers le cadre analytique du marché.

L'enjeu qui se pose désormais à nous est de proposer un dispositif méthodologique qui permette d'analyser *empiriquement* ce marché de l'opinion sur internet. Nous n'avons pas trouvé dans la littérature existante de méthodes répondant à nos exigences théoriques. Nous avons alors construit notre propre dispositif méthodologique.

⁵⁷ Pierre Bourdieu, *La Distinction, op. cit.*, p. 465-466.

Notre méthodologie : panéliser un échantillon représentatif d'internautes inscrits sur le réseau social Twitter

Afin de répondre à ces exigences théoriques, le dispositif méthodologique imaginé devait satisfaire à une double exigence : tirer profit des analyses de *web-tracking* sans en reproduire les principales limites.

Échantillonner de façon aléatoire une population d'individus inscrits sur un espace d'autopublication numérique particulier – en l'occurrence Twitter – et lui administrer un court questionnaire d'une vingtaine de questions d'identification socio-démographiques et politiques permet de concilier ces deux exigences. Une fois le questionnaire complété par les individus tirés au sort, plus aucune interaction n'a lieu entre le chercheur et les enquêtés. Le chercheur peut ainsi, à l'aide d'un logiciel spécialisé, archiver et collecter en continu et de façon exhaustive les messages publiés par les enquêtés et les rattacher systématiquement aux propriétés sociologiques et politiques de leurs auteurs, et ce sans jamais interagir avec eux.

Ce dispositif permet d'abord de tirer profit des avantages des logiciels de *web-tracking*. Ainsi, les opinions collectées ne sont pas des réponses à des questions « imposées » aux enquêtés, mais des messages publiés par les enquêtés eux-mêmes, sans que nous leur ayons demandé de le faire. Les internautes ne sont donc pas mis en demeure de répondre à des questions qu'ils ne se sont possiblement jamais posées. Nous n'avons donc pas interrogé ce panel à plusieurs reprises dans le temps, mais simplement enregistré et analysé passivement leur production de messages au fil des mois.

De plus, deuxième avantage, en autorisant une recension exhaustive des messages publiés par chacun des membres du panel, notre dispositif permet, à l'aide du logiciel *Ami Opinion Tracker*, d'opérer un traitement quantitatif des données qualitatives collectées. Les données collectées ne sont pas en effet des réponses à des questions fermées, mais des messages écrits, qui peuvent être représentés soit sous forme de pourcentage soit sous forme de verbatim.

Enfin, troisième avantage tiré des analyses de *web-tracking*, les données publiées sur les réseaux sociaux sont collectées en temps réel, et chacun des messages est précisément horodaté, à la seconde près. On peut donc mesurer la distribution du nombre d'opinions

publiées dans le temps et analyser ainsi les aspects *conjoncturels* de la production des opinions numériques, évoqués plus haut.

Tableau 0.1 Les avantages et limites des analyses de web-tracking et leur contournement par notre dispositif de panélisation du réseau social Twitter

Avantages des analyses de <i>web-tracking</i>	
Données collectées non générées par la situation d'enquête	
Traitement quali-quantitatif et quanti-qualitatif des données collectées	
Données précisément horodatées et collectées en temps réel	
Limites des analyses de <i>web-tracking</i>	Contournement des limites par notre dispositif
Agrège des données hétéroclites	Analyse centrée sur un seul terrain : Twitter
Analyse des flux de messages	Analyse centrée sur un panel d'individus
Absence de données socio-démographiques sur les auteurs des messages	Administration d'un questionnaire socio-démographique d'entrée dans le panel

Mais notre dispositif de panélisation permet également de pallier les trois principales limites des analyses de *web-tracking*. Ainsi, nous avons vu qu'une de ces limites est de collecter et d'agréger des données hétéroclites (billets de blog, tweets, commentaires sur des sites d'information, etc.), difficilement comparables entre elles. Afin de pallier cette limite, nous avons fait le choix de n'étudier qu'un seul espace d'autopublication numérique, le réseau social Twitter, que l'on peut donc considérer comme *notre terrain de recherche*. Twitter est un site internet créé aux États-Unis en 2006. Il permet à ses utilisateurs de publier gratuitement des messages dont la taille ne peut dépasser une limite de 140 caractères : les tweets. En 2014, Twitter revendiquait 241 millions d'utilisateurs actifs dans le monde et 500 millions de tweets publiés chaque jour⁵⁸. En France, en 2013, 94% des internautes déclaraient connaître le réseau social et 17% seulement déclaraient y être inscrits⁵⁹. Le différentiel entre le taux de notoriété et le taux d'inscription sur Twitter est le plus important de l'ensemble des réseaux sociaux⁶⁰. On peut sans doute expliquer ce phénomène par une forte présence d'agents issus du champ médiatique sur ce site. En effet, un très grand nombre de journalistes, d'éditorialistes et de présentateurs d'émissions radiodiffusées ou télévisées sont inscrits sur Twitter et évoquent en retour le site internet dans leurs titres respectifs, ce qui explique alors

⁵⁸ <https://about.twitter.com/company>

⁵⁹ IFOP, « Observatoire des réseaux sociaux », 2013.

⁶⁰ *Ibidem*.

sans doute sa forte notoriété, et ce en dépit du faible taux de Français qui y sont inscrits. C'est donc à ce réseau social particulier que nous avons appliqué notre dispositif méthodologique. Cette focalisation sur un seul espace permettra de mieux saisir les logiques sociales spécifiques qui structurent la production et la diffusion des messages sur ce site. Mais surtout, nous tenterons de le démontrer tout au long de notre démonstration, Twitter est un terrain de recherche qui permet de rendre particulièrement manifeste l'existence d'un marché de l'opinion.

Une deuxième limite des logiciels de *web-tracking* est de centrer leur analyse sur des flux de messages, et non sur les individus qui en sont les auteurs. Si nous avions fait le choix de ne pas mettre en application notre dispositif de panélisation, nous aurions ainsi pu connaître chaque jour le nombre total de tweets politiques publiés sur l'ensemble du réseau social, et comparer l'évolution de ce résultat dans le temps, sans toutefois connaître le nombre d'auteurs uniques de ces messages et surtout sans savoir si ces auteurs étaient les mêmes d'un jour à l'autre. Pour mieux comprendre notre propos, on peut citer ici l'exemple d'une élection à deux tours. S'il s'en tient aux résultats exhaustifs délivrés par le Ministère de l'Intérieur, un politiste pourra connaître le nombre de votes au premier et au second tour de l'élection et constater ainsi une baisse ou une hausse du nombre de votants d'un tour à l'autre. Cependant, il ne pourra déterminer dans quelle mesure les votants du premier tour ont une nouvelle fois voté au second et inversement. Seule une analyse d'un échantillon panélisé d'électeurs représentatifs peut permettre au chercheur de connaître ces informations relatives au *turn-over* électoral. Nous avons donc appliqué cette même logique à l'analyse du réseau social Twitter. Nous n'avons pas analysé l'ensemble des inscrits qui y parlent de politique, mais seulement un échantillon représentatif. Nous pouvons ainsi connaître le nombre de messages publiés chaque jour par le panel, le nombre d'individus qui en sont les auteurs et le nombre d'individus identiques qui ont publié ces messages d'un jour à l'autre.

Enfin, en permettant de situer socialement et politiquement les auteurs des messages collectés, notre dispositif de panélisation permet de pallier la limite la plus fondamentale des analyses de *web-tracking*. En effet, nous l'avons vu, les messages publiés sur internet et collectés par les logiciels ne sont le plus souvent pas associés à un auteur dont on connaît les propriétés sociologiques et politiques. Il est donc dans ce cas impossible pour le sociologue d'analyser les verbatims recueillis et les opinions collectées au prisme des propriétés sociologiques de leurs auteurs. En administrant aux enquêtés un court questionnaire comprenant une vingtaine de questions socio-démographiques (âge, sexe, PCS, niveau de

diplôme, PCS du père, de la mère, etc.) et politiques (orientation politique, pratiques militantes, niveau d'intérêt pour la politique, trajectoire de vote, etc.)⁶¹, le chercheur peut alors situer très précisément les auteurs des opinions numériques dans l'espace social *réel*. L'administration de ce questionnaire se déroule lors de la phase de constitution du panel. Le chercheur doit ensuite « se faire oublier » des enquêtés nouvellement panélisés, afin de ne pas altérer leur future production de tweets.

Enfin, précisons qu'afin de pallier une autre limite méthodologique, celle-ci relative à toute enquête par questionnaire, nous avons constitué en parallèle de ce panel d'individus qui ont accepté de répondre à notre questionnaire – le panel « répondants », composé de 608 individus (soit un taux d'acceptation de 6%) – un panel de non répondants. Ce panel « non répondants » est constitué d'un échantillon représentatif d'individus ($n = 620$) qui n'ont pas répondu à notre questionnaire. Nous avons alors cherché à qualifier sociologiquement ce panel à partir des informations socio-démographiques parcellaires que les individus indiquent sur les espaces d'autopublication numérique sur lesquels ils sont inscrits, telles que leur sexe, leur âge, leur profession ou leur statut social, etc. Ce panel de non répondants se compose donc à la fois d'individus pour lesquels nous disposons d'une ou plusieurs informations socio-démographiques mais également d'individus pour lesquels nous n'en possédons aucune. L'analyse de ce panel « non répondants » nous permettra de déterminer dans quelle mesure les individus qui ont accepté de répondre à notre questionnaire sont représentatifs de l'ensemble de la population des inscrits.

À l'aide de notre logiciel, nous avons alors archivé et analysé l'exhaustivité des tweets publiés par les panels « répondants » et « non répondants » durant 11 mois, du 1er mars 2013 au 31 janvier 2014. L'expérimentation de ce dispositif méthodologique nous a permis de générer des données et des résultats de différentes natures, comme des indices mesurant le nombre de tweets politiques⁶² publiés par individu, le nombre de tweets politiques publiés par individu et par jour, le nombre de tweets politiques publiés quotidiennement par chacun des panels (« répondants » et « non répondants »), mais aussi des sociographies de ces populations, des analyses de leurs descriptions de profil⁶³ et de leurs pratiques d'abonnements à des comptes de médias, de journalistes, de femmes et d'hommes politiques, etc. Enfin, nous avons complété ces données par des entretiens semi-directifs réalisés avec 19 membres de

⁶¹ Cf. questionnaire Annexe 2 p. 467.

⁶² Nous entendons par tweet politique un tweet qui contienne une occurrence définie comme politique, Cf. chapitres 2 et 4.

⁶³ Cf. chapitre 3.

notre panel « répondants ». L'objectif de ces entretiens était d'analyser les représentations de ces individus sur leurs pratiques sociales en ligne et de confronter ces données *déclaratives* aux pratiques *observées* directement en ligne.

En résumé, c'est donc à partir d'un cadre théorique inspiré des thèses de Pierre Bourdieu et à partir d'un dispositif méthodologique que nous avons nous-même construit que nous avons choisi d'analyser ce nouveau matériau empirique – les autopublications numériques – et de mobiliser pour ce faire des outils issus du secteur professionnel des études marketing – les logiciels de *web-tracking*.

Notre méthode de panélisation du réseau Twitter permet-elle de rendre compte de l'existence d'un marché de l'opinion ?

La principale interrogation de cette recherche doctorale peut alors se résumer de la manière suivante : *notre dispositif méthodologique de panélisation d'une population d'internautes inscrits sur le réseau social Twitter permet-il de rendre compte de manière empirique de l'existence d'un marché de l'opinion ?*

Si l'on considère, en suivant la thèse de Bourdieu, qu'il existe un marché de l'opinion, où se rencontrent une « offre » et une « demande », on peut alors admettre que les sondages d'opinion ne constituent pas un instrument adapté à l'analyse de ce marché. En effet, les sondages d'opinion, en reposant sur l'hypothèse d'un consensus entre sondés et sondeurs sur les questions qui méritent d'être posées dans le débat public⁶⁴, tendent à assimiler de manière relativement indifférenciée « demande » et « offre » d'opinions. Tous les individus étant considérés comme politiquement compétents, il ne saurait exister de différence ni de frontière entre les agents sociaux situés à l'intérieur et ceux situés à l'extérieur du champ de production idéologique.

Notre dispositif d'enquête permet au contraire de dissocier empiriquement et analytiquement « offre » et « demande » d'opinions. Twitter constitue en effet selon nous un double observatoire : un observatoire du champ de production de l'opinion – soit un observatoire de l'« offre » – mais également un observatoire du « pays légal » (ou un

⁶⁴ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

observatoire de la « demande ») ; « pays légal »⁶⁵ que l'on peut définir en suivant les termes de Pierre Bourdieu comme l'ensemble des agents sociaux, fortement politisés, qui se sentent autorisés à s'exprimer politiquement tout en se situant à l'écart – ou pour certains à la périphérie – du champ de production de l'opinion.

Twitter, un observatoire de l'« offre » d'opinions

Twitter peut être en effet appréhendé comme un observatoire de l'« offre » d'opinions. Nous l'avons dit, la très grande majorité des médias (qu'ils soient télévisés, radiophoniques, de presse écrite ou de presse numérique), des agents médiatiques (journalistes, animateurs, éditorialistes, etc.), mais aussi des partis politiques et des professionnels de la représentation politique, ont, en 2014, un compte Twitter. Pour ne prendre que l'exemple des médias, la plupart relaie sur leurs comptes, *via* des tweets, les articles parus sur leur site internet (sites internet où se trouvent répertoriés, en accès gratuit ou payant, l'ensemble de leur production parue sur leur support d'origine). Comptabiliser le nombre de tweets comprenant par exemple l'occurrence « mariage pour tous » publiés sur le compte Twitter du journal *Le Monde* revient alors à mesurer le volume d'articles publiés par ce titre de presse à ce sujet. L'offre d'informations et d'opinions n'est donc pas produite « sur » ou « pour » Twitter, mais peut être mesurée « *via* » Twitter.

Afin d'analyser l'« offre » d'opinions dans son ensemble, du moins telle qu'elle se donne à voir sur Twitter, nous mobiliserons le « Panel Twitter audience politique », produit dans le cadre d'une recherche collective menée parallèlement à la nôtre⁶⁶. Ce panel permettra ainsi de mesurer l'activité de production du champ de l'information et de l'opinion sur Twitter, bref, de mesurer l'« offre » d'opinions. Nous pourrions alors confronter ce panel à notre propre panel représentatif, qui permet quant à lui de mesurer l'activité de la « demande ».

⁶⁵ Pierre Bourdieu, *La Distinction, op. cit.*, p. 480.

⁶⁶ Contrairement aux panels représentatifs « répondants » et « non répondants », le panel « Twitter audience politique » n'a pas été réalisé dans le cadre de cette recherche doctorale. Nous tenons ici à remercier l'ensemble des personnes qui ont contribué à sa réalisation : Emmanuelle Rimbault, Jean-Yves Dormagen, Marie Neihouser, Rémi Andris, Geoffrey Della Valentina, Émeline Rochefeuille et Lysne Fröydis.

Twitter, un observatoire de la « demande » d'opinions

Notre panel représentatif d'agents sociaux inscrits sur Twitter permet en effet de mesurer l'activité de « la demande ». Nous verrons ainsi que les individus tirés au sort qui figurent dans notre panel représentatif sont très fortement politisés, produisent des tweets politiques, « parlent » de politique sur le réseau social, mais n'appartiennent pas au champ de production de l'opinion et de l'information. Ces agents sociaux ne produisent pas l'information, ils la commentent, ils la « consomment ».

Pour être tout à fait précis, il convient ici de dissocier *production de message* et *production d'opinion*. Les individus qui ont répondu à notre questionnaire sont ainsi tous des « producteurs » de tweets politiques : ils composent et publient eux-mêmes des messages d'au maximum 140 caractères relatifs à des objets politiques. Pour autant, ils ne sont pas des producteurs d'opinions. Nous démontrerons ainsi que les opinions ou les informations évoquées, commentées ou discutées dans leurs messages ont été produites par des agents appartenant aux champs médiatiques et politiques. Cependant, la *production d'un message politique* sur Twitter rend manifeste *le mode de consommation d'une opinion ou d'une information*.

Twitter matérialise en effet certains modes de consommation des opinions et des informations politiques, modes de consommation que l'on qualifiera d'« ostentatoires ». Retweeter un message publié par un compte de média, de journaliste ou d'homme politique constitue ainsi un mode de consommation (parmi d'autres) des opinions ou des informations politiques, tout comme le fait d'évoquer dans un tweet publié en première personne un sujet inscrit à l'agenda médiatique.

L'analyse des comptes à l'origine des retweets politiques publiés par les membres de nos panels représentatifs ainsi que l'analyse des volumes de production du panel « audience politique » (« l'offre ») et du panel représentatif (« la demande ») permettront ainsi de rendre compte empiriquement de l'existence d'un marché de l'opinion sur Twitter ; un marché régi à la fois par des logiques structurelles et conjoncturelles.

Plus encore, nous tenterons de démontrer, à l'appui de données empiriques générées par notre dispositif méthodologique, que le niveau de l'activité de production de l'« offre » d'opinions est statistiquement corrélé à celui de la demande, et que l'on peut déduire de cette corrélation une relation de causalité, c'est-à-dire affirmer que le niveau de production de

l'offre détermine le niveau de consommation de la demande. Dit autrement, plus les agents du champ de production de l'information et de l'opinion produisent des « biens » politiques (production mesurée sur Twitter *via* le volume de tweets publiés par ces agents) et plus la « demande » consomme ces biens (consommation mesurée *via* le volume de tweets et de retweets par des agents situés à l'extérieur de ces champs).

Plan de thèse

Notre démonstration s'articulera autour de six chapitres : deux premiers chapitres théoriques et méthodologiques, trois chapitres empiriques et enfin un sixième chapitre conclusif.

Dans notre premier chapitre, nous reviendrons sur les usages scientifiques et commerciaux des autopublications numériques. Nous verrons ainsi que ce sont d'abord les secteurs professionnels des études de marché et de la veille économique qui ont investi ce nouveau matériau et créé leurs outils d'analyse et de collecte, les logiciels de *web-tracking*. Puis, nous verrons que ces nouveaux outils se sont par la suite diffusés au sein du champ académique, ou plutôt au sein de certaines de ses disciplines, comme notamment les sciences de l'information et de la communication ou les sciences du marketing. À l'inverse, la science politique et la sociologie politique n'ont accordé à ces outils qu'un intérêt très limité. De plus, les travaux académiques qui analysent les autopublications numériques mobilisent des cadres théoriques pour certains relativement éloignés de ceux de la sociologie politique.

Nous défendrons ensuite dans notre deuxième chapitre que les théories de Pierre Bourdieu relatives à la production des opinions politiques et à la déconstruction de la notion d'opinion publique constituent un cadre particulièrement opérant pour analyser et rendre intelligibles ces nouvelles données empiriques. Nous verrons notamment comment les propriétés spécifiques des autopublications numériques et les possibilités méthodologiques offertes par les outils de *web-tracking* permettent au sociologue de construire un dispositif d'analyse des opinions qui ne reproduisent pas les mêmes biais épistémologiques que les sondages d'opinions et les enquêtes par questionnaire. C'est dans cette perspective que nous exposerons notre propre dispositif méthodologique de panélisation du réseau social Twitter ainsi que les conditions concrètes de sa réalisation.

Dans une seconde partie, nous présenterons les résultats issus de l'expérimentation de notre dispositif méthodologique.

Dans notre troisième chapitre, nous proposerons ainsi au lecteur une sociologie de nos enquêtés et démontrerons que Twitter constitue un véritable observatoire du « pays légal ». Les individus qui publient des tweets politiques – qu'ils aient ou non répondu à notre questionnaire – présentent en effet des propriétés sociologiques et politiques tout à fait singulières. Les producteurs de tweets politiques sont ainsi majoritairement des hommes, diplômés, urbains, issus de milieux sociaux relativement aisés.

Dans notre quatrième chapitre, nous analyserons la production de tweets politiques de cette population. Nous verrons ainsi que le nombre de tweets politiques publiés est très inégal d'un individu à l'autre, mais également, au niveau des panels pris dans leur ensemble, d'un jour à l'autre. Nous tenterons alors de mettre en lumière les facteurs dispositionnels et conjoncturels qui président à la production de tweets politiques. Du côté des facteurs dispositionnels, nous démontrerons que ce sont principalement l'âge et le niveau d'intérêt pour la politique qui expliquent ces disparités de production individuelles. Concernant les facteurs contextuels, nous verrons que c'est le niveau de production de tweets politiques des institutions et des agents issus du champ de production de l'opinion et de l'information qui expliquent les disparités temporelles de production de tweets politiques. Dit autrement, sur Twitter – que l'on peut appréhender comme un marché de l'opinion – le niveau de l'« offre » détermine celui de la « demande ».

Nous reviendrons ensuite plus en détail dans notre cinquième chapitre sur ce marché de l'opinion, en l'analysant plus particulièrement sous l'angle de la demande. Nous mettrons ainsi en avant les logiques sociales et politiques de consommation des opinions politiques sur Twitter. Si l'on conçoit en effet Twitter comme un marché, il convient alors de s'interroger sur les logiques qui orientent les individus à consommer certains biens plutôt que d'autres. Nous verrons alors que malgré l'immédiateté et la gratuité des biens informationnels disponibles sur le réseau, la consommation de ces « biens » répond toujours à des logiques sociologiques et politiques très structurantes.

Enfin, dans un sixième chapitre, nous achèverons notre démonstration en résumant les apports théoriques de notre dispositif méthodologique mais questionnerons également ses principales limites. Nous verrons ainsi que si elle permet bien de rendre compte de façon empirique des logiques structurelles et conjoncturelles du « marché de l'opinion » que permet

de rendre manifestes Twitter, notre méthode de panélisation se trouve en revanche relativement inadaptée à saisir d'une part les logiques de tension et de concurrence qui parcourent l'« offre », et d'autre part les logiques de consommation horizontales – ou les logiques de diffusion de pair à pair – qui structurent la « demande ». Nous présenterons alors pour terminer quelques pistes de réflexions relatives à l'élaboration de dispositifs méthodologiques alternatifs permettant de pallier ces limites.

**PREMIÈRE PARTIE : LES AUTOPUBLICATIONS NUMÉRIQUES
COMME SOURCE DE RENOUVELLEMENT THÉORIQUE ET
MÉTHODOLOGIQUE DE L'ANALYSE DES OPINIONS POLITIQUES**

Dans cette première partie nous questionnerons l'intérêt heuristique des autopublications numériques pour les sciences sociales et plus particulièrement pour la sociologie politique. Nous verrons que les messages publiés par les agents sociaux sur internet constituent un véritable matériau d'analyse du monde social. Les autopublications numériques présentent néanmoins un certain nombre de spécificités par rapport à d'autres types de données empiriques, comme les verbatims recueillis lors d'entretiens d'enquête ou les réponses à des questionnaires fermés.

Le premier enjeu de cette partie sera alors d'analyser les spécificités de ces données numériques et de questionner la manière dont elles ont été appréhendées par les agents académiques et économiques qui les ont étudiées. Nous verrons que les théories et les méthodologies employées sont relativement éloignées des traditions d'analyse des opinions en sociologie politique (chapitre 1). Le second enjeu sera de montrer ce que l'analyse des autopublications numériques pourrait apporter aux recherches de sociologie politique. Nous démontrerons alors que notre dispositif méthodologique de panélisation du réseau social Twitter permet de tirer profit des spécificités de ce nouveau matériau tout en adoptant un cadre analytique plus proche des traditions d'étude des opinions en sociologie politique (chapitre 2).

CHAPITRE 1 - LES USAGES COMMERCIAUX ET SCIENTIFIQUES DES AUTOPUBLICATIONS NUMÉRIQUES

En 2011, 35,55 % des Français âgés de 16 à 74 ans déclaraient publier du contenu sur les réseaux sociaux⁶⁷. Encore inexistante au début des années 1990, l'autopublication numérique serait aujourd'hui pratiquée par plus d'un tiers de la population française. Le développement des espaces d'autopublication (plateformes de blogs, forums électroniques, réseaux sociaux, etc.) a permis aux internautes ne disposant pas de compétences informatiques spécifiques de publier facilement et sans intermédiaire des contenus écrits, visuels, audio ou vidéo⁶⁸. Le contenu de ces publications est varié et embrasse toutes les dimensions du monde social. Chaque jour, des milliers d'individus commentent ainsi des sujets d'actualité, des aspects de leur vie privée ou des produits de consommation.

L'étude de ces messages, avis, commentaires, jugements ou opinions – que l'on désignera à ce stade sous le terme générique d'« autopublications » afin de faciliter le travail ultérieur de définition et de délimitation de la notion d'« opinion numérique » – peut être appréhendée comme un nouveau moyen d'accès à la connaissance du monde social. Les autopublications sont en effet autant d'indices de pratiques, d'attitudes et d'opinions exprimées par les internautes. De plus, le développement d'instruments de mesure des données numériques à grande échelle a permis de rationaliser l'étude de ce nouveau matériau, en faisant des autopublications un objet quantifiable.

Si ces nouveaux instruments sont aujourd'hui mobilisés dans des travaux académiques, c'est bien souvent à l'intérieur des champs professionnels du marketing, de la communication et des études de marché qu'ils ont été créés⁶⁹. À l'instar des sondages⁷⁰ ou des *focus groups*⁷¹, l'histoire de ces nouvelles méthodes d'enquête se situe à la croisée des champs scientifiques et commerciaux.

⁶⁷ Eurostat, « Statistiques sur la société de l'information », DEPS, 2013.

⁶⁸ Dominique Cardon *et alii*, « Présentation », *Réseaux*, 3/2006 (n° 137), p. 9-25.

⁶⁹ Dominique Boullier et Audrey Lohard, *Opinion mining et Sentiment analysis : Méthodes et outils*, Marseille, OpenEdition Press, 2012.

⁷⁰ Loïc Blondiaux, *La Fabrique de l'opinion. Une histoire sociale des sondages*, Paris, Le Seuil, 1998.

⁷¹ Sophie Duchesne et Florence Haegel, *L'entretien collectif. L'enquête et ses méthodes*, Paris, Armand Colin, 2008.

Paradoxalement, alors que l'autopublication numérique constitue bien une nouvelle forme possible d'expression des opinions, les agents – scientifiques et commerciaux – qui ont investi ces nouveaux terrains et ces nouveaux instruments ne sont pas issus du champ de l'opinion. Les chercheurs et les professionnels spécialistes des sondages et de l'opinion publique n'ont en effet accordé un intérêt que marginal aux autopublications. Dans le champ scientifique, la sociologie politique a consacré très peu de travaux à cet objet⁷², contrairement à d'autres disciplines – en premier lieu desquelles les sciences de l'information et de la communication – qui en ont fait un axe de recherche majeur.

L'objet de ce premier chapitre est moins d'expliquer ce désintérêt que d'identifier les cadres théoriques, les notions et les concepts mobilisés par des disciplines scientifiques et des acteurs commerciaux jusqu'alors relativement étrangers aux problématiques sociologiques de l'opinion publique. La réflexion théorique sur ces nouveaux instruments de mesure des opinions s'inscrit-elle et prolonge-t-elle le débat épistémologique initié dès les années 1970 en France autour de la question des sondages, de l'opinion publique et de la compétence politique ?

Nous reviendrons brièvement dans une première section sur la sociogenèse de cette nouvelle pratique sociale et les principaux facteurs qui ont présidé à sa démocratisation (1). Nous aborderons ensuite les particularités heuristiques des autopublications numériques, appréhendées cette fois en tant que matériau d'étude (2). Enfin, nous tenterons d'objectiver l'intérêt différencié des agents scientifiques et commerciaux pour ce nouveau matériau en nous attachant à identifier les notions, les concepts et les définitions de l'opinion mobilisés (3).

⁷² On ne recensait en 2013, au moment de l'écriture de cette thèse, qu'un seul article prenant pour objet des autopublications paru dans la *Revue Française de Science Politique* : Sylvain Paraisie et Jean-Philippe Cointet, « La presse en ligne au service de la démocratie locale. Une analyse morphologique de forums politiques », *Revue Française de Science Politique*, 2012/1 (Vol. 62), p. 45-70.

1. L'autopublication numérique, une nouvelle pratique sociale

Si la volonté de démocratiser la publication et la diffusion de contenus informationnels a constitué un des principes au fondement de la création d'internet⁷³, il a fallu attendre le début des années 2000 et la démocratisation de l'accès à internet ainsi que le développement d'outils simplifiés de publication numérique pour que l'autopublication se démocratise réellement⁷⁴ (1.1). Mais en se démocratisant, l'autopublication s'est également diversifiée, devenant moins une pratique journalistique amateur qu'une forme ordinaire d'expression et d'interaction sociale (1.2).

1.1 Les facteurs de démocratisation

Selon Patrice Flichy, les « *pionniers du web* » avaient pour objectif de démocratiser l'écriture, la publication et la diffusion des informations en « *court-circuitant les filtres éditoriaux* » et en abolissant la frontière entre consommateurs et producteurs d'informations⁷⁵. Les outils proposés aux internautes devaient être alors « *suffisamment simples d'accès pour être appropriables par tout un chacun* ». Mais, comme l'a rappelé Dominique Cardon, les premiers outils de publication numérique nécessitaient en réalité la possession d'une certaine compétence technique, une maîtrise du langage informatique. L'autopublication sur internet était dès lors quasi-exclusivement pratiquée par des individus possédant cette culture informatique, excluant ainsi une très large partie de la population. Jusqu'à la fin des années 1990, la démocratisation de la publication était restée à l'état d'utopie et ne s'était pas concrétisée.

Enclenchée au début des années 2000, la dynamique de démocratisation de l'autopublication a été provoquée par deux phénomènes concomitants qui ont connu chacun un développement exceptionnellement rapide : la démocratisation de l'accès à internet et le développement d'outils simplifiés d'autopublication.

Une première condition, nécessaire mais non suffisante, à la démocratisation de l'autopublication numérique a été en effet la démocratisation de l'accès à internet. Alors

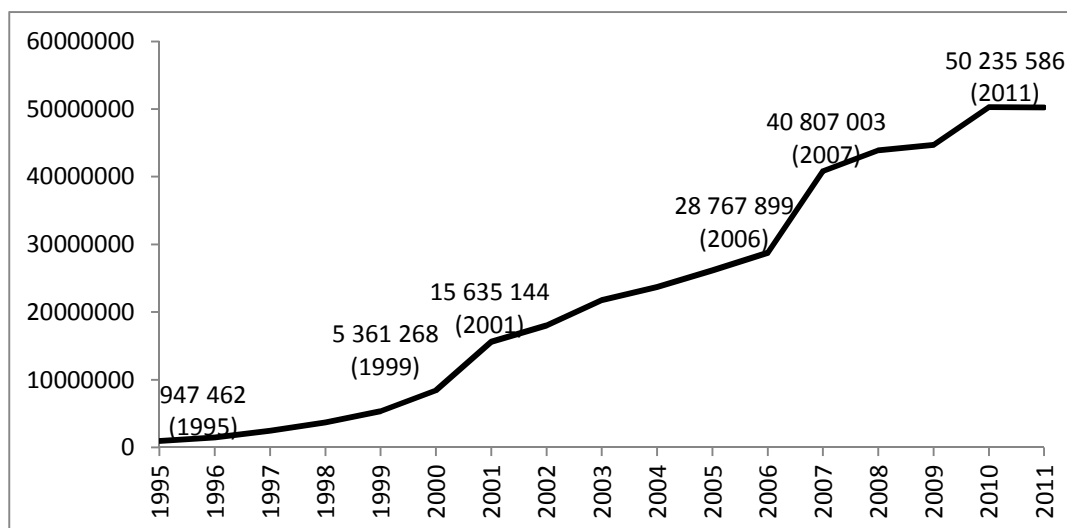
⁷³ Patrice Flichy, *L'imaginaire d'internet*. Paris, La Découverte, 2001.

⁷⁴ Dominique Cardon *et alii*, « Présentation », *Réseaux*, art. cit.

⁷⁵ Patrice Flichy, *L'imaginaire d'internet*, *op. cit.*

qu'en 1995 la France comptait moins d'un million d'internautes, plus de 75 % des Français disposaient d'un accès à internet à leur domicile en 2011, seize ans plus tard⁷⁶.

Graphique 1.1 Évolution du nombre d'internautes en France de 1995 à 2011



Source : Banque mondiale

Se connecter à internet est devenu, pour une grande majorité de Français, une pratique sociale quotidienne, routinière. En augmentant la vitesse de navigation tout en diminuant les coûts d'accès, le développement du haut débit dans les années 2000 a ainsi considérablement impacté la fréquence et la durée moyenne de connexion au réseau. En seulement 5 ans, la part des internautes déclarant se connecter tous les jours à internet est passée de 26,04 % en 2006 à 61,74 % en 2011⁷⁷.

Si la diffusion d'internet a été extrêmement rapide, elle ne s'est pas déployée à la même vitesse dans toutes les catégories de la population. La pénétration beaucoup plus lente du réseau dans les classes populaires, les territoires ruraux ou les populations âgées a ainsi longtemps été au centre du débat public et académique sur la « fracture numérique »⁷⁸. Aujourd'hui, même si des différences substantielles demeurent⁷⁹, l'ensemble des catégories

⁷⁶ Données disponibles sur le site internet de la Banque mondiale, 2013. <http://donnees.banquemondiale.org>.

⁷⁷ Eurostat, « Statistiques sur la société de l'information », DEPS, 2013.

⁷⁸ On peut citer entre autre ici les numéros spéciaux des revues *Les cahiers du numériques*, Vol. 5, 2009/1 « Fracture numérique et justice sociale » et *Réseaux*, n° 127-128, 2004/5-6, « La fracture numérique ».

⁷⁹ INSEE, enquête « Technologies de l'information et de la communication », avril 2010.

socioprofessionnelles sont très majoritairement équipées d'un accès à internet. En 2010, plus de trois ouvriers sur quatre disposaient d'une connexion à internet à leur domicile, même si ce taux restait encore inférieur de 17 points (78,9 %) à celui des cadres (96,1 %). Seules les catégories les plus âgées de la population restent aujourd'hui encore relativement sous-équipées. 44,3 % des 60-74 ans et seulement 8,3 % des plus de 75 ans disposaient en 2010 d'une connexion à internet à leur domicile. À l'inverse, ce taux s'élevait à 93,3 % pour les 15-29 ans. L'accès à internet décroît donc significativement en fonction de l'âge. Cependant, tout laisse à penser que ces différences sont davantage liées à un effet de génération qu'à un effet d'âge et qu'elles disparaîtront donc à moyen terme.

Tableau 1.1 Taux de pénétration d'internet dans les foyers Français par PCS (en %)

	Internet à domicile	Connexion quotidienne
Agriculteurs	70,4	61,0
Artisans et commerçants	84,0	76,0
Cadres et prof. libérales	96,1	91,9
Prof. intermédiaires	89,5	84,5
Employés	81,4	74,6
Ouvriers	73,3	72,9
Étudiants	95,2	81,9
Retraités	30,7	71,1
Total	64,4	78,6

Source : INSEE

Si la démocratisation de l'accès à internet a été une condition nécessaire à la démocratisation de l'autopublication, elle n'a pas été une condition suffisante. Un autre facteur a favorisé la démocratisation de cette pratique sociale : la création d'outils simplifiés d'autopublication numérique.

L'apparition des plateformes de blogs à la fin des années 1990 puis des réseaux sociaux dans les années 2000 a en effet permis aux individus ne disposant pas de compétences informatiques spécifiques de publier facilement et sans intermédiaire du contenu sur internet⁸⁰. Ces nouveaux dispositifs ont été précisément conçus pour simplifier et automatiser la publication numérique. Contrairement aux sites personnels, les blogs permettent de publier

⁸⁰ Dominique Cardon *et alii*, « Présentation », *Réseaux*, art. cit.

des articles à partir d'une interface s'apparentant à un traitement de texte simplifié, et ne nécessitent aucune maîtrise du langage informatique. En hybridant les fonctions de publication et de communication⁸¹, ces nouveaux outils ont fait de l'autopublication une pratique ludique et interactive de sociabilité. Ils ont ainsi fortement abaissé le seuil minimal de compétence technique nécessaire à la publication de contenus sur internet.

Dès leur apparition, ces dispositifs ont connu un développement extrêmement rapide. Alors qu'on ne comptabilisait que vingt-trois blogs en 1999, on en recensait près de vingt-un million en 2005 et plus de cent cinquante-six millions en 2011⁸². Même si ces données sont à prendre avec beaucoup de précaution – leur communication participant d'un véritable enjeu commercial – elles n'en restent pas moins révélatrices de la vitesse de diffusion de ces nouveaux outils de publication. Les réseaux sociaux ont connu un succès tout aussi rapide. Alors que l'inscription sur le réseau social Facebook n'a été ouverte au grand public qu'en 2006, 37 % des internautes français y étaient déjà inscrits trois ans plus tard⁸³. En 2012, ce sont plus de la moitié d'entre eux (54 %) qui déclaraient y être inscrits. Créé en 2006, le site de micro-blogging Twitter a quant à lui connu une croissance moins forte. En 2012, seuls 12% des internautes Français déclaraient y être inscrits.

À des échelles différentes, ces réseaux sociaux rassemblent des catégories de population sociologiquement très diverses. En 2012, 53 % des internautes ouvriers interrogés et 63 % des cadres déclaraient posséder un compte Facebook. Les espaces d'autopublications ont donc pénétré l'ensemble des catégories socioprofessionnelles. Au final, en 2011, 35,55 % des Français âgés de 16 à 74 ans déclaraient publier du contenu sur les réseaux sociaux.

⁸¹ Dominique Cardon *et alii*, « Présentation », *Réseaux*, art. cit.

⁸² Philippe Martin, « L'état de la blogosphère 2008 selon Technorati indique 133 millions de blogs », article en ligne, consulté le 13 juillet 2013 : <http://www.paperblog.fr/1114128/l-etat-de-la-blogosphere-2008-selon-technorati-indique-133-millions-de-blogs/>

⁸³ IFOP, « Observatoire des réseaux sociaux », novembre 2012.

Tableau 1.2 Taux d'inscription à Facebook et Twitter (en %) par PCS

	Facebook	Twitter
Artisans et commerçants	47	7
Cadres	63	15
Professions intermédiaires	54	18
Employés	54	11
Ouvriers	53	11
Ensemble des répondants	54	12

Source : IFOP 2012

Les espaces d'autopublications et les réseaux sociaux en particulier, en simplifiant et en hybridant les fonctions de publication et de communication ont donc amené un public plus divers socialement à pratiquer l'autopublication.

2. La diversification de la pratique sociale d'autopublication

En se démocratisant, l'autopublication s'est également diversifiée, devenant moins une pratique journalistique amateur qu'une forme ordinaire d'expression et d'interaction sociale.

Les premiers blogs et sites personnels – ou du moins la représentation qu'en donnaient les médias et certains travaux académiques⁸⁴ – concevaient en effet l'autopublication comme une production amateur voire semi-professionnelle de contenus informationnels ou culturels, de façon relativement fidèle aux aspirations des pionniers d'internet. Cette pratique de l'autopublication prenait alors racine dans une tradition plus ancienne, celle du journalisme alternatif, dans la continuité du pamphlet, de la radio libre ou des fanzines⁸⁵. Internet a néanmoins bouleversé l'économie de cette presse alternative, en augmentant considérablement la production et la diffusion de ces publications tout en diminuant considérablement les coûts.

L'apparition des réseaux sociaux a transformé, nous l'avons vu, l'autopublication en une activité ludique et sociale de partage et d'échange de contenus numériques. En France, la plateforme Skyblog, créée en 2002, a considérablement rajeuni et féminisé le public des

⁸⁴ Dominique Cardon *et alii*, « Présentation », *Réseaux*, art. cit.

⁸⁵ Dominique Cardon et Fabien Granjon, *Médiactivistes*. Paris, Les presses de Science Po, 2010.

blogueurs. Le succès de cette plateforme a ainsi mis en exergue que la composition sociologique aussi bien que le contenu de la majorité des blogs différaient de la conception « élitiste » de l'autopublication. Davantage centrées sur la vie quotidienne des adolescents, ces autopublications ne s'inscrivaient pas dans un projet éditorial et leurs auteurs ne cherchaient pas nécessairement à être lus par un large public⁸⁶.

Depuis, de nombreux espaces concurrents d'autopublication se sont créés durant la seconde moitié des années 2000, disposant chacun d'un public avec des caractéristiques sociologiques spécifiques⁸⁷. Même si elles dépendent souvent pour partie de la thématique du réseau social et de la sociologie du public qui le fréquente, le contenu des autopublications est infiniment varié et concerne aussi bien des aspects de la vie privée que des sujets d'actualité, de société, des produits de consommation ou des activités culturelles, etc.

Le passage d'une pratique « élitiste » de l'autopublication, conçue comme une production journalistique amateur, à une pratique plus diversifiée socialement d'expression et d'interaction « ordinaire » est en réalité plus idéale-typique qu'empirique, car exagérément linéaire. La pratique de l'autopublication est si diverse qu'il serait d'ailleurs plus juste de raisonner en termes de *continuums* à plusieurs dimensions plutôt que de mobiliser des catégories binaires pour caractériser les différentes catégories de messages publiés par les internautes. Cette opposition rend néanmoins compte du phénomène de diversification sociale qu'a connu cette pratique, sous l'impulsion conjointe de la démocratisation de l'accès à internet et de l'apparition des réseaux sociaux.

L'autopublication a donc connu un développement exceptionnellement rapide. La démocratisation de l'accès à internet et le développement d'outils simplifiés de publication numérique en ont fait une pratique sociale de masse. En se démocratisant, l'autopublication s'est également diversifiée. Aujourd'hui, quotidiennement, des milliers de Français publient des messages sur les blogs, les forums de discussions ou les réseaux sociaux. Le contenu de ces messages est infiniment varié : actualité, politique, pratiques culturelles, vie privée, etc. L'étude de ces autopublications peut alors constituer un véritable matériau d'enquête pour étudier le monde social.

⁸⁶ Dominique Cardon, « Le design de la visibilité. Un essai de cartographie du web 2.0 », *Réseaux*, 2008/6, n°152.

⁸⁷ IFOP, « Observatoire des réseaux sociaux », novembre 2012.

2. Les spécificités heuristiques d'un nouveau matériau d'étude du monde social

À l'instar des réponses de sondages ou des verbatims recueillis lors d'entretiens d'enquête, les autopublications publiées quotidiennement sur les blogs, les réseaux sociaux, les forums ou les sites d'information constituent un véritable matériau empirique pour les chercheurs en sciences sociales. Les messages publiés sur internet sont en effet des indices révélateurs de pratiques, d'attitudes ou d'opinions exprimées par des internautes.

Les autopublications numériques présentent néanmoins certaines particularités heuristiques. À la différence d'autres matériaux d'enquête, elles ne sont pas des données générées artificiellement par le questionnement ou l'interaction avec un chercheur. Leur analyse peut donc se prévaloir de ne pas reproduire les artefacts inhérents aux enquêtes déclaratives. Mais si les chercheurs n'altèrent pas l'« authenticité » des verbatims recueillis, ils ignorent cependant bien souvent leurs conditions sociales de production. Cette absence de données contextuelles peut alors donner lieu à deux stratégies de recherche. La première consiste à compléter les informations manquantes par un travail empirique de production *ex post* de données de « première main ». La seconde à faire partiellement ou totalement abstraction de ces données contextuelles afin de privilégier un traitement quantitatif (2.1).

Le possible dépassement de l'opposition entre approches qualitatives et quantitatives constitue un deuxième apport heuristique des autopublications numériques. En raison de leur abondance et de leur facilité d'accès, de collecte et de traitement, les autopublications apparaissent comme un matériau particulièrement adapté à des méthodes combinant approches qualitatives et quantitatives. Le développement de nombreux indicateurs et instruments de mesure des autopublications permet ainsi d'opérer un traitement quantitatif de données qualitatives mais aussi une analyse qualitative de données quantitatives (2.2).

Enfin, l'activité de publication continue des internautes et l'automatisation de l'opération de collecte rendent possible une étude dynamique et en temps réel des autopublications. L'objectivation et la modélisation de la circulation des autopublications dans le temps et dans l'espace constituent un troisième apport heuristique de ce nouveau matériau d'étude du monde social (2.3).

2.1. Un matériau produit hors laboratoire

Contrairement aux matériaux empiriques générés par les méthodologies « classiques » des sciences sociales, les autopublications apparaissent comme des données non stimulées, non impactées par l'intervention du chercheur (2.1.1). Mais n'étant précisément pas produites et récoltées en situation de « laboratoire », les conditions sociales et contextuelles de production des autopublications numériques peuvent alors partiellement ou totalement échapper au chercheur (2.1.2). Face à ce manque de données, deux stratégies de recherche peuvent être alors adoptées (2.1.3).

2.1.1. Des verbatims non altérés par la situation d'enquête

Les autopublications présentent en effet la particularité de ne pas être des données produites en situation de laboratoire. Formule empruntée aux « sciences dures », la situation de « laboratoire » représente symboliquement en sciences sociales la phase de production et de collecte de données empiriques, qu'elles soient de nature qualitative (entretiens individuels ou collectifs, observations ethnographiques, etc.) ou quantitative (sondages, recensements, etc.). L'usage de matériaux empiriques en sciences sociales, théorisé dès le dix-neuvième siècle par Émile Durkheim, repose sur la volonté de pratiquer les sciences sociales selon des règles analogues aux sciences de la nature. Cependant, à la différence des sciences dures, les objets des sciences sociales – les hommes en société – sont « *des êtres qui pensent, parlent et font circuler des récits, des êtres qui peuvent s'approprier le discours des scientifiques et modifier leurs comportements en conséquence* »⁸⁸. L'adaptabilité des hommes à leur environnement et à la science – dont ils sont à la fois le sujet et l'objet – constitue précisément « le malheur »⁸⁹ des sciences sociales. Les enquêtés peuvent attribuer au chercheur des intentions qui ne sont pas les siennes mais qui vont conditionner leurs réponses, leur manière de parler, ce qu'ils vont choisir de montrer ou au contraire de dissimuler, etc.⁹⁰. Dès lors, tout

⁸⁸ Michel Grossetti, « Plus largement, les sciences sociales... », *SociologieS* [En ligne], Débats, La situation actuelle de la sociologie, mis en ligne le 27 janvier 2012, consulté le 14 octobre 2013. URL : <http://sociologies.revues.org/3788>

⁸⁹ Pierre Bourdieu, Jean-Claude Chamboredon et Jean-Claude Passeron, *Le métier de sociologue. Préalables épistémologiques*, Paris, Éditions de l'EHESS, 1968.

⁹⁰ Philippe Baumard, « Quelles approches avec quelles données ? » in Marie-Laure Gavard-Perret, David Gotteland, Christophe Haon et Alain Jolibert (dir.), *Méthodologie de la recherche en gestion*, Paris, Nathan, 1998.

recours au questionnement ou à l'observation produit des effets propres à la situation d'enquête, qui ne se seraient sans doute pas produits dans d'autres situations. Ces effets sont donc des artefacts, c'est-à-dire des purs produits générés par la situation d'enquête. Chaque méthodologie employée par les sciences sociales, qu'elles soit qualitative ou quantitative, produit des artefacts spécifiques. L'objectivation – ou l'absence d'objectivation – de ces biais méthodologiques a donné lieu à des nombreux travaux et controverses académiques.

L'entretien est un exemple d'une situation sociale littéralement extraordinaire, car dissemblable, en raison de son formalisme, des interactions sociales ordinaires. Pour Pierre Bourdieu, la situation d'enquête repose sur une double dissymétrie. La première concerne l'établissement des règles du jeu : « *c'est l'enquêteur qui engage le jeu et institue la règle du jeu; c'est lui qui, le plus souvent, assigne à l'entretien, de manière unilatérale et sans négociation préalable, des objectifs et des usages parfois mal déterminés, au moins pour l'enquêté* »⁹¹. Elle se double de plus d'une dissymétrie sociale, lorsque l'enquêteur occupe une position sociale plus élevée que l'enquêté, ou, inversement, lorsque l'enquêteur est dominé⁹². La possession d'un plus grand capital culturel – pouvant se manifester durant l'entretien par un niveau de langue plus élevé de l'enquêteur – peut conduire à une situation de violence symbolique, bridant alors la parole des enquêtés.

Les méthodes quantitatives produisent, elles aussi, des artefacts. Les biais déclaratifs générés par les sondages d'opinion ont donné lieu à une longue controverse scientifique, initiée en France dès les années 1970 par Madeleine Grawitz⁹³ et Pierre Bourdieu⁹⁴. Un des principaux biais des sondages⁹⁵ réside selon ces auteurs dans l'imposition de problématique, à savoir le fait de « *mettre en demeure [les enquêtés] de répondre à des questions qu'ils ne se sont pas posées* »⁹⁶. Cette critique émane de la remise en cause d'un des postulats sur lequel repose les sondages d'opinion : « *tout le monde n'interprète pas les questions de la même façon* »⁹⁷ et, plus encore, tout le monde n'est pas en mesure d'avoir un avis sur tous les

⁹¹ Pierre Bourdieu, *La misère du monde*, Paris, Seuil, 1993.

⁹² Monique Pinçon-Charlot, Michel Pinçon, *Voyage en grande bourgeoisie*, Paris, PUF, 1997.

⁹³ Madeleine Grawitz, « Les sondages d'opinion publique », *Revue des travaux de l'Académie des sciences morales et politiques*, 125^{ème} année, 1^{er} semestre 1972, p. 99-108.

⁹⁴ Pierre Bourdieu, « L'opinion publique n'existe pas », *Les temps modernes*, n°318, 1972, p. 1292-1309.

⁹⁵ Si ces critiques sont principalement adressées aux sondages d'opinion commerciaux, la proximité voire la confusion structurelle des instituts de sondage et de certains Instituts d'Étude politique, confère, plus ou moins implicitement, aux sondages universitaires, du moins tels qu'ils sont pratiqués par la science politique, les mêmes biais. Cf. Pierre Bourdieu, « Les doxosophes », *Minuit*, 1972.

⁹⁶ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

⁹⁷ Madeleine Grawitz, « Les sondages d'opinion publique », art. cit.

sujets⁹⁸. De nombreux travaux américains⁹⁹ et français¹⁰⁰ ont de longue date démontré empiriquement l'inégale distribution sociale de la compétence politique, entendue comme la capacité à maîtriser et manipuler les schèmes de perception, d'évaluation et de classification du champ politique.

Seule la pratique très controversée, et relativement rare, de l'observation ethnographique « clandestine » – ou « *incognito* » – est à même de produire des matériaux empiriques (dont les enregistrements dissimulés sont l'archétype) non impactés par l'intervention ou l'observation du chercheur, les enquêtés ignorant par définition sa présence. Pour se rendre « invisible », le chercheur peut dissimuler son identité en se présentant à l'enquêté sans mentionner sa profession et ses intentions ou, plus rarement, en ne développant aucune interaction avec l'enquêté¹⁰¹. Ainsi, afin de ne pas troubler le déroulement des interactions observées, le sociologue américain Laud Humphreys, dans son étude des « pissotières », lieu de rencontre homosexuelle anonyme, a dissimulé son identité de chercheur et a adopté une posture de « voyeur », socialement tolérée. Il n'a ainsi noué aucune interaction directe avec les enquêtés. La dissimulation de l'identité du chercheur et la pratique d'enregistrement clandestin posent néanmoins de lourds problèmes déontologiques et éthiques¹⁰² dont il est d'ailleurs intéressant de noter qu'ils se posent aujourd'hui en des termes relativement identiques dans le débat académique sur la politique de recherche à adopter face aux données récoltées sur internet¹⁰³.

À l'instar des enregistrements clandestins, les autopublications numériques constituent un matériau empirique non stimulé par l'intervention du chercheur. Cependant, alors que le recours à l'observation clandestine est souvent légitimé comme relevant d'un choix par défaut, comme étant le seul dispositif d'enquête à même d'étudier des terrains sensibles voire dangereux, l'étude des autopublications numériques permet d'investiguer des objets variés et ne se réduit pas à l'étude de terrains sensibles. Les autopublications offrent ainsi la possibilité

⁹⁸ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

⁹⁹ Philip Converse, « The Nature of Mass Beliefs in Mass Publics », in David Apter (dir.), *Ideology and Discontent*, New York, The Free Press, 1964, p. 206-261 ; John R. Zaller, *The Nature and Origins of Mass Opinion*, Cambridge, Cambridge University Press, 1992.

¹⁰⁰ Daniel Gaxie, *Le cens caché. Inégalités culturelles et ségrégation politique*, Paris, Édition du Seuil, 1978 ; Pierre Bourdieu, *La Distinction, Critique sociale du jugement*, Paris, Les Éditions de Minuit, 1979.

¹⁰¹ Laud Humphreys, *Le commerce des pissotières. Pratiques homosexuelles anonymes dans l'Amérique des années 1960*, Paris, La découverte, 2007.

¹⁰² Christophe Dargère, *L'observation incognito en sociologie, Notions théoriques, démarche réflexive, approche pratique et exemples concrets*, Paris, L'Harmattan, 2012.

¹⁰³ Pour avoir un aperçu des termes des débats, cf. Guillaume Latzko-Toth et Serge Proulx, « Enjeux éthiques de la recherche sur le Web », in Christine Barats (dir.), *Manuel d'analyse du Web*, Paris, Arman Colin, 2012.

d'appréhender des objets classiques des sciences sociales à partir d'un matériau empirique inédit et « authentique ».

« Authentique » ne signifie pas que les autopublications sont la manifestation pure et inaltérée d'opinions ou de sentiments « privés » jusqu'alors enfouis dans le for intérieur de leurs auteurs. Nous le verrons, comme cela est le cas avec les discussions politiques hors-ligne¹⁰⁴, les auteurs des autopublications adaptent plus ou moins consciemment le contenu de leur message en fonction des autres individus qui prennent part à la discussion et/ou de leurs lecteurs, mais aussi en fonction des contraintes d'énonciation parfois imposées par les espaces d'autopublication (usage ou non de l'anonymat, publication visible par tous ou par un public restreint et choisi, taille limitée ou non du nombre de caractères, etc.)¹⁰⁵. Qualifier de plus « naturelles » ou de plus « authentiques » les autopublications ne signifie donc pas que leur expression n'est pas contrainte par des dispositifs sociotechniques ou par des normes sociales, mais simplement qu'elle n'est pas altérée par le dispositif d'enquête. Sauf à adopter volontairement une stratégie de recherche « ouverte »¹⁰⁶, les auteurs des autopublications ne peuvent se savoir lus et analysés par des chercheurs.

Les autopublications peuvent donc se prévaloir de ne pas être des données générées artificiellement par une situation d'enquête. Cependant, précisément parce que la production de ces données ne se déroule pas en laboratoire, sous la focale du chercheur, les conditions sociales de production de ces messages et l'identité socio-démographique de leurs auteurs peuvent alors partiellement ou totalement lui échapper.

2.1.2 Des données de seconde main décontextualisées et désincarnées

Les autopublications numériques ne sont pas collectées dans le cadre d'un dispositif d'enquête dont le chercheur maîtrise et contrôle les différents paramètres. Elles ne peuvent donc être considérées comme des données de première main. Les autopublications numériques se distinguent néanmoins des données dites de « seconde main ». Les données de

¹⁰⁴ Nina Eliasoph, *L'évitement du politique. Comment les Américains produisent l'apathie dans la vie quotidienne*, Paris, Economica, 2010 ; Vincent Goulet, *Médias et classes populaires. Les usages ordinaires des informations*, Paris, INA Édition, 2010.

¹⁰⁵ Dominique Boullier et Audrey Lohard. *Opinion mining et Sentiment analysis : Méthodes et outils*, op. cit.

¹⁰⁶ Nous entendons par stratégie « ouverte » le fait de se présenter aux enquêtés « à visage découvert » en leur présentant explicitement les enjeux de la recherche.

seconde main sont des données collectées et traitées par un tiers (chercheurs, journalistes, archivistes, etc.), ou une chaîne de tiers, à partir de « sources primaires ». Ainsi les résultats du recensement de la population par l'INSEE, avant de se présenter sous la forme de tableaux statistiques, ont d'abord été le produit d'interactions sociales entre des enquêteurs et des enquêtés à qui a été administré un questionnaire. Les données de seconde main sont collectées et traitées en fonction de problématiques et d'un cadre théorique qui ne sont pas nécessairement ceux du chercheur qui les mobilise « en seconde main ».

Les autopublications numériques se distinguent donc sur ce point des données de seconde main. Les messages publiés sur internet n'ont pas été collectés, traités et analysés par d'autres chercheurs. Ils se présentent sous la forme de verbatims bruts, publiés directement par leurs auteurs, généralement sans médiation ni intermédiaire. Les messages publiés sur internet s'apparentent donc davantage à des « sources primaires ». Mais comme cela peut être le cas avec des données de seconde main, le chercheur ignore souvent les caractéristiques sociologiques de ces sources.

Les autopublications numériques se présentent en effet au chercheur comme des verbatims décontextualisés et désincarnés. Si l'heure et la date de publication des messages sont des informations facilement disponibles, les motivations et l'identité sociale de leurs auteurs sont en revanche des données qui échappent bien souvent à l'analyse. Alors que dans les enquêtes déclaratives les variables socio-démographiques des enquêtés (âge, sexe, PCS, niveau de diplôme, préférence politique, etc.) sont au centre de l'analyse et permettent de situer socialement les opinions, les attitudes et les comportements (déclarés) des enquêtés, sur internet, ces données ne sont pas systématiquement disponibles.

Sur les forums de discussion, l'anonymat et l'usage de pseudonyme est une pratique majoritaire. Si les effets sociaux de l'anonymat sur la prise de parole en ligne ont donné lieu à de nombreux travaux, concluant souvent à un effet désinhibant voire abréactif¹⁰⁷, une autre conséquence de l'anonymat, peut-être évidente mais assez peu soulignée, réside dans la difficulté à identifier et à qualifier les auteurs des autopublications. La nature du pseudonyme choisie n'est toutefois pas neutre socialement et peut être parfois un indice de caractérisation sociologique. Ainsi, Sylvain Paraisie et Jean-Philippe Cointet, dans une étude statistique et lexicométrique de forums mis en place par un quotidien régional à l'occasion des élections municipales de 2008, ont démontré que la nature des pseudonymes adoptés par les

¹⁰⁷ Sherry Turkle, *Life on the Screen. Identity in the Age of Internet*, New York, Simon & Shuster, 1995.

usagers des forums était déterminée par la taille de la commune dans laquelle ils résidaient (le journal avait mis en place un forum propre à chaque commune de la région). Dans les communes de moins de 2 000 habitants les pseudonymes faisaient ainsi majoritairement référence à la localité, par la mention du lieu de résidence (« *un vrai gonnehemois* ») et/ou d'une fonction, d'un engagement associatif (« *un footballeur de St-Souplet* ») ou d'une responsabilité locale (« *le maire* ») tandis que dans les communes de plus de 10 000 habitants les pseudonymes étaient davantage critiques (« *il faut changer* ») et associés à des références nationales (« *socialiste en colère* »)¹⁰⁸. Si les pseudonymes peuvent donc dans certaines configurations se révéler être des indices de caractérisation sociologique, ces données sont parcellaires et leur traitement non-systématisable.

Le recours à l'anonymat est en revanche moins développé sur les réseaux sociaux. Facebook et Google+ ont ainsi imposé à leurs utilisateurs de déclarer leur identité légale. Cependant, cette règle est loin d'être totalement respectée¹⁰⁹, l'usage de pseudonyme est encore répandu. De plus, si les réseaux sociaux invitent – et ont un véritable intérêt économique à le faire – les internautes à publier de nombreuses informations sociodémographiques (sexe, âge, ville de résidence, ville de naissance, niveau d'étude et établissements fréquentés, professions, etc.) mais aussi politiques, culturelles ou religieuses, les usagers des réseaux sociaux restent libres de ne pas dévoiler ces informations.

Même sur les réseaux sociaux, où le recours à l'anonymat est moins répandu, les données socio-démographiques et politiques des auteurs des autopublications ne sont disponibles que de façon parcellaire. Face à ce constat empirique, deux stratégies s'offrent alors aux chercheurs.

2.1.3 Les stratégies de (non) qualification sociodémographique des auteurs des autopublications numériques

La première stratégie consiste à compléter les informations manquantes par une phase de production *ex post* de données de première main. La production de ces données peut être réalisée à partir de deux méthodes : des enquêtes par questionnaire en ligne visant

¹⁰⁸ Sylvain Paraisie et Jean-Philippe Cointet, « La presse en ligne au service de la démocratie locale. Une analyse morphologique de forums politiques », art. cit.

¹⁰⁹ Dominique Boullier et Audrey Lohard, *Opinion mining et Sentiment analysis : Méthodes et outils*, op. cit.

explicitement à caractériser sociologiquement les enquêtés et/ou un profilage des auteurs à partir d'un croisement de données collectées sur internet.

La première méthode consiste à administrer aux enquêtés – les auteurs des autopublications du corpus étudié – un questionnaire en ligne¹¹⁰. Il s'agit alors d'interroger directement les enquêtés sur leurs attributs socio-démographiques. Ce faisant, le principal attrait heuristique des autopublications numériques, à savoir l'absence d'interaction entre enquêteurs et enquêtés, peut être alors partiellement altéré. Le moment choisi pour administrer le questionnaire est alors déterminant. S'il intervient en amont de la constitution du corpus, les autopublications étudiées peuvent être potentiellement impactées par le protocole de recherche. En revanche, cette situation peut être évitée si l'interrogation des enquêtés intervient en aval de la collecte¹¹¹.

Une deuxième méthode consiste à *profiler* les auteurs des autopublications à partir d'un croisement de données recensées sur internet¹¹². Si les informations socio-démographiques déclarées des enquêtés sont insuffisantes ou incomplètes, le chercheur peut alors, à partir des données parcellaires dont il dispose rechercher des informations complémentaires sur d'autres espaces d'autopublication où peuvent être inscrits les enquêtés. En moyenne, un internaute serait membre de 3,5 réseaux sociaux¹¹³. Des individus peuvent communiquer assez peu d'informations sur un réseau social et être plus prolixes sur d'autres. Les réseaux sociaux dits « professionnels » (Linkedin, Viadeo, etc.), sur lesquels le profil des inscrits s'apparente à des *curriculum vitae* en ligne, constituent par exemple des sources d'informations particulièrement riches pour déterminer le niveau d'étude ou la catégorie socioprofessionnelle des enquêtés. Les informations recueillies sur les différents espaces d'autopublications peuvent être ensuite croisées et regroupées, afin de compléter les données socio-démographiques manquantes. À défaut ou en complément de ces données socio-démographiques, des informations relatives aux goûts culturels des enquêtés peuvent constituer des indices sociologiques révélateurs de leur catégorie sociale d'appartenance. Les

¹¹⁰ Concernant l'administration de questionnaires en ligne, cf. notamment Arnaud Mercier, « Étudier les internautes par des questionnaires en ligne », in Amandine Degand et Benoît Grevisse (dir.), *Journalisme en ligne*, Bruxelles, de Boeck, 2012, p. 292-297 ; Dieudonné Tchente et alii, « Pour une approche interdisciplinaire des TIC. Le cas des réseaux sionumériques », *Document numérique* 2011/1 (Vol. 14) ou encore Gersende Blanchard, Simon Gadras et Stéphanie Wojcik, « Analyser la participation politique en ligne : des traces numériques aux pratiques sociales » in Christine Barats (dir.), *Manuel d'analyse du Web*, Paris, Arman Colin, 2012.

¹¹¹ Nous aborderons plus en détail ces questions dans le chapitre 2, lors de la présentation de notre méthodologie de panélisation de la twittosphère politique.

¹¹² Dominique Boullier et Audrey Lohard, *Opinion mining et Sentiment analysis : Méthodes et outils*, op. cit.

¹¹³ IFOP, « Observatoire des Réseaux Sociaux », Novembre 2012.

goûts culturels des individus sont en effet, on le sait, fortement discriminants et socialement déterminés¹¹⁴. Dès lors, par un renversement méthodologique, l'étude inductive des goûts culturels exprimés par les enquêtés peut fournir des informations relatives à leurs attributs sociaux. Si cette méthode présente l'avantage de ne pas nécessiter une interaction directe avec les enquêtés, elle ne permet pas néanmoins un recueil systématisé des informations nécessaires : des individus peuvent ne pas être inscrits sur d'autres réseaux sociaux ou, si ils le sont, ne pas y dévoiler davantage d'informations.

Ces deux méthodes de production de données sociodémographiques complémentaires sur les auteurs des autopublications numériques présentent un coût relativement élevé et de surcroît proportionnel à la taille du corpus étudié. Plus le nombre d'autopublications collectées est important, plus la recherche d'informations complémentaires sur leurs auteurs sera chronophage. Une autre stratégie possible consiste alors, pour reprendre les termes de Dominique Boullier et Audrey Lohard, à « *accepter d'abandonner la référence socio-démographique* »¹¹⁵. Selon ces auteurs, puisque « *ce qui est étudié [les autopublications] ne fait plus référence à des émetteurs bien identifiés dont les chercheurs posséderaient les caractéristiques socio-démographiques classiques et considérées comme définitives de la personne* »¹¹⁶, les sciences sociales n'auraient plus d'autres choix que de « *faire [leur] deuil et pour longtemps des repères socio-démographiques sur le web actuel et pour cela changer de paradigme* »¹¹⁷.

De plus, le développement de logiciel permettant de traiter des quantités très importantes de données numériques a offert aux chercheurs la possibilité de collecter de façon exhaustive – et non de façon échantillonnée comme avec les méthodes quantitatives traditionnelles – les messages publiés sur les espaces d'autopublication. Dès lors, face à l'attrait de corpus exhaustifs de très grande taille, l'opération de qualification « manuelle » (c'est-à-dire non automatisée) des données socio-démographiques peut, pour certains chercheurs, apparaître secondaire.

Nous le verrons dans la prochaine section de ce chapitre, les stratégies et arbitrages de recherche adoptés entre une démarche « manuelle » et potentiellement chronophage de qualification sociologique des autopublications numériques et une analyse automatisée et

¹¹⁴ Pierre Bourdieu, *La Distinction, op. cit.*

¹¹⁵ Dominique Boullier et Audrey Lohard. *Opinion mining et Sentiment analysis : Méthodes et outils, op. cit.*

¹¹⁶ *Ibidem.*

¹¹⁷ *Ibidem.*

exhaustive d'importants volumes de données dépendent des intérêts spécifiques des agents – scientifiques et commerciaux – qui ont investi ces nouveaux terrains d'étude. Néanmoins, une particularité heuristique des autopublications numériques est d'être précisément un matériau adapté à des méthodes combinant approches quantitatives et qualitatives.

2.2. Un matériau adapté à des méthodes combinant approches quantitatives et qualitatives

En raison de leur abondance et de leur facilité d'accès, de collecte et de traitement, les autopublications apparaissent en effet comme un matériau particulièrement adapté à des méthodes combinant approches qualitatives et quantitatives. Des logiciels spécialisés dans la récolte et le traitement de données numériques¹¹⁸ permettent ainsi d'opérer un traitement quantitatif des autopublications. Cette quantification peut prendre la forme d'analyses statistiques et lexicométriques d'importants volumes de verbatims recueillis sur internet. Par ailleurs, les autopublications se trouvent de plus en plus souvent associées à des données quantitatives : nombre de pages vues, de like¹¹⁹, de retweet, etc. Ces données quantitatives peuvent être alors qualifiées par l'analyse des verbatims qui leurs sont associées. L'analyse des autopublications autorise donc un traitement quantitatif de données qualitatives mais aussi une qualification de données quantitatives.

L'opposition entre approches quantitatives et qualitatives renvoie aux dualismes fondateurs des sciences sociales : méthodes nomothétique et compressive, holisme et individualisme, positivisme et constructivisme, objectivisme et subjectivisme, etc. Si ces dualismes sont toujours prégnants¹²⁰, de nombreux travaux cherchent aujourd'hui à les dépasser, prônant une plus grande souplesse intellectuelle et une combinaison des méthodes et des approches. De nombreux auteurs, notamment américains, ont en effet expérimenté des méthodologies combinant approches qualitatives et quantitatives¹²¹ tandis que le développement des statistiques et des mathématiques en sciences sociales a rendu plus poreuse la distinction conceptuelle entre données qualitatives et quantitatives. Les

¹¹⁸ On peut citer les logiciels *Ami Opinion Tracker*, *Digimind*, *Spotter*. Pour une présentation détaillée du logiciel *Ami Opinion Tracker*, logiciel mobilisé pour cette recherche, cf. Annexe 6, p. 487.

¹¹⁹ Cf. glossaire p. 464.

¹²⁰ Daniel Gaxie, « Appréhender le politique à l'aune des expériences sociales », *Revue Française de Science Politique*, vol. 52, n°2-3, avril-juin 2002

¹²¹ Donald Campbell et Donald Fiske, « Convergent and discriminant validation by the multitrait-multimethod matrix », *Psychological Bulletin*, Vol. 56(2), 1959, p. 81-105.

autopublications apparaissent comme un matériau particulièrement adapté aux méthodes combinant approches quantitatives et qualitatives (1).

Le développement de logiciels permettant de collecter, d'archiver et d'opérer un traitement quantitatif des autopublications numériques a généré la création de nombreux instruments de mesure des autopublications numériques. Certains sont purement quantitatifs – indicateurs de volumétrie, d'audience et d'« influence » – tandis que d'autres sont davantage qualitatifs ou plus exactement quanti-qualitatifs : indicateurs lexicométriques de co-occurrences de termes, indicateurs de tonalité, etc. Reposant sur une analyse automatique du langage, ce second type d'indicateur pose néanmoins de véritables problèmes épistémologiques et heuristiques (2).

2.2.1 Les autopublications au prisme du dualisme quantitatif/qualitatif

L'opposition entre méthodes qualitatives et quantitatives constitue en effet l'un des dualismes fondateurs des sciences sociales, étudié dans tous les manuels de méthodologie¹²². Cette distinction est néanmoins équivoque, car elle repose en réalité sur différents critères dont aucun ne permet une distinction absolue entre les deux approches¹²³. On peut retenir ici deux principaux critères distinctifs : l'orientation de la recherche et la nature des données¹²⁴. Pour chacun d'eux, les autopublications confirment et renforcent la porosité entre approche quantitative et qualitative et *in fine* l'adaptabilité de ce matériau à des méthodes mixtes.

Un premier critère de distinction entre méthodes quantitatives et qualitatives généralement mis en avant concerne l'orientation de la recherche. On a ainsi tendance à associer l'approche qualitative à une phase d'exploration et l'approche quantitative à une phase de vérification¹²⁵. Silverman distingue ainsi schématiquement deux écoles en sciences sociales : l'une orientée vers la génération qualitative d'hypothèses et l'autre orientée vers le test quantitatif d'hypothèses¹²⁶. Pour Glaser et Strauss, cette distinction est factice car il n'y « *pas de conflit fondamental entre les buts et les potentialités des méthodes [...] qualitatives et quantitatives. [...]. Chacune des formes de données est utile pour la vérification et la*

¹²² Comme par exemple le manuel de Madeleine Grawitz, *Méthodes des sciences sociales*, Paris, Dalloz, 2000.

¹²³ Philippe Baumard, « Quelles approches avec quelles données ? », art. cit.

¹²⁴ *Ibidem*.

¹²⁵ *Ibidem*.

¹²⁶ David Silverman, *Interpreting Qualitative Data*, London, Sage, 1993.

génération de théorie»¹²⁷. Ainsi, les progrès des statistiques et de l'informatique et plus récemment le développement des logiciels de veille et de *tracking* ont accru les potentialités de l'approche quantitative comme démarche exploratoire¹²⁸. Ces logiciels permettent en effet de collecter de façon automatique et systématique des données (articles de presse, messages publiés sur les forums, posts de blogs, tweets, statuts Facebook, etc.) issues de sites internet préalablement définis, puis de les archiver et de suivre leur évolution en temps réels. *Ami Opinion Tracker*, logiciel mobilisé dans le cadre de cette recherche, nous a ainsi permis de collecter de façon exhaustive et automatique, sur une durée de deux ans, l'intégralité des posts de blogs et des tweets politiques (c'est-à-dire contenant des termes relatifs aux noms propres d'hommes politiques, de partis, d'institutions, etc.)¹²⁹. Le logiciel peut ainsi générer des analyses statistiques et lexicométriques sur les millions d'autopublications collectées. Ces analyses quantitatives très générales sur le terrain étudié (distribution du nombre de tweets dans le temps, comptes les plus retweetés, expressions les plus fréquemment employées, etc.) permettent, selon une démarche purement inductive, d'alimenter la phase exploratoire d'élaboration d'hypothèses. Internet a donc accru le recours à des approches quantitatives comme démarches exploratoires.

De manière symétrique, une approche qualitative peut également parvenir, en multipliant les études de cas selon une logique de réplication, à une plus grande montée en généralité théorique, à accroître la validité externe de l'analyse¹³⁰. La nature numérique et donc facilement accessible – à la fois géographiquement et temporellement – des espaces d'autopublication facilite la multiplication des études de cas en baissant le coût d'accès aux terrains. Sur internet, les données générées par les enquêtés sont en effet accessibles indépendamment du lieu géographique où se trouve le chercheur¹³¹. Internet facilite donc également le recours à une approche qualitative comme démarche vérificatoire.

En favorisant le recours à des approches quantitatives exploratoires et en facilitant la multiplication des études de cas comme démarche de vérification, les autopublications numériques rendent donc plus poreuse la frontière conceptuelle entre approche quantitative et qualitative.

¹²⁷ Barney Glaser et Anselm Strauss, *The Discovery of Grounded Theory*, New York, Aldine Publishing Company, 1967, cité dans David Silverman, *Interpreting Qualitative Data*, London, Sage, 1993.

¹²⁸ Doris Graber, *Processing the News: How People Tame the Information Tide*, University Press of America, 1988.

¹²⁹ Pour consulter la liste des 121 mots-clés retenus, Cf. Annexe 7, p. 495.

¹³⁰ Philippe Baumard, « Quelles approches avec quelles données ? », art. cit.

¹³¹ Guillaume Latzko-Toth et Serge Proulx, « Enjeux éthiques de la recherche sur le Web », in Christine Barats (dir.), *Manuel d'analyse du Web*, Paris, Arman Colin, 2012.

La nature des données est un second critère de distinction mis en avant pour opposer approches qualitatives et quantitatives. Si selon Miles et Huberman « *les données qualitatives [...] se présentent sous formes de mots plutôt que de chiffres* »¹³², l'introduction du traitement statistique en sciences sociales a rendu plus poreuse cette distinction, en permettant de convertir des données qualitatives en variables discrètes¹³³ classables sur des échelles nominales ou ordinales. Ces variables dites « qualitatives » autorisent ainsi des calculs d'appartenance, de fréquence ou de classement. Les autopublications, données textuelles et donc qualitatives, peuvent être ainsi converties, après une opération de codage, c'est-à-dire d'attribution à une classe d'équivalence, en variables qualitatives nominales (comme, par exemple, le classement d'autopublications en fonction de leur thématique) ou ordinales (un classement de tonalité selon une échelle de Likert par exemple). Mais les autopublications numériques peuvent être également transformées en données purement quantitatives. Un *verbatim* peut ainsi être converti en une donnée chiffrée, métrique, comme par exemple un nombre de caractères, ou bien un nombre d'occurrences (nombre de l'occurrence « contre » ou « gauche » dans un même message par exemple). Ces données chiffrées peuvent être ensuite converties en variables « quantitatives », classables sur des échelles métriques d'intervalles ou de ratios, autorisant ainsi un traitement statistique plus sophistiqué et une analyse lexicométrique des autopublications.

Les données chiffrées parfois associées aux autopublications (nombre de pages vues ou de visiteurs uniques, nombre de retweet, de like, etc.) peuvent être elles aussi traitées quantitativement, mais aussi qualitativement, en les croisant avec des variables qualitatives. Une variable quantitative « nombre de retweet d'un message » peut ainsi être croisée avec une variable qualitative (messages favorables ou défavorables au sujet du Premier ministre par exemple) pour ainsi qualifier la tonalité des messages les plus repris au sujet du Premier ministre. De plus, les logiciels de collecte et de traitement des autopublications proposent une interface dynamique et interactive autorisant un aller/retour formel permanent entre des données qualitatives et leur conversation en données quantitatives, et inversement. En un clic, un pourcentage de messages négatifs peut ainsi laisser place au corpus de verbatims correspondant. Données qualitatives aisément quantifiables, les autopublications rendent donc encore plus perméable la délimitation entre données qualitatives et quantitatives.

¹³² Matthew B. Miles et A. Michael Huberman, *Analyse des données qualitatives : Recueil de nouvelles méthodes*, Bruxelles, De Boeck, 2003.

¹³³ Alain Desrosières, « Entre réalisme métrologique et conventions d'équivalence : les ambiguïtés de la sociologie quantitative », *Genèses*, 2001/2 (n°43), p. 112-127.

Les autopublications apparaissent donc comme un matériau particulièrement adapté à des méthodes combinant approches qualitatives et quantitatives. Campbell et Fiske ont théorisé l'usage simultané de méthodes qualitatives et quantitatives par le biais de la triangulation¹³⁴. « *L'achèvement de construits utiles et hypothétiquement réalistes dans une science passe par l'utilisation de méthodes multiples focalisées sur le diagnostic d'un même construit à partir de points d'observation indépendants, à travers une sorte de triangulation*¹³⁵ ». À partir d'une même problématique, le chercheur aura alors recours à des méthodologies différentes, dont « *le jeu différentiel sera source d'apprentissages pour le chercheur, [...] permettant un dialogue différencié entre ce qui est observé (l'objet de la recherche) et les deux façons de le symboliser [les données qualitatives et quantitatives]*¹³⁶ ». Étant à la fois un matériau qualitatif et quantitatif, les autopublications favorisent et renforcent cette logique de triangulation. Elles peuvent ainsi symboliser de deux manières (sous la forme de verbatims ou sous la forme de données statistiques) un même phénomène. Le double statut des autopublications favorisent ainsi un renforcement du dialogue entre approches qualitatives et quantitatives.

Tant du point de vue de l'orientation de la recherche que de la nature des données, les autopublications confirment donc la porosité des approches qualitatives et quantitatives et *in fine* leur adaptabilité à des méthodes mixtes. L'étude quanti-qualitative ou quali-quantitative des autopublications peut se réaliser à partir de différents indicateurs et instruments de mesure.

2.2.2 Les instruments de mesure des autopublications numériques

Les logiciels de collecte, de stockage et de traitement des données numériques proposent de nombreux indicateurs et instruments de mesure des autopublications. Nous le verrons dans la prochaine section, ces logiciels n'ont pas été conçus pour les sciences sociales mais pour une série très diversifiée d'agents économiques qui ont pour point commun de commercialiser des services centrés sur l'analyse d'autopublications (suivi de *e*-réputation,

¹³⁴ Donald T. Campbell et Donald W. Fiske, « Convergent and discriminant validation by the multitrait-multimethod matrix », *Psychological Bulletin*, Vol. 56(2), 1959, p. 81-105.

¹³⁵ *Ibidem*.

¹³⁶ Philippe Baumard, « Quelles approches avec quelles données ? », art. cit.

études de marchés, veille stratégique, etc.)¹³⁷. Les indicateurs et instruments de mesure proposés par ces logiciels, même s'ils n'ont pas été pensés à partir de problématiques de recherche, proposent néanmoins des données quantitatives intéressantes pour les sciences sociales. On peut distinguer des indicateurs purement quantitatifs – indicateurs de volumétrie, d'audience et d'« influence » – et des indicateurs « quanti-qualitatifs » – indicateurs de tonalité et calculs de co-occurrences.

Les indicateurs de volumétrie mesurent la production d'autopublication pour un temps *t*. Les logiciels de veille et de *tracking* permettent en effet de recenser et comptabiliser en temps réel et de façon exhaustive les publications des internautes sur les réseaux sociaux, les forums de discussion ou les plateformes de blog. Ces indicateurs représentent sous forme numérique ou graphique la distribution du nombre de publications dans le temps, avec comme unité la seconde, la minute, l'heure, le mois ou l'année. L'analyse volumétrique peut être également filtrée par mot-clé. Ne seront alors comptabilisées que les autopublications contenant un mot-clé ou une combinaison de mots-clés préalablement définie. Dans le cas d'une analyse par mots-clés politiques, ces données volumétriques peuvent être des indices sociologiques révélateurs de certains phénomènes sociaux et politiques, comme le niveau général de politisation des discussions politiques ordinaires¹³⁸ ou l'intensité d'une campagne électorale¹³⁹. La production d'autopublications numériques peut être également appréhendée comme la manifestation empirique d'autres pratiques médiatiques. De nombreux internautes commentent ainsi simultanément sur internet – et plus particulièrement sur le réseau social Twitter – des émissions qu'ils regardent en direct à la télévision¹⁴⁰. Dès lors, la quantification des tweets publiés durant une émission télévisée peut être un indice de l'intensité de l'attention médiatique des téléspectateurs/commentateurs. Ainsi, une étude commerciale de TNS Sofres analysait les réactions des internautes sur Twitter durant une émission politique où était invité Nicolas Sarkozy et mettait en relation le nombre de tweets publiés minute par minute et les propos tenus au même moment par le Président¹⁴¹. L'unité de mesure de ces indicateurs volumétriques peut être l'autopublication ou l'individu, c'est-à-dire l'auteur de l'autopublication. Lorsque ces deux informations sont disponibles, des ratios peuvent être alors calculés. Ces ratios constituent alors des indicateurs de l'intensité de production des

¹³⁷ Dominique Boullier et Audrey Lohard. *Opinion mining et Sentiment analysis : Méthodes et outils*, op. cit.

¹³⁸ Vincent Goulet, *Médias et classes populaires*, op. cit.

¹³⁹ Céline Braconnier et Jean-Yves Dormagen, *La démocratie de l'abstention. Aux origines de la démobilisation en milieu populaire*, Paris, Gallimard, 2007.

¹⁴⁰ Valérie Jeanne-Perrier, « Parler de la télévision sur Twitter : une réception oblique à partir d'une conversation médiatique », *Communication & langages*, 2010, p. 127-147.

¹⁴¹ <http://www.tns-sofres.com/etudes-et-points-de-vue/intervention-televisee-de-nicolas-sarkozy-analyse-twitter>

internauts, elle-même déterminée par certaines propriétés sociologiques. Un ratio élevé de messages politiques publiés par individu est ainsi souvent corrélé à un investissement militant, ou du moins à un très fort intérêt pour la politique¹⁴². Entièrement automatisés, les indicateurs de volumétrie offrent donc des informations sociologiques intéressantes et exploitables pour la recherche.

Contrairement aux indicateurs volumétriques, les mesures d'audience et d'« influence » sont le plus souvent réalisées à partir de données métriques publiées directement par les espaces d'autopublication ou par des moteurs de recherche. Chaque espace d'autopublication fournit des indicateurs métriques qui leurs sont propres. Ainsi, les plateformes de blogs rendent généralement public le nombre de pages vues (qui correspond au nombre absolu de pages consultées sur un blog), le nombre de visiteurs uniques et le *PageRank*. Le *PageRank* est un algorithme créé par le moteur de recherche Google dont la finalité est de mesurer la popularité d'une page web. Directement inspiré de la sociométrie et du *Science Citation Index*¹⁴³, le *PageRank* conçoit le lien hypertexte¹⁴⁴ comme un acte de reconnaissance et un signe d'autorité, de manière analogue aux références bibliographiques publiées dans un article scientifique. Le *PageRank* créé dès lors « *une rupture décisive en faisant de la qualité des informations présentes sur les sites non pas une propriété interne qui doit être recherchée à l'intérieur du document par une analyse toujours plus fine de son contenu lexical, mais une propriété externe qui est façonnée par les attributions respectives que les sites se font les uns envers les autres en se reconnaissant*¹⁴⁵ ». L'algorithme du *PageRank* n'attribue pas à tous les liens hypertextes le même poids : il les pondère en fonction d'un indice d'autorité. Ainsi, schématiquement, un article de blog cité une fois par le site internet du *Monde* aura un *PageRank* plus élevé que s'il avait été cité par deux blogs eux-mêmes très peu référencés.

Sur le réseau social Twitter, quatre données métriques sont systématiquement disponibles. Trois d'entre elles sont relatives aux comptes des utilisateurs, il s'agit du nombre total de tweets publiés, du nombre d'abonnements et du nombre d'abonnés. Sur Twitter, les utilisateurs ont la possibilité de s'abonner à d'autres utilisateurs, afin de voir apparaître sur

¹⁴² Nous proposerons une étude empirique de ce phénomène dans notre chapitre 3.

¹⁴³ Dominique Cardon, « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », *Réseaux*, 2013/1 (n° 177), p. 63-95.

¹⁴⁴ Cf. Glossaire p. 464.

¹⁴⁵ Dominique Cardon, « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », art. cit.

leurs *timeline*¹⁴⁶ les tweets des utilisateurs auxquels ils se sont abonnés. Chaque utilisateur a donc un nombre délimité d'abonnements (comptes auxquels il s'est abonné) et d'abonnés (utilisateurs qui se sont abonnés à son compte). Le nombre de retweet est quant à lui une donnée associée au tweet (et non au compte). Par un simple clic, un utilisateur peut retweeter un message, c'est-à-dire publier à l'identique ce message sur son propre compte. Le nombre de retweets d'un message correspond au nombre total de reprises d'un même tweet par différents utilisateurs. À partir de ces quatre données, toute une série d'indicateurs synthétiques et de ratios peuvent être calculés : nombre moyen de retweet par utilisateur, nombre moyen de retweets pondéré par le nombre d'abonnés, ratio abonnés/abonnements, etc. Tout comme le *PageRank*, ces données et ces indicateurs synthétiques peuvent être qualifiés de mesure d'« influence », de « réputation », de « popularité », d'« autorité » ou de « viralité ». L'usage parfois indifférencié de ces notions et leur plus ou moins grande conceptualisation est corrélée, nous le verrons dans la prochaine section, aux intérêts scientifiques et commerciaux des agents qui les mobilisent.

Les indicateurs purement quantitatifs de volumétrie, d'audience et d'« influence » se distinguent d'une autre catégorie d'indicateurs davantage qualitatifs : les calculs de co-occurrences de termes et les indicateurs de tonalité. Les calculs d'occurrences et de co-occurrences proposés par les logiciels de collecte et de traitement des autopublications numériques sont directement importés des analyses et logiciels de lexicométrie¹⁴⁷. La numérisation à grande échelle des textes sur internet et le développement d'outils de traitement de données textuelles toujours plus puissants ont donné une nouvelle actualité aux analyses lexicométriques¹⁴⁸. À partir d'un corpus défini préalablement, ces logiciels permettent des calculs automatisés de fréquence d'occurrences et de co-occurrences de termes, mesurées à partir de coefficients statistiques de significativité. Fréquemment mobilisées pour analyser les discours du personnel politique¹⁴⁹, les analyses lexicométriques sont en revanche moins souvent appliquées à des verbatims d'individus non-professionnels de la politique, politisés ou non. Une caractérisation sociologique fine des auteurs des autopublications peut ainsi affiner et enrichir l'analyse lexicométrique de leurs publications,

¹⁴⁶ Cf. Glossaire p. 464.

¹⁴⁷ Christine Barats *et alli*, « Approches textométriques du web : corpus et outils », in Christine Barats (dir.), *Manuel d'analyse du Web*, *op. cit.*

¹⁴⁸ Damon Mayaffre, « De la lexicométrie à la logométrie » :

http://hal.archives-ouvertes.fr/docs/00/55/19/21/PDF/18_Mayaffre_Astrolabe_2005.pdf.

¹⁴⁹ Damon Mayaffre, « Débat de l'entre-deux-tours 2007. La conquête de l'espace lexical Mayaffre Damon, 2007, « Vocabulaire et discours électoral de Sarkozy. Entre modernité et pétainisme », *La Pensée*, n° 352, 2007, p. 65-80.

en mettant en relations des variables textuelles telles que le niveau de technicité politique des termes employés et des variables individuelles, telles que le niveau de politisation ou le statut social de leurs auteurs.

Les instruments de tonalité sont, comme les analyses lexicométriques, issus du traitement automatique du langage (TAL) et s'inscrivent dans la sous-discipline académique du *sentiment analysis*¹⁵⁰. Ces indicateurs, dont la pertinence et la fiabilité sont critiquées tant du point de vue académique que commercial¹⁵¹, proposent d'attribuer aux autopublications ou à des fragments d'autopublications une tonalité, généralement négative, neutre ou positive. Cette tonalité peut être attribuée à un objet (dénommé « cible » ou « entité », par exemple le téléphone « iPhone 4 » ou François Hollande) ou à des caractéristiques de cet objet (« *feature* » ou « attribut »). Cependant, l'analyse automatique de tonalité connaît de nombreuses lacunes techniques et méthodologiques. Les plus graves concernent leur incapacité à repérer de nombreuses figures rhétoriques comme l'humour, le sarcasme, les métaphores, etc.¹⁵².

Dès lors, face aux difficultés de la lexicométrie et du codage automatique de tonalité à mesurer scientifiquement des données qualitatives, une autre option consiste pour le chercheur à opérer une codification manuelle des autopublications. Si le recours au codage « humain » repose sur l'idée selon laquelle les machines n'interpréteront jamais aussi bien le langage humain que les hommes, le codage « manuel » n'est néanmoins pas une opération neutre, univoque et objective. Différentes recherches¹⁵³, corroborées par nos propres observations, ont constaté que différents annotateurs n'attribuent pas tous la même polarité à un même message. Qu'ils soient chercheurs ou simples enquêteurs, les annotateurs interprètent les messages selon des schèmes de pensée et des perceptions politiques sélectives qui leurs sont propres. S'il est généralement admis, avec les enquêtes par questionnaire, que les enquêtés n'interprètent pas tous les questions de la même manière¹⁵⁴, il est en revanche plus rare de dresser le même constat pour l'interprétation de leurs réponses par les enquêteurs. Dans le cas des enquêtes par sondage, le travail d'interprétation codificatoire est précisément assuré en amont par les enquêtés eux-mêmes, lorsqu'ils cochent des modalités de réponse

¹⁵⁰ Bo Pang et Lillian Lee, « Opinion Mining and Sentiment Analysis », *Foundations and Trends in Information Retrieval*, vol. 2, n°1-2, p. 1-135.

¹⁵¹ Dominique Boullier et Audrey Lohard. *Opinion mining et Sentiment analysis : Méthodes et outils*, op. cit.

¹⁵² *Ibidem*.

¹⁵³ Theresa Wilson, Janyce Wiebe et Paul Hoffman, « Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis », *Proceedings of Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing*, (HLT/EMNLP 2005).

¹⁵⁴ Madeleine Grawitz, « Les sondages d'opinion publique », art. cit.

prédéfinies. Dans le cas des recherches qualitatives, le travail d'interprétation des entretiens ou des observations ethnographiques est très souvent assuré par un seul chercheur¹⁵⁵, qui n'a donc pas à confronter ses analyses aux interprétations d'autres chercheurs. Nos propres expériences d'analyse collective de données qualitatives ont alors mises au jour des interprétations divergentes, contradictoires voire conflictuelles¹⁵⁶. Le codage manuel des autopublications offre néanmoins des gages de scientificité plus probants que le codage automatique. Les différences interprétatives peuvent être en effet objectivées et analysées, le chercheur, contrairement à l'ordinateur, pouvant expliciter et légitimer ses choix.

Si la codification – manuelle ou automatique – de données qualitatives entraîne nécessairement une réduction heuristique du matériau étudié, les potentialités offertes par les logiciels de collecte et de traitement des données numériques permettent d'opérer une dialectique continue entre la représentation qualitative des autopublications, matérialisée sous forme de verbatims – utiles à la compréhension et à la signification des propos des enquêtés – et leur représentation quantitative, sous forme de données statistiques – nécessaires à l'explication et à la quantification de phénomènes sociaux. De plus, la codification des autopublications numériques permet de représenter graphiquement les dynamiques temporelles et spatiales de circulation des autopublications.

2.3 L'objectivation et la modélisation de la circulation temporelle et spatiale des autopublications

Même si elles leur apportent de nouvelles perspectives, les autopublications n'ont pas précédé l'apparition des méthodes combinant approches quantitatives et qualitatives. En revanche, les possibilités d'objectivation et de modélisation des dynamiques de circulation des autopublications dans le temps et dans l'espace constituent un apport heuristique quant à lui totalement inédit pour les sciences sociales.

L'automatisation et la systématisation de l'opération de collecte ainsi que le caractère abondant et continu de l'activité de publication des internautes permettent un suivi en temps réel des autopublications numériques. L'étude de ce matériau introduit dès lors un nouveau

¹⁵⁵ Olivier Rozenberg et François Backman, « Mortels transferts ? La pratique des entretiens collectifs par les Grands Instituts de Sondages Internationaux et par la recherche en Science Sociale », *Focus groupes, entretiens collectifs : état des lieux - Journée d'étude organisée dans le cadre du groupe MOD de l'AFSP* - 8 avril 2009.

¹⁵⁶ Expériences réalisées dans le cadre du projet ANR Imagiweb.

rapport à la temporalité pour les sciences sociales. Le chercheur peut désormais étudier la réaction des individus en temps réel à toute sorte de *stimuli*. Mais les possibilités d'archivage offertes par les logiciels de veille et de *tracking* permettent également d'étudier la réaction des individus à des faits passés au moment où ces faits se sont déroulés et non pas, comme avec les enquêtes déclaratives, en faisant appel à leur mémoire. La collecte et l'archivage en temps réel des autopublications autorisent alors une étude dynamique des « mouvements » et des « flux » d'opinion, là où les sondages étaient cantonnés à des analyses statiques, « photographiques », pour reprendre la métaphore utilisée à l'envi par les sondeurs pour symboliser leur instrument (1).

Le développement des outils de cartographie du web et de modélisation des réseaux¹⁵⁷ permettent de représenter graphiquement les dynamiques temporelles de circulation des autopublications en y intégrant de plus une dimension spatiale. En effet, le web peut être représenté comme un espace composé de documents possédant une topologie qu'il est possible de visualiser¹⁵⁸. En plus de représenter graphiquement la dynamique de diffusion d'un message dans le temps, les graphs du web permettent également de visualiser sa circulation entre différents individus ou entre différents groupes d'individus, aux caractéristiques politiques et sociales plus ou moins homogènes (2).

2.3.1 Un nouveau rapport à la temporalité

En sciences sociales, la temporalité de l'objet précède généralement celle de l'enquête. À l'exception des recherches ethnographiques qui peuvent, par le recours à l'observation, synchroniser la phase de production de données avec la temporalité des objets étudiés, les enquêtes en sciences sociales n'ont en général d'autre choix que d'interroger des individus sur des faits passés. Ainsi, dans le cas des enquêtes électorales « sorties des urnes », même lorsque la phase de production de données (l'administration du *QSU* à la sortie du bureau de vote) suit de quelques minutes seulement le fait social étudié (le passage dans l'isoloir et l'acte de vote), la question « pour quel candidat avez-vous voté ? » renvoie déjà à un acte passé, et l'enquêté devra alors faire appel à sa mémoire immédiate pour y répondre.

¹⁵⁷ Duncan J. Watts, « The "new" science of networks », *Annual Review of Sociology*, 2004, 30:243–70.

¹⁵⁸ Jean-Christophe Plantin, « D'une carte à l'autre : le potentiel heuristique de la comparaison entre graphe du web et carte géographique », in Christine Barats (dir.), *Manuel d'analyse du Web*, op. cit.

Sur Twitter, une large majorité de messages a trait à des faits ou à des informations récents, précédant de quelques heures voire de quelques minutes la publication des tweets. Le réseau social est en effet très sensible à l'actualité. Le déroulement d'un événement, les réactions des internautes à cet événement, la matérialisation écrite de ces réactions et leur collecte par un logiciel peuvent alors se produire de manière quasi-simultanée et en temps réel. Avec les fonctionnalités de lexicométrie et de codage automatique de la tonalité, les phases de production et de traitement des données peuvent même être conduites quasi-simultanément. Cette simultanéité peut présenter de réels intérêts pour les études commerciales de veille stratégique. Le service d'information du gouvernement (SIG), dont la principale mission consiste à « analyser l'évolution de l'opinion publique », s'est ainsi doté d'un logiciel et d'analystes dédiés à la surveillance du réseau Twitter. Un des principaux objectifs de cette veille en temps réel réside dans la détection rapide de « signaux faibles », à savoir des informations ou des rumeurs – fondées ou non – potentiellement sensibles ou polémiques, afin d'adapter au plus vite, le cas échéant, un plan de communication pour les contrecarrer¹⁵⁹. Si la synchronisation des opérations de collecte et de traitement des données présente un réel intérêt pour les études commerciales, cet intérêt est en revanche nul voire négatif pour les sciences sociales. Un des principes constitutifs de la recherche est en effet de ne pas faire du temps un obstacle à la vérification scientifique des faits. La phase de production de données et leur phase de traitement ne peuvent être qu'asynchrones.

En revanche, l'automatisation et la systématisation des opérations de collecte et d'archivage des autopublications permettent aux chercheurs d'étudier les réactions, les discours et les représentations des internautes sur des faits ou des événements passés au moment même où se sont déroulés ces faits, ou du moins au moment où les internautes ont produit un discours sur ces faits ; et non, comme c'est le cas avec les enquêtes par questionnaire, en faisant appel à leur mémoire de ces faits *a posteriori*, plusieurs jours, plusieurs mois voire plusieurs années après.

L'usage des propos rétrospectifs recueillis en entretien pose de nombreux problèmes épistémologiques¹⁶⁰. Il est en effet admis que les propos tenus par des enquêtés sur des faits ou des événements passés sont généralement peu fiables. D'abord parce que les individus peuvent mentir, pour diverses raisons, sur leur passé. En sociologie électorale, l'omission

¹⁵⁹ Jeanne Dussueil, « Comment le gouvernement écoute ce qui se dit sur Twitter », *Challenges*, 23 novembre 2012.

¹⁶⁰ Gilles Pinson, « Peut-on vraiment se passer de l'entretien en sociologie de l'action publique ? », *Revue Française de Science politique*, 2007/5 (Vol. 57), p. 555-597.

d'un vote pour un candidat ou un parti d'extrême-droite est un cas récurrent. Mais les enquêtés peuvent également sincèrement ne pas se souvenir de faits ou d'événements passés, même s'ils les ont vécus et même si ces faits se sont déroulés dans un passé proche voire immédiat. Ainsi retracer des parcours de vote *a posteriori* peut se relever être une opération extrêmement périlleuse, tant la mémoire des enquêtés peut s'avérer défaillante. En plus des problèmes de mensonges et d'oubli, la reconstruction du passé et la re-hiérarchisation des faits *a posteriori* peuvent également entraver la fiabilité des propos tenus sur des événements passés. « *L'interviewé peut mettre en avant des événements et des faits qui lui paraissent faire davantage sens en fonction de la perception qu'il a rétrospectivement du cours des événements. Perception qui est bien souvent largement influencée par les discours historiques produits par les groupes sociaux dans lesquels l'individu est impliqué* »¹⁶¹. Philippe Hamman a ainsi montré, à partir de trois vagues d'entretien avec des anciens ouvriers d'une usine de faïencerie comment les discours étaient à chaque fois influencés par la conjoncture dans laquelle ils étaient produits¹⁶². La mémoire humaine repose sur des mécanismes très complexes qui empêchent l'enquêté de restituer ses souvenirs à la manière d'un magnétophone. Pour reprendre les termes de Ian Hacking « *la mémoire [humaine] n'est pas comme un enregistrement vidéo* »¹⁶³.

Or, avec les autopublications et leurs logiciels de collecte et d'archivage automatique, le chercheur peut désormais substituer la mémoire informatique à la mémoire humaine. Le chercheur peut ainsi étudier des propos publiés par des internautes sur des faits passés au moment où ces propos ont été produits et ainsi éviter les biais des entretiens rétrospectifs. Si, dans une situation d'entretien, des enquêtés ne se souviennent pas des tweets ou des messages qu'ils ont publiés dans le passé, le chercheur en a néanmoins connaissance et peut, s'il le souhaite, les rappeler aux enquêtés. La possession de ces informations par l'enquêteur peut alors donner lieu à diverses stratégies d'enquête voire à des innovations méthodologiques.

L'archivage systématique et exhaustif des autopublications permet également au chercheur de mener autant d'allers/retours qu'il le souhaite entre ses questionnements, ses hypothèses, son terrain, ses données et leur traitement. Avec les enquêtes par questionnaire, toutes les problématiques et hypothèses testées doivent être formalisées avant la passation du

¹⁶¹ *Ibid.*

¹⁶² Philippe Hamman, « Quand le souvenir fait le lien... De la délimitation des domaines de validité des énoncés recueillis par le sociologue en situation d'entretien », *Sociologie du travail*, 44 (2), 2002, p. 175-191.

¹⁶³ Ian Hacking, *Rewriting the Soul: Multiple Personality and Sciences of Memory*, Princeton University Press, 1995, p. 250.

questionnaire. Une fois le questionnaire administré, il sera alors plus compliqué pour le chercheur de tester des hypothèses nouvelles ou complémentaires auxquelles il n'avait pas pensées. Étant produites par les enquêtés en amont de la phase d'élaboration des hypothèses, les autopublications permettent en revanche une plus grande souplesse et une véritable dialectique entre les différentes étapes la recherche. À partir d'un premier traitement des données, le chercheur pourra en déduire de nouvelles hypothèses et les tester sur les mêmes corpus, ou sur des nouvelles données.

La collecte et l'archivage en temps réel des autopublications permettent enfin d'étudier l'opinion de manière dynamique¹⁶⁴, là où les sondages ne proposent qu'une étude statique, « photographique ». Ainsi pour Dominique Boullier et Audrey Lohard, « *l'opinion sondée comme agrégat des expressions individuelles peut en effet devenir limitée à la valeur d'une photo quand la conversation décrite par l'opinion mining propose, elle, le film, l'image animée avec rembobinage à volonté. L'opinion sondée sera dès lors considérée comme un état obtenu à un temps T de la société déjà faite, alors que l'opinion mining traitant les conversations visera à rendre compte de la société en train de se faire* »¹⁶⁵. Selon ces auteurs, le passage d'une mesure par sondage à une mesure numérisée des autopublications (l'« *opinion mining* ») se traduit par une double substitution. La première concerne la nature de l'opinion, avec le passage d'une opinion conçue comme un agrégat individuel à une opinion appréhendée comme le produit d'une discussion. La seconde a trait à la temporalité, avec le passage d'un temps figé à un temps continu, d'une opinion « déjà faite » à une opinion « en train de se faire ». La substitution d'une représentation processuelle et collective de la formation des opinions à une conception individuelle et privée n'est pas propre à l'étude des autopublications. Des recherches reposant sur des entretiens collectifs¹⁶⁶ ou des travaux issus des théories délibératives de l'opinion¹⁶⁷ ont en effet déjà expérimenté des instruments à même de mesurer l'opinion dans sa dimension conversationnelle. De plus, si internet permet bien de matérialiser des discussions politiques¹⁶⁸, de nombreuses autopublications ne

¹⁶⁴ Précisons d'emblée, comme nous l'avons fait dans l'introduction de ce chapitre, qu'à ce stade de la démonstration nous mobiliserons la notion d'« opinion » sans la conceptualiser, en la réduisant à une tonalité associée à une autopublication (positive, neutre, négative). Nous verrons dans la prochaine section que les définitions et conceptualisations de l'opinion varie selon les acteurs qui les mobilisent.

¹⁶⁵ Dominique Boullier et Audrey Lohard, *Opinion mining et Sentiment analysis : Méthodes et outils*, *op. cit.*

¹⁶⁶ Sophie Duchesne et Florence Haegel, *L'entretien collectif*, *op. cit.*

¹⁶⁷ James Fishkin, *Democracy and Deliberation*, Yale, Yale University Press, 1991 ; Loïc Blondiaux, « Sondages et délibération. Une épistémologie alternative de l'opinion publique », *Politix*, Volume 15, Numéro 57, 2002, p. 167-180.

¹⁶⁸ Fabienne Greffet et Stéphanie Wojcik, « Parler politique en ligne. Une revue des travaux français et anglo-saxons », *Réseaux*, 2008/4 (n° 150), p. 19-50.

s'insèrent pas dans le cadre d'un échange collectif, ou du moins d'un échange manifeste sur le web. En revanche, la substitution d'un temps figé à un temps continu est quant à elle spécifique à l'étude des autopublications. Si l'étude par panel avait déjà introduit une analyse sérielle de l'opinion, centrée sur l'évolution ou la permanence des opinions au cours du temps, l'unité temporelle retenue était, pour des raisons de coût et de faisabilité, au mieux le jour, voire la semaine ou le mois. La véritable nouveauté avec les autopublications est que l'unité temporelle devient la seconde et la photographie devient donc un film.

2.3.2. Les représentations graphiques des dynamiques de messages

En plus d'autoriser un suivi en temps réel des autopublications numériques, les logiciels de collecte et de traitement permettent également de représenter graphiquement ces dynamiques, non pas sous la forme de courbes statiques, mais à partir de graphes dynamiques incluant une dimension spatiale.

Internet peut en effet être appréhendé comme un territoire informationnel « *dans lequel les proximités entre sites et groupes de sites peuvent être représentées en un système d'attraction et de répulsion par des liens hypertextes que les sites nouent entre eux*¹⁶⁹ ». Les blogs publient en effet des « liens sortants » qui dirigent les internautes vers d'autres blogs et sites internet. Il a été empiriquement démontré que les liens sortants référencés par un blog peuvent être généralement appréhendés comme des liens affinitaires¹⁷⁰. Dans le cas de la blogosphère politique, les blogs de gauche ont ainsi tendance à davantage publier des liens vers d'autres blogs de gauche plutôt que vers des blogs de droite. Dès lors, « *la proximité relationnelle des sites définit leur ancrage « territorial », leur position dans la géographie du web, en rapprochant spatialement des sites présentant des patterns de liens entrants et sortants proches* »¹⁷¹. Contrairement aux cartes géographiques qui possèdent une certaine fixité, pour les cartographies du web, « *l'espace n'existe que par le système de signe qui le fait advenir* »¹⁷².

¹⁶⁹ Dominique Cardon, Guilhem Fouetillou et Clémence Lerondeau, « Esquisse de géographie de la blogosphère politique (2007-2009), in Fabienne Greffet (dir.) *Continuerlalutte.com, Les partis politiques sur le web*, Paris, Les Presses de Science Po. 2011.

¹⁷⁰ Jean-Christophe Plantin, « D'une carte à l'autre : Le potentiel heuristique de la comparaison entre graphe du web et carte géographique », in Christine Barats (dir.), *Manuel d'analyse du Web, op. cit.*

¹⁷¹ *Ibidem.*

¹⁷² *Ibidem.*

Illustration 1.1 : Cartographie de la blogosphère politique en 2007

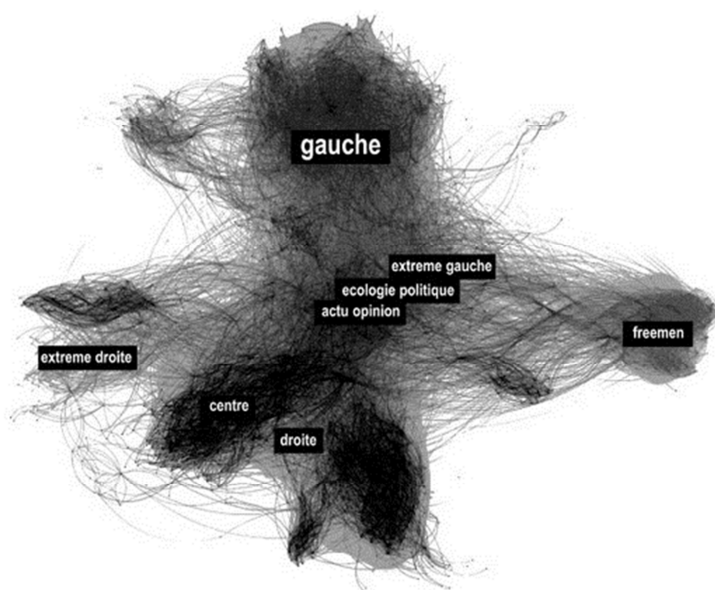


Illustration issue de *Continuerlalutte.com*, chapitre 3¹⁷³

De nombreux travaux ont entrepris de cartographier les réseaux socionumériques. Les recherches sur les graphes du web s'inscrivent d'ailleurs dans la lignée de la sociologie des réseaux et notamment des travaux de Milgram¹⁷⁴ sur les « les petits mondes » ou de Granovetter sur la force des liens faibles¹⁷⁵. Internet a donné une nouvelle actualité et de nouvelles perspectives à la sociologie des réseaux, en permettant de tester des hypothèses anciennes ou nouvelles sur des corpus particulièrement riches, issus des réseaux socionumériques¹⁷⁶.

Les cartographies des réseaux socionumériques permettent de représenter graphiquement deux types d'information : les propriétés des relations interindividuelles et les propriétés individuelles des comptes. La possibilité de représenter graphiquement les réseaux de relations constituent le principal apport méthodologique de la théorie des graphes¹⁷⁷. Sur le réseau socionumérique Facebook, les réseaux de relations entre individus sont matérialisés

¹⁷³ Dominique Cardon, Guilhem Fouetillou et Clémence Lerondeau, « Esquisse de géographie de la blogosphère politique (2007-2009), in Fabienne Greffet (dir.) *Continuerlalutte.com, Les partis politiques sur le web*, op. cit.

¹⁷⁴ Stanley Milgram et Jeffrey Travers, « An Experimental Study of the Small World Problem », *Sociometry*, 1969, Vol.32, n°4. (1), p. 425-443

¹⁷⁵ Mark S. Granovetter, « The Strength of Weak Ties », *American Journal of Sociology*, Vol. 78, n°6, Mai 1973.

¹⁷⁶ Jean-Christophe Plantin, « D'une carte à l'autre : Le potentiel heuristique de la comparaison entre graphe du web et carte géographique », in Christine Barats (dir.), *Manuel d'analyse du Web*, op. cit., p. 233.

¹⁷⁷ Pierre Mercklé, *Sociologie des réseaux sociaux*, Paris, La Découverte, 2011.

par la structure des liens affinitaires et électifs d'« amitié » tandis que sur Twitter, ces réseaux sont symbolisés par des structures d'« abonnements ». Ces liens sont représentés graphiquement par des « arêtes » ou des « arcs » reliant différents « nœuds » ou « points », c'est-à-dire différents comptes.

Les propriétés individuelles des comptes peuvent être représentées graphiquement par la couleur ou le diamètre des « nœuds ». Ainsi, dans la cartographie de la blogosphère politique proposée par Dominique Cardon *et alii*, l'appartenance politique des blogs est symbolisée par une couleur attribuée à chaque nœud. Les blogs politiques de gauche sont symbolisés par des nœuds colorés en rose tandis que les blogs de droite le sont par des nœuds bleus, etc. Les variables qualitatives (sexe, PCS, thématique, etc.) se prêtent ainsi davantage à une coloration graphique tandis que les variables quantitatives telles que le nombre d'amis, le nombre d'abonnés ou l'âge peuvent être mesurées par le diamètre des nœuds. La visualisation de ces graphes peut apporter des renseignements relatifs au degré d'ouverture des réseaux et, corrélativement, à la plus ou moins grande homogénéité sociale des groupes.

Les graphes du web peuvent donc représenter graphiquement l'état d'une population à temps t . La cartographie de la blogosphère politique représente ainsi l'état de la population exhaustive des blogs politiques français en 2007. Elle est donc une représentation statique et non dynamique de la blogosphère. Mais les graphes du web peuvent également représenter graphiquement la circulation d'un message ou l'évolution d'une thématique, d'une information, d'une opinion ou d'une rumeur au sein de cette population. En représentant la circulation de ces messages, les graphes deviennent alors dynamiques, la représentation temporelle vient s'ajouter à la représentation spatiale. Ces représentations dynamiques prennent alors la forme de schémas animés interactifs. Le chercheur peut, comme avec un film, revenir en arrière, accélérer, faire des arrêts sur image et alors se mouvoir sur le graphe, le manipuler, en jouant par exemple entre les différents niveaux de zoom, en passant d'une vision synoptique à une vue plus resserrée, etc.

Ces graphes dynamiques permettent de visualiser la circulation d'un message entre différents groupes sociaux dont, on l'a vu, on peut définir l'homogénéité ou au contraire l'hétérogénéité sociale. Comment se diffuse un message, une rumeur, une opinion, une information ? Quant et par qui est-elle reprise ? Quels sont les caractéristiques individuelles des individus ou des groupes qui créent et/ou diffusent et reprennent ces informations ? L'information change-t-elle de nature au cours de sa diffusion entre des groupes aux priorités

différentes¹⁷⁸ ? L'étude des graphes du web permettent en théorie d'apporter des éléments de réponse à ces problématiques de recherche.

Néanmoins, dans certaines configurations, la valeur heuristique des graphes et cartographies du web peut être limitée. Les logiciels de *crawling* fournissent la possibilité de générer des graphes de façon automatique à partir de millions de données. Cette surabondance peut d'abord rendre les cartes totalement illisibles. Au-delà de cet aspect formel, la lecture d'un graphe généré automatiquement peut s'avérer heuristiquement très pauvre si le chercheur n'a pas une connaissance suffisante des modalités de constitution du corpus, de la nature et de la fonction des algorithmes calculés par les logiciels, etc. Comme le note Jean-Christophe Plantin, « rien ne ressemble plus, à première vue, à un lien qu'un autre lien »¹⁷⁹. Pourtant ces liens ne relèvent pas du hasard mais d'un calcul algorithmique. Le chercheur, pour donner du sens à ces liens, pour comprendre s'ils sont dirigés ou non, et le cas échéant, pour interpréter le sens de cette direction, doit avoir une connaissance et une maîtrise suffisantes de son logiciel et de son corpus. À la différence d'un histogramme ou d'un tableau croisé, la compréhension et le sens que l'on tire d'un graphe semblent *a priori* moins immédiats.

Les autopublications numériques offrent donc à la fois de réels avantages épistémologiques mais présentent également certaines limites.

Dès lors, l'étude scientifique de ce nouveau matériau conduit les chercheurs à arrêter un certain nombre d'arbitrages. La taille des corpus étudiés (échantillons raisonnés ou corpus exhaustifs), la qualification (ou l'absence de qualification) des auteurs des messages, le degré d'automatisation de l'analyse (manuelle, semi-automatisée ou exclusivement automatisée) et le choix des méthodes (qualitatives, quantitatives ou mixtes) constituent les quatre principaux enjeux de l'analyse des autopublications numériques.

Les arbitrages et stratégies de recherche adoptés ainsi que les théories et les concepts mobilisés pour étudier les autopublications numériques dépendent des intérêts scientifiques et/ou économiques des agents qui ont investi ce nouveau matériau.

¹⁷⁸ Henry Jenkins, « If it doesn't spread, it's dead: Media Viruses and Memes », *Confessions of an Aca fan – The blog of Henry Jenkins*, février 2009.

¹⁷⁹ Jean-Christophe Plantin, « D'une carte à l'autre : Le potentiel heuristique de la comparaison entre graphe du web et carte géographique », in Christine Barats (dir.), *Manuel d'analyse du Web*, *op. cit.*

3. Un nouveau champ de l'opinion ?

Les autopublications numériques possèdent, au-delà de leur valeur scientifique, une valeur économique. Dès le milieu des années 2000, un nouveau secteur d'activité s'est en effet progressivement constitué autour de l'étude de ce nouveau matériau¹⁸⁰. Parallèlement et souvent antérieurement aux disciplines académiques – c'est d'ailleurs pour cette raison que nous aborderons dans l'ordre de notre démonstration les travaux scientifiques seulement après avoir évoqué le secteur économique –, ces acteurs commerciaux ont créé, légitimé et parfois « théorisé » de nouvelles méthodes et de nouveaux instruments de mesure des autopublications numériques. L'histoire de ces méthodes et de ces instruments se situe donc à la croisée des champs scientifique et économique.

Bien que ces champs obéissent à des logiques propres, le même phénomène de diversification des acteurs « traditionnels » de l'opinion peut y être observé. Dans le champ commercial, une nouvelle série d'acteurs est ainsi venue concurrencer les instituts de sondage et tenter de disputer leur monopole. Dans le champ académique, alors que les sciences de l'information et de la communication, les sciences de gestion et les sciences du langage ont fait de l'étude des autopublications un axe de recherche majeur, la sociologie politique n'y a au contraire accordé qu'un intérêt très limité (3.1).

L'arrivée de ces nouveaux acteurs et la relative marginalisation des acteurs « traditionnels » du champ de l'opinion ont conduit à la mise en circulation de définitions et d'approches théoriques concurrentes voire parfois contradictoires des opinions numériques. Les trois grandes approches théoriques des autopublications ne conçoivent ni ne mesurent les opinions selon les mêmes problématiques, les mêmes cadres théoriques, les mêmes instruments et les mêmes méthodes. Ces approches théoriques sont de plus relativement éloignées des problématiques sociologiques de l'opinion publique (3.2).

¹⁸⁰ Dominique Boullier et Audrey Lohard, *Opinion mining et Sentiment analysis : Méthodes et outils*, op. cit.

3.1. L'intérêt différencié des acteurs scientifiques et commerciaux

Si les autopublications constituent un matériau commun aux études scientifiques et commerciales, les méthodes et les approches adoptées ne sont pas les mêmes. Alors que dans le champ académique, et plus particulièrement dans les travaux de sociologie et de science de l'information et de la communication, les recherches sur les autopublications revêtent souvent une dimension « artisanale » et privilégient une analyse « manuelle » et qualitative de petits corpus de données, la production des études commerciales est quant à elle davantage « industrielle » – bien que peu standardisée – et centrée sur des analyses essentiellement quantitatives, partiellement ou entièrement automatisées, d'importants volumes de données (3.1.1).

Les logiciels et les instruments de mesure des autopublications numériques ont donc été créés, légitimés et théorisés au sein d'un nouveau sous-secteur économique marqué par l'arrivée d'une série diversifiée d'acteurs (agence de communication, de *e*-réputation, producteurs de logiciels, etc.) venue concurrencer les instituts de sondage. La numérisation des terrains et des méthodes, en réduisant le coût de production des données, a contribué à abaisser le coût d'entrée dans le champ des études d'opinion (3.1.2).

La diffusion parfois plus tardive des logiciels et instruments de mesure des autopublications dans le champ académique s'est principalement orientée vers des disciplines ayant développé un rapport précoce à l'informatique, comme la sociologie des réseaux, les sciences du langage ou les sciences de l'information et de la communication. (3.1.3).

3.1.1. Un matériau commun aux études scientifiques et commerciales

Les autopublications numériques sont un matériau commun aux études scientifiques et commerciales. Les sciences sociales n'en ont pas le monopole. Cette situation n'est cependant pas inédite. La majorité des méthodes quantitatives et qualitatives sont en effet communes aux disciplines académiques et aux études d'opinion et de marché commercialisées par les instituts d'études et de conseil.

La technologie du sondage d'opinion est sans doute l'exemple le plus connu d'importation dans le champ académique de méthodes issues du champ commercial des

études. Loïc Blondiaux, dans son histoire sociale des sondages¹⁸¹, a ainsi rappelé l'origine de cette technique d'enquête, issue de l'univers du marketing et des études de marché aux États-Unis dans les années 1930, puis a montré comment les sondages d'opinion ont dès leur naissance été conçus et légitimés comme des outils de recherche appliquée, notamment sous l'impulsion de Paul Lazarsfeld à l'*Office of Radio Research* puis au *Bureau of Applied Social Research*. En France comme aux États-Unis, les « pères fondateurs » des sondages ont ainsi entrepris d'apporter une légitimité à la fois commerciale et scientifique à l'instrument¹⁸². Cette double appartenance restera une propriété constitutive des sondages. Pierre Bourdieu, dans sa critique des sondages, s'adressera ainsi tout autant aux sondeurs professionnels des instituts qu'aux « politologues » universitaires, désignés sous la qualification commune de « doxosophes »¹⁸³. De manière analogue, Sophie Duchesne et Florence Haegel ont rappelé comment les *focus group* ont été à l'origine un outil créé par la recherche appliquée, notamment sous l'influence de Robert Merton aux États-Unis dans les années 1940, puis comment le marketing s'est emparé de la méthode dans les années 1950 jusqu'à sa redécouverte par les sciences sociales dans les années 1980¹⁸⁴.

L'étude des autopublications, en se situant à la croisée des champs scientifique et commercial, s'inscrit donc dans une trajectoire historique similaire à celle des sondages ou des *focus groups*. Cependant, à la différence des réponses de questionnaire ou des propos recueillis en entretien, qui n'ont de réalité empirique que par les méthodes et les instruments qui les mettent au jour, les autopublications ont une existence propre, antérieure et indépendante de leurs instruments d'étude. On peut ainsi dissocier conceptuellement et pratiquement le matériau étudié – les verbatims publiés par les internautes sur les espaces d'autopublication – des méthodes et des instruments qui les mesurent.

L'usage de logiciels de veille et de *tracking* a longtemps été un élément distinctif entre les études scientifiques et les études commerciales des autopublications numériques. En effet, alors que les recherches académiques en sciences sociales, revêtaient – et revêtent toujours dans leur majorité – une dimension « artisanale », adoptant ainsi une analyse manuelle de petits corpus de données, les études commerciales ont dès leur apparition eu recours à des logiciels industriels de veille et de *tracking*, proposant ainsi des analyses essentiellement

¹⁸¹ Loïc Blondiaux, *La Fabrique de l'opinion*, op. cit.

¹⁸² Loïc Blondiaux, « Paul F. Lazarsfeld (1901-1976) et Jean Stetzel (1910-1987) et les sondages d'opinion : genèse d'un discours scientifique », *Mots*, Vol. 23, n°23, 1990, p. 5-23.

¹⁸³ Pierre Bourdieu, « Les doxosophes », art. cit.

¹⁸⁴ Sophie Duchesne et Florence Haegel, *L'entretien collectif*, op. cit.

quantitatives, partiellement ou entièrement automatisées d'impressionnants volumes de données. Ces éléments d'opposition entre pratique scientifique et pratique commerciale avaient déjà été observés par François Backman et Olivier Rozenberg dans leur étude comparée de la pratique des *focus groups* en instituts d'études et en recherche¹⁸⁵. Les deux auteurs avaient en effet notamment constaté « *des écarts d'échelle* » entre les grands instituts et les laboratoires de recherche, avec d'un côté une pratique industrielle et de l'autre une pratique artisanale.

Les logiciels de *tracking* et les instruments de mesure des autopublications, avant d'être importés dans le champ académique, ont donc pour la plupart été créés, légitimés et parfois théorisés au sein des champs professionnels des études, du marketing et de la communication.

3.1.2. La création d'un nouveau secteur d'activité économique marqué par l'arrivée de nouveaux acteurs, concurrents des instituts de sondage

L'étude des autopublications numériques constitue un secteur d'activité économique à part entière. Un secteur caractérisé par une surabondance de l'offre, une grande hétérogénéité des acteurs, des frontières floues – entre études, communication, marketing, veille et conseil – et des perspectives de développement et de rentabilité incertaines. Il ne s'agira pas ici de dresser une sociogenèse de ce champ, un chapitre n'y suffirait pas. Nous nous bornerons à évoquer les principales causes et quelques conséquences de l'ouverture du champ professionnel de l'opinion à des acteurs issus de la communication, du marketing ou de la veille.

Dominique Boullier et Audrey Lohard ont recensé une vingtaine de termes différents pour qualifier les études et services proposés autour des autopublications numériques¹⁸⁶ : « *e-réputation* », « *veille marketing* », « *veille digitale* », « *référencement* », « *community management* », « *communication interactive* », « *veille en intelligence économique* », « *buzz tracking* », « *brand monitoring* », « *marketing communautaire* », « *web marketing* », « *marketing digital* », « *sentiment analysis* », « *social analytics* », « *relation blogueurs* », « *e-*

¹⁸⁵ Olivier Rozenberg et François Backman, « Mortels transferts ? La pratique des entretiens collectifs par les Grands Instituts de Sondages Internationaux et par la recherche en Science Sociale », art. cit.

¹⁸⁶ Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis*, op. cit.

influence », « *sentiment monitoring* », « *veille d'image* », « *réputation digitale* », « *online sentiment* », « *mesure de la visibilité* », « *suivi de notoriété* » ou encore « *buzz monitoring* », etc. Cette profusion de termes et d'appellations est révélatrice de la très grande hétérogénéité des acteurs qui ont investi ce secteur. On peut sommairement identifier cinq grandes catégories d'acteurs : les instituts d'études et de conseils (GfK, Ipsos, Médiamétrie, BVA, etc.), les instituts « *pure player* », c'est-à-dire spécialisés exclusivement dans les études numériques (Linkfluence, Trendybuzz, Semiocast, etc.), les agences de communication et de publicité (Publicis, Havas, etc.), les « *web agency* », spécialisées dans toutes les activités liées à internet (création de site internet, référencement et mise en valeur de contenu, etc.) et enfin les éditeurs de logiciels – qui commercialisent pour certains des études (Lingway, Digimind, Spotter, etc.). En 2011, on recensait ainsi plus de 160 entreprises proposant des études numériques au sens large¹⁸⁷.

Le secteur des études numériques regroupe donc des acteurs, des professions et des secteurs d'activité qui étaient jusqu'alors relativement cloisonnés : le secteur de la veille (concurrentielle, économique, juridique), le secteur des études (d'opinion, de marché, de marketing), le secteur du conseil (en vente, en stratégie financière, etc.) et le secteur de la communication et de la publicité. La matérialisation et la concentration de données de différentes natures sur un même espace – numérique – peuvent constituer un élément d'explication de la plus grande porosité de ces frontières sectorielles sur internet. Sur un même réseau socionumérique – Twitter – un analyste réalisant une veille pourra trouver des informations qui n'ont pas encore été reprises par les agences de presse, un chargé d'étude pourra « sonder » en temps réel les réactions des internautes à des événements en direct tandis qu'un chargé de communication pourra publier massivement des tweets pour promouvoir la marque ou le produit de son client. De plus, les acteurs de ces différents secteurs ont souvent recours aux mêmes logiciels de veille et de *tracking* pour collecter et traiter leurs données. Le logiciel *Ami Opinion Tracker*, initialement conçu pour la veille et l'intelligence économique, est aujourd'hui commercialisé et présenté comme « *un outil parfaitement adapté aux métiers des études et de la communication, en plus d'un savoir-faire hérité en matière de veille et d'Intelligence économique*¹⁸⁸ ». Cette concentration des données de différentes natures sur un même espace numérique et leur traitement par des logiciels identiques constituent donc des

¹⁸⁷ Nous qualifions de « numériques » les études commerciales prenant pour objet les autopublications numériques. Ces études n'incluent pas les études CAWI.

¹⁸⁸ <http://www.amisw.com/fr/produits/ami-entreprise-intelligence/logiciel-e-reputation.htm>

éléments d'explication de la plus grande porosité des secteurs de la veille, des études, du conseil ou de la communication sur internet.

Cette plus grande porosité a ainsi conduit des acteurs venus d'autres secteurs à produire également des études numériques. En disposant des mêmes logiciels et en ayant accès aux mêmes données, des agences de communication et des entreprises de veille ont ainsi pu réaliser à leur tour des études numériques. Ce phénomène d'ouverture a de plus été renforcé par une baisse relative des coûts de production des études sur internet. Les coûts de production des études numériques sont en effet bien moins élevés que ceux d'une étude par sondage ou d'une étude qualitative commercialisée en institut. Cette baisse des coûts concerne à la fois la phase de production des données mais aussi celle de l'analyse. Alors qu'une étude par sondage nécessite la mobilisation d'une vingtaine d'enquêteurs en moyenne et peut durer plusieurs jours, la phase de collecte des autopublications numériques peut être réalisée par un seul enquêteur, ou directement par le chargé d'étude, en quelques heures seulement. Le coût de l'analyse des données est quant à lui difficile à estimer et est variable d'une société à l'autre. À partir d'un même corpus de données, cette phase d'analyse peut durer en effet plusieurs jours, plusieurs heures voire moins d'une heure. L'automatisation de l'analyse des données par les logiciels d'*opinion mining* et de *sentiment analysis*, le degré de qualification du chargé d'étude, l'absence de standardisation des méthodes ainsi que la relative méconnaissance par les clients des potentialités et des limites des études numériques sont quelques éléments d'explication de la grande plasticité des pratiques, de la qualité et *in fine* des coûts de production des études numériques. La livraison des résultats peut ainsi prendre la forme d'un rapport d'étude analytique détaillé de plusieurs pages, d'un simple « *reporting* » présentant quelques graphiques et données chiffrées accompagnés de commentaires descriptifs, ou encore d'un « tableau de bord » mis à la disposition du client et présentant en temps réel l'évolution de la volumétrie, de la tonalité ou de la distribution des résultats par mot-clé ou selon différentes variables (types de sources, catégories thématiques, etc.). Au final, alors que la production d'une étude « classique » nécessite une forte division du travail et d'importants moyens humains (une batterie d'enquêteurs, plusieurs statisticiens, un ou deux chargés d'études) et logistiques (location d'une plateforme téléphonique, etc.), une étude numérique peut être réalisée par un seul chargé d'étude (sans qu'il ait d'ailleurs la qualification et le titre de chargé d'étude), avec comme seul outil de production un logiciel de veille et de *tracking*. Cette baisse des coûts de production a donc abaissé le coût d'entrée dans

le champ professionnel des études d'opinion et a permis à une grande diversité d'acteurs de produire et commercialiser des études numériques.

L'ouverture du secteur a donc logiquement entravé le monopole des instituts de sondage. Cette ouverture, notamment aux acteurs de la communication, a eu pour conséquence une absence totale de standardisation des études, que cela soit au niveau de leurs dénominations – le rapide inventaire dressé plus haut nous en a donné un aperçu – au niveau des concepts et notions mobilisés (« opinion », « réputation », « influence », « sentiment », etc.) ou encore au niveau des méthodes employées.

Parmi cette offre abondante et hétérogène, Dominique Boullier et Audrey Lohard distinguent deux principales postures : « *observer pour comprendre* » et « *observer pour agir* »¹⁸⁹. La première posture consiste à appréhender internet comme un terrain d'observation, où s'expriment et circulent des opinions formulées par des internautes « citoyens » ou « consommateurs ». Cette posture attribue donc aux études numériques des fonctions similaires à celles des études « classiques » : comprendre et anticiper les besoins des consommateurs, évaluer la réception d'un produit ou d'une campagne de communication, etc. On peut, parmi les études adoptant cette première posture, distinguer deux principales catégories : les études quantitatives de tonalité et les études qualitatives ou cartographiques de « communautés en ligne ».

Les études quantitatives de tonalité sont les études numériques qui se rapprochent formellement le plus des études par sondage, certains auteurs anglo-saxons les qualifiant même de « *semantic polls* »¹⁹⁰. Ces études proposent généralement une analyse volumétrique de citations (combien d'internautes, sur différents supports du web, ont cité l'objet étudié – un produit, une marque, une entreprise, etc.), suivie d'une analyse – manuelle, automatisée ou partiellement automatisée – de la tonalité des messages évoquant l'objet. Est alors associé à l'objet étudié un pourcentage d'avis positif, neutre ou négatif. Ces études, majoritairement qualifiées d'« *e-réputation* », peuvent prendre également pour objet des hommes politiques. Ainsi, l'institut TNS Sofres et la société SemioCast, en collaboration avec le site *TF1.fr* et le journal gratuit *Métro*, publia mensuellement de janvier 2011 à juin 2012 le « Twittoscope », un « *nouveau baromètre d'opinion* »¹⁹¹ établissant chaque mois le classement des

¹⁸⁹ Dominique Boullier et Audrey Lohard, *Opinion mining et Sentiment analysis : Méthodes et outils*, op. cit.

¹⁹⁰ Nick Anstead, « Social media metrics: measuring manifestations of the public in online spaces », Colloque international *Regards critiques sur la participation politique en ligne*, Paris, 19-20 juin 2013.

¹⁹¹ Études en ligne : http://www.tns-sofres.com/_assets/files/surveys/2011.02.09-twittoscope-analyse.pdf

personnalités politiques les plus citées sur Twitter et mesurait pour chacune d'elle la tonalité des tweets l'évoquant (avec une échelle en quatre items : rejet, critique, neutre, positif). À l'instar des sondages d'opinion, les commentaires du « baromètre » précisaient les variations mensuelles de la courbe d'avis positifs ou d'avis négatifs des tweets prenant pour objet les hommes politiques étudiés.

Le second type d'études adoptant « une posture de compréhension » propose quant à elles une analyse qualitative ou cartographique de « communautés en ligne ». Il s'agit alors d'identifier des « communautés d'internautes » échangeant plus ou moins régulièrement des messages sur des espaces d'autopublication thématiques (des forums ou réseaux sociaux consacrés au sport dans le cas d'une étude réalisée pour un équipementier sportif, par exemple, etc.). L'étude consiste alors à identifier les internautes de cette communauté, analyser leurs discussions et éventuellement cartographier leur réseau de relations. Ces études reposent sur les théories du « marketing viral ». Le marketing viral, issu comme nous le verrons de la rencontre entre les théories sociologiques de l'influence interpersonnelle, de leur application dans le domaine du marketing (« *word-of-mouth marketing* ») et de l'analyse mathématique des réseaux socionumériques, appréhende les internautes comme des individus en capacité d'influencer leur réseau de relations sur internet et ainsi de faire et défaire la réputation d'une marque ou d'un produit¹⁹². Les consommateurs seraient ainsi devenus des « consom'acteurs »¹⁹³. Ce jeu de mot, formé de la contraction de « consommateurs » et d'« acteurs » traduirait ainsi le changement de perception ontologique des individus par le marketing, passés du statut de récepteurs passifs à celui de promoteurs actifs. Dès lors, il deviendrait essentiel pour les entreprises d'identifier ces « leaders d'opinion » et d'analyser leurs opinions afin d'anticiper les réactions futures des consommateurs.

Ces deux catégories d'études, bien que reposant sur des théories et des méthodologies différentes, ont pour point commun de circonscrire les études numériques à leur fonction d'observation et de compréhension des phénomènes sociaux. Si les conclusions et recommandations de l'étude peuvent inciter le commanditaire à agir, en faisant appel à une agence de communication pour corriger et réorienter l'image de sa marque par exemple, cela ne se fera que dans un second temps, un temps distinct de celui de l'étude.

¹⁹² Kévin Mellet, « Aux sources du marketing viral », *Réseaux*, 2009/5-6 (n° 157-158), p. 267-292.

¹⁹³ Laurent Florès, « Web 2.0 et études de marché : vers une nouvelle génération d'études de marché ? », *Revue Française du Marketing*, 220 (2008), p. 7-16.

La seconde posture (« observer pour agir ») adoptée par les acteurs du secteur conçoit internet comme un terrain d'action où il est possible de convaincre les internautes et d'influer sur leurs opinions et intentions d'achat. Ces études reposent en effet sur le postulat selon lequel les individus se constituent des opinions sur un objet (une marque, un produit, une entreprise, etc.) à partir d'une lecture des commentaires d'autres internautes ou à partir d'un échange ou d'une discussion avec ces internautes. De nombreuses études qui ont ainsi démontré que les consommateurs prenaient en compte dans leur décision d'achat sur internet les commentaires formulés par d'autres internautes-consommateurs¹⁹⁴. Après avoir observé les opinions circulant sur un objet, l'enjeu est alors de promouvoir l'objet, soit en créant artificiellement des messages positifs et en améliorant leur référencement, soit en associant des « leaders d'opinion » à la promotion de l'objet. Les études et services de « *gestion d'e-réputation* » consistent à produire de faux avis d'internautes et à améliorer leur référencement. L'objectif est donc que le consommateur, lorsqu'il consulte les avis et commentaires publiés sur un site d'achat, lise prioritairement les commentaires positifs. Le second type d'étude, qualifié de « *community management* » consiste à identifier et mobiliser des communautés d'internautes autour d'un objet afin d'en faire des promoteurs de l'objet. Il ne s'agit donc plus d'observer passivement les communautés mais bien de les mobiliser pour les associer à une démarche de promotion de l'objet. Les études numériques adoptant cette seconde posture, qu'elles proposent de créer artificiellement de faux avis de consommateurs ou d'associer les internautes à la stratégie de promotion de la marque, se situent alors davantage dans une stratégie de communication et de promotion que dans une démarche de connaissance des consommateurs et de leurs besoins, envies ou attentes.

Si la majorité des entreprises du secteur adoptent soit l'une soit l'autre de ces postures, Dominique Boullier et Audrey Lohard notent cependant que « *sous l'effet de la concurrence, les offres semblent tendre peu à peu à se ressembler et à se rapprocher vers une sorte de modèle « 360 »* »¹⁹⁵. Ces « offres 360 » proposent d'étudier et de mesurer les opinions des internautes sur un objet donné, de produire des opinions fictives ou d'encourager les enquêtés à publier des messages pour améliorer la « réputation » de l'objet et enfin de mesurer à nouveau cette dernière. Ces offres fusionnent donc des activités d'étude, de production

¹⁹⁴ Selon diverses études synthétisées par le site Olimeo.com, en 2010, 86% des internautes feraient « confiance aux avis » des consommateurs sur les sites d'achat : <http://frenchweb.fr/infographie-importance-des-avis-des-consommateurs-sur-le-web/20070>

¹⁹⁵ Dominique Boullier et Audrey Lohard, *Opinion mining et Sentiment analysis : Méthodes et outils, op. cit.*

communicationnelle et de valorisation marketing qui étaient jusqu'alors relativement cloisonnées.

Classiquement, une entreprise ou une organisation souhaitant réorienter une campagne de communication commande à un institut une étude, qualitative ou quantitative, pour évaluer la campagne existante. À partir des résultats de l'étude, l'entreprise peut ensuite faire appel à une agence de communication pour réorienter sa campagne. Enfin, avant de lancer la nouvelle campagne, l'entreprise pourra la tester en amont en commandant une nouvelle étude. Les délais entre chaque étape du processus peuvent durer plusieurs semaines voire plusieurs mois. Sur internet, ces délais peuvent être réduits à quelques heures et la réalisation de l'ensemble des opérations (étude exploratoire en amont – plan de communication – étude évaluatrice en aval) réalisé par un seul prestataire. Une même entreprise peut ainsi proposer à ses clients de connaître le volume et la tonalité des messages qui circulent sur une de ses marques (« *opinion mining* », « *sentiment analysis* », « suivi d'*e*-réputation »), entreprendre une stratégie de communication pour contrecarrer la diffusion d'éléments négatifs et/ou pour augmenter la visibilité de sa marque (« opération de référencement », « gestion d'*e*-réputation » et « *community management* »), puis enfin évaluer les effets de cette opération de référencement.

Cette fusion des activités d'étude et de production communicationnelle introduit une rupture fondamentale dans l'étude – commerciale – de l'opinion. Cette rupture concerne l'« impératif de sincérité » des matériaux étudiés, érigée en norme jusqu'alors par les professionnels des études. Les enquêtes d'opinion et de marketing reposent en effet sur la croyance selon laquelle les opinions sont des objets « purs » que le dispositif d'enquête ne doit en aucun cas altérer¹⁹⁶. Les enquêteurs ou les animateurs de *focus group* doivent ainsi tout faire pour ne pas orienter les propos des enquêtés¹⁹⁷. L'inéligibilité de certains individus aux sondages d'opinion en raison de leurs professions jugées trop proches du métier des études – les professionnels de la publicité, du marketing, les journalistes, voire les professionnels du secteur sur lequel porte l'enquête – est un exemple significatif de cette volonté de ne recueillir que des opinions « authentiques » et en aucun cas orientées pour les besoins de l'enquête ou pour satisfaire les commanditaires. Avec ces « offres 360 », l'authenticité des opinions étudiées et la neutralité de la situation d'enquête ne sont plus des

¹⁹⁶ Loïc Blondiaux, « Sondages et délibération. Une épistémologie alternative de l'opinion publique », *Politix*, Vol. 15, n°57, 2002, p. 167-180.

¹⁹⁷ Olivier Rozenberg et François Backman, « Mortels transferts ? La pratique des entretiens collectifs par les Grands Instituts de Sondages Internationaux et par la recherche en Science Sociale », art. cit.

impératifs, l'enjeu principal étant de veiller à la bonne « réputation » d'un objet, quitte à ce que le prestataire de l'étude ou les enquêtés eux-mêmes prennent part à ce travail de réputation.

Ce changement de posture résulte de la diffusion dans le champ des études des théories du « marketing viral ». D'une manière analogue au champ marketing, où le consommateur est passé d'un statut de récepteur passif à un statut de promoteur actif, les enquêtés seraient passés du statut passif de « répondants » au statut actif de « partenaires de l'étude » : *« durant trop longtemps, il semblerait que la « ressource » répondants ait été considérée comme un dû, sans véritablement chercher à valoriser l'individu répondant comme autre chose qu'un répondant, non comme un véritable partenaire d'échange, tout à fait capable de créer et d'influencer. Tout cela conduit à un besoin de revalorisation de l'échange entre questionnant et répondant, le premier devant plus que jamais valoriser le second pour maximiser la qualité des idées issues d'une collaboration équitaine »*¹⁹⁸. Dans cette optique, les études ne doivent donc pas se contenter d'observer et d'analyser passivement les communautés en ligne et les « leaders d'opinion » qui s'y expriment, elles doivent interagir avec eux pour en faire des partenaires de l'étude et de la marque. Il ne s'agit pas seulement de connaître l'avis des « leaders d'opinion » sur un nouveau produit, mais de valoriser ces leaders et leur signifier que l'on prend en considération ce qu'il pense et qu'on les associe à l'opération de promotion de l'objet.

L'introduction des théories du marketing viral dans le champ des études numériques permet également de légitimer l'occultation de la problématique de représentativité des opinions étudiées sur internet. En appliquant les lois de l'échantillonnage probabiliste, les sondages ont pu se prévaloir de prédire l'opinion en interrogeant seulement une partie (l'échantillon) et non le tout (l'ensemble de la population française). L'« impératif de représentativité »¹⁹⁹ est ce qui donne donc le droit aux sondages de parler au nom du tout. Sur internet, non seulement la population des internautes n'est pas représentative de la population française, mais il est de plus très difficile comme nous l'avons vu de qualifier sociologiquement les auteurs des autopublications. Les études numériques, en substituant les théories de l'influence à celle de la représentativité, éludent leur incapacité à qualifier sociologiquement les individus et partant à mesurer leur représentativité. Il ne s'agit plus alors

¹⁹⁸ Laurent Florès, « Web 2.0 et études de marché : vers une nouvelle génération d'études de marché ? », art. cit.

¹⁹⁹ Alain Desrosières, *La politique des grands nombres. Histoire de la raison statistique*, Paris, Éditions La Découverte, 1993.

de savoir si la population étudiée est représentative, mais seulement de savoir si les enquêtés sont influents, c'est-à-dire en capacité d'orienter les décisions (d'achat, de vote, de consommation de bien culturel, etc.) de leurs réseaux de connaissances. L'individu n'est plus alors défini en priorité par ses caractéristiques sociodémographiques mais par des données mesurant son « influence »²⁰⁰. Or, comme nous l'avons vu, contrairement aux caractéristiques sociodémographiques, les données métriques d'influence sont systématiquement disponibles sur les réseaux socionumériques.

Les théories du marketing viral ont donc réactualisé et importé dans le champ des études d'opinion les problématiques sociologiques de l'influence personnelle. Ces théories et problématiques étaient jusqu'alors totalement absentes du logiciel théorique des études par sondages. La diffusion et l'importation de ces théories peut être ainsi perçue comme une conséquence de l'ouverture du champ professionnel de l'opinion aux acteurs du marketing et de la communication. Ces acteurs, en intégrant le champ, y ont en effet importé leurs propres théories, concepts, méthodes et instruments.

En concentrant sur un même espace des données de différentes natures et en abaissant les coûts de production des études, internet a ouvert le champ professionnel de l'opinion à des acteurs venus de secteurs différents. Dans le champ académique, si le même phénomène d'ouverture peut être observé, il n'obéit cependant pas aux mêmes logiques.

3.1.3. Désintérêt de la sociologie politique et informatisation des sciences sociales

Dans le champ académique, les disciplines spécialisées dans l'étude des opinions – la sociologie politique et la science politique – n'ont accordé un intérêt que très limité aux autopublications numériques. À ce jour, on ne dénombre dans la *Revue Française de Science Politique* qu'un seul article prenant pour objet des autopublications numériques et adoptant des méthodes issues de l'*opinion mining*. La recherche de Sylvain Parasie et Jean-Philippe Cointet, « La presse en ligne au service de la démocratie locale. Une analyse morphologique de forums politiques »²⁰¹ est en effet le seul article paru dans la revue à mobiliser comme

²⁰⁰ En oubliant ou en omettant, comme nous le verrons, que le statut, l'âge et le sexe sont en réalité pour Lazarsfeld et Katz trois facteurs déterminants de l'influence cf. Elihu Katz et Paul Lazarsfeld, *Influence personnelle*, Paris, Arman Colin, 2008.

²⁰¹ Sylvain Parasie et Jean-Philippe Cointet, « La presse en ligne au service de la démocratie locale. Une analyse morphologique de forums politiques », art. cit.

matériau empirique des autopublications numériques. Les deux auteurs, afin d'étudier le rôle que peut jouer une organisation de presse dans les espaces politiques locaux ont mené une analyse statistique et lexicométrique d'un corpus de messages publiés sur des forums mis en place par le journal *La Voix du Nord* à l'occasion des élections municipales de 2008. La méthodologie adoptée est relativement proche de celle des études numériques : analyse de la distribution du nombre de commentaires et du nombre d'auteurs par forum, traitement lexical « *partiellement automatisé* » du corpus collecté, et calculs de significativité entre données *offline*²⁰² (taille de la commune) et données *online* (catégorie de pseudonymes utilisés). À partir d'un objet et d'une problématique « légitimes » de science politique (le rôle d'une organisation de presse dans la structuration de l'espace politique local), cette recherche adopte un matériau empirique et une méthodologie issus des études numériques. On pourrait également citer l'article de Thierry Barboni et Éric Treille, « L'engagement 2.0. Les nouveaux liens militants au sein de l'e-parti socialiste »²⁰³ paru dans le volume 60 de la revue en 2010. Cependant, bien qu'il prenne pour objet un espace d'autopublication numérique – le site la *Coopol*, mis en place par le Parti socialiste – cet article n'analyse pas les messages publiés sur cet espace et n'adopte pas pour cela une méthodologie spécifiquement numérique. *Politix*, autre revue de référence de la discipline, n'a publié quant à elle à ce jour aucune recherche mobilisant des méthodes issues de l'*opinion mining*. Or, comme le signalait Pierre Favre, à l'occasion du centième numéro de la revue : « *les revues sont l'une des pièces maîtresses du dispositif scientifique dans les sciences sociales. Les effets de diffusion des résultats de recherche et de reconnaissance disciplinaire, des articles publiés dans les revues sont supérieurs aux effets des contributions à des livres collectifs et des communications des colloques et congrès. C'est en écrivant dans les revues centrales de la discipline, les fameuses revues « à comité de lecture », que l'on se fait connaître et reconnaître* »²⁰⁴.

L'absence presque totale d'articles parus dans ces revues « à comité de lecture » témoigne donc bien de la place relativement marginale accordée jusqu'ici par la science politique aux autopublications numériques. Cependant, certains projets de recherche interdisciplinaires commencent à voir le jour. Le *médialab*, laboratoire de recherche attaché à l'*Institut d'Études Politiques* de Paris associe des chercheurs en sociologie, en science politique et en informatique. Centré sur les théories de l'acteur-réseau et les travaux de Bruno

²⁰² Nous entendons par données *offline* toutes données non-relatives à internet.

²⁰³ Thierry Barboni et Éric Treille, « L'engagement 2.0. Les nouveaux liens militants au sein de l'e-parti socialiste », *Revue française de science politique*, 2010/6 (Vol. 60), p. 1137-1157.

²⁰⁴ Pierre Favre, « *Politix*, 1988-2012 : changement de génération, basculement de paradigmes », *Politix*, 2012/4 (n° 100), p. 41-62.

Latour, quatre principaux axes de recherche y sont développés : le développement des méthodologies centrées sur l'étude des données numériques, la cartographie des controverses, l'adaptation de l'analyse textuelle aux sciences humaines et sociales et enfin l'analyse visuelle des réseaux. On peut également citer le projet de recherche *Imagiweb*²⁰⁵, dans lequel s'inscrit ce travail doctoral. Ce projet, financé par l'Agence Nationale de la Recherche, associe des politistes (le laboratoire CEPPEL), des informaticiens (laboratoires ERIC, LIA et Xerox) et des linguistes (département recherche et développement d'EDF). L'objectif du projet est d'associer aux analyses automatisées des opinions sur internet (*opinion mining* et *sentiment analysis*) une approche sociologique (caractérisation sociologique des sources, comparaison de la tonalité des tweets à des données de sondage, etc.). On peut également citer les projets ANR « Politique des algorithmes »²⁰⁶ « Webfluence »²⁰⁷ ou encore « enpolitique.com »²⁰⁸

Si ces projets de recherche aboutiront sans doute à terme à une augmentation du nombre de publications relatives à l'étude des opinions numériques, ces approches et ces méthodes restent pour le moment relativement marginales en science politique. À la différence du phénomène observé dans le champ professionnel des études, la numérisation des opinions n'a pas rendu plus poreuses les frontières académiques entre les acteurs traditionnels de l'opinion – la sociologie politique et la science politique – et d'autres disciplines. Au contraire, les autopublications numériques, qui constituent pourtant une nouvelle forme d'expression des opinions, n'ont pas été étudiées par la science politique et la sociologie politique, mais par d'autres disciplines, comme la linguistique, l'informatique ou la sociologie. Plusieurs hypothèses peuvent être évoquées pour expliquer l'herméticité de la sociologie politique à ces nouveaux objets et à ces nouvelles technologies numériques : une relative méconnaissance des chercheurs de ces nouveaux outils informatiques, une absence de formation en informatique, une présence relativement rare d'informaticien au sein des laboratoires de sciences humaines et sociales²⁰⁹, une délégitimation des objets numériques en raison de leur nouveauté et/ou de leur « mainmise » par des disciplines concurrentes, une réticence envers des méthodes et des concepts forgés par d'autres disciplines voire par des acteurs non académiques, etc. Nous n'aurons pas ici les moyens de valider – ou d'invalider –

²⁰⁵ ANR « Imagiweb : Images sur le Web : analyse de la dynamique des images sur le Web 2.0 » (IMAGIWEB - ANR-12-CORD-0002).

²⁰⁶ ANR « Politique des algorithmes » (ALGOPOL - ANR 2012 CORD 01804).

²⁰⁷ ANR « Webfluence » (WEBFLUENCE - ANR-08-SYSC-009).

²⁰⁸ *enpolitique.com* est un projet de recherche international mené par une équipe de chercheurs universitaires québécois et français. Ce projet est financé par l'Agence Nationale de la Recherche (France) et le Fonds de recherche du Québec, Société et culture (FQRSC) pour la période 2011-2014.

²⁰⁹ Dominique Boullier et Audrey Lohard. *Opinion mining et Sentiment analysis : Méthodes et outils, op. cit.*

empiriquement ces hypothèses. Nous nous contenterons simplement de constater que cette herméticité aux nouvelles technologies numériques ne concerne pas l'ensemble des sciences humaines et sociales. D'autres disciplines ont au contraire développé des liens très étroits avec la recherche informatique. Et ce sont ces mêmes disciplines qui ont accordé très tôt un vif intérêt pour les autopublications numériques et, *ipso facto*, pour les opinions numériques.

Quatre disciplines ont développé un lien étroit avec la recherche informatique : la linguistique, avec le TAL (traitement automatique des langues), la sociologie et les sciences de gestion avec la nouvelle science des réseaux (« *new science of networks* ») et le marketing viral, et les sciences de l'information et de la communication.

La linguistique a dès les années 1950, bien avant la création d'internet, développé un rapport étroit avec l'informatique, en créant une sous-discipline dédiée à l'étude informatique du langage²¹⁰. Le traitement automatique des langues (TAL), ou traitement automatique du langage naturel (TALN), se trouve ainsi à la croisée de quatre disciplines : la linguistique, l'informatique, les mathématiques et l'intelligence artificielle. L'avènement des réseaux sociaux et des espaces d'autopublications numériques ont conduit de fait l'analyse automatique du langage à prendre pour objet les opinions numériques. L'*opinion mining* (la détection automatique des opinions sur internet) et le *sentiment analysis* (détection automatique de la tonalité des messages), sont en effet directement issues du TAL.

Les sciences de gestion et la sociologie ont également développé des liens étroits avec les mathématiques et l'informatique autour de « *the new science of network* ». Les réseaux siconumériques ont été très largement étudiés par ces quatre disciplines, soit à partir de cadre théorique propre à chaque discipline, soit à partir d'une approche réellement interdisciplinaire²¹¹.

Enfin, le rapport des sciences de l'information et de la communication à l'informatique est quasiment tautologique, cette discipline étant par essence pluridisciplinaire, à la croisée des sciences sociales et des sciences de l'ingénieur. Les premiers travaux consacrés à Internet remontent ainsi au début des années 1990. La question de l'effet d'internet sur la participation politique et l'interrogation du concept d'espace public ont été les deux principales problématiques structurant les travaux de cette discipline.

²¹⁰ Marcel Cori et Jacqueline Léon, « La constitution du TAL. Étude historique des dénominations et des concepts », *TAL*, Vol. 43 – n° 3/2002, p. 21-55.

²¹¹ Dieudonné Tchuente *et alii*, « Pour une approche interdisciplinaire des TIC. Le cas des réseaux siconumériques », art. cit.

Ces disciplines académiques ont donc étudié, parallèlement ou conjointement aux acteurs du champ professionnel des études numériques, les autopublications – et parfois les opinions numériques – avec leur propre cadre théorique, problématiques, concepts et méthodes.

3.2. Les trois grandes traditions d'étude des autopublications numériques

Parmi l'abondante littérature consacrée à internet et aux autopublications numériques, on peut distinguer sommairement trois grandes approches²¹².

La première, principalement issue des travaux de chercheurs en science de l'information et de la communication et de sociologues, américains notamment, appréhende les autopublications comme des indices – ou des « traces »²¹³ – de participation politique. Ces travaux, parus dès les années 1990, s'inscrivent dans la problématique des effets d'internet sur la participation politique. Les autopublications numériques apparaissent alors comme des éléments de preuve mobilisés dans des travaux empiriques souhaitant valider ou invalider l'hypothèse d'une démocratisation de la participation. Ces recherches, souvent empreintes d'un arrière-plan normatif, privilégient des approches microsociologiques et souvent qualitatives des autopublications numériques (3.2.1).

La seconde approche appréhende les autopublications numériques comme des contenus dont on questionne le potentiel de « viralité ». L'objet étudié est moins le contenu du message que les données chiffrées relatives à sa diffusion (nombre de retweet, nombre de like) et les caractéristiques des réseaux où il se diffuse. Ces travaux se situent à la croisée des théories du marketing viral, des théories sociologiques de l'influence interpersonnelle et de l'analyse mathématique et informatique des réseaux socionumériques. Les autopublications sont ici appréhendées à partir de méthodes quantitatives et font appel à des technologies d'étude spécifiquement numériques (*crawling*, cartographie, etc.) (3.2.2).

Enfin, la troisième approche conçoit les autopublications – et plus particulièrement les opinions numériques – comme des éléments textuels qu'il s'agit de détecter, d'isoler et de

²¹² Nous n'évoquerons pas ici les travaux appréhendant l'autopublication comme une pratique journalistique ou culturelle amateur. Nous nous concentrons sur les publications numériques « ordinaires », c'est-à-dire dissociées d'un projet éditorial.

²¹³ Le terme de « traces » est souvent mobilisé dans les travaux relatifs à internet. Il désigne l'ensemble des manifestations visuelles sur internet.

codifier (en fonction de leur tonalité, de leur thématique, etc.). Les recherches en *opinion mining* et en *sentiment analysis*, longtemps cloisonnées aux frontières du TAL, sont aujourd'hui mobilisées dans des travaux questionnant les liens entre les opinions exprimées sur les réseaux socionumériques et les opinions mesurées par les sondages d'opinion ou les consultations électorales (3).

3.2.1. Les autopublications numériques comme indices de participation politique

La question des effets d'internet sur la participation politique a été au centre des débats académiques des années 1990 et 2000. Comme à l'arrivée de chaque nouvelle technologie de communication, des auteurs ont ainsi questionné les potentialités offertes par l'outil pour « revitaliser la démocratie », « réduire la distance entre élus et citoyens » ou encore « redynamiser la participation politique »²¹⁴. Cette approche techno-déterministe a alors clivé la littérature et opposé « cyber-optimistes »²¹⁵ et « cyber-réalistes »²¹⁶.

Les premiers mobilisent deux principaux arguments pour défendre la thèse selon laquelle internet revitaliserait la démocratie : d'une part, en tant que média, internet diffuserait une information rapide, peu chère et non soumise aux pouvoirs économiques et politiques. D'autre part, en tant que moyen de communication, internet « [brouillerait] *les indices sociaux présents dans toute communication et [participerait] ainsi à l'égalité des échanges en ligne. L'espace public en ligne est ainsi perçu comme une arène de débat et de discussion libérée des contraintes et des inconvénients qui perturbent les échanges en face-à-face* »²¹⁷. Les seconds, sceptiques quant à la capacité d'internet à produire de réels effets sur la participation politique, répondent au premier argument évoqué qu'il existe de réelles inégalités sociales d'accès à internet²¹⁸ et que la consommation d'information y est

²¹⁴ Laurence Monnoyer-Smith, « La participation en ligne, révélateur d'une évolution des pratiques politiques ? », *Participations*, 2011/1 (n° 1), p. 156-185 et Roza Tsagarroussianou, « Electronic democracy in practice... One, two, three... countless variants », *Hermès*, 26-27, 2000, p. 233-246.

²¹⁵ Yochai Benkler, *La Richesse des réseaux. Marchés et libertés à l'heure du partage social*, Lyon, Presses Universitaires de Lyon, 2009.

²¹⁶ Cass Sunstein, *Republic.com*, Princeton, Princeton University Press, 2001.

²¹⁷ Laurence Monnoyer-Smith, « La participation en ligne, révélateur d'une évolution des pratiques politiques ? », art. cit. ; Pierre Levy, *L'intelligence collective. Pour une anthropologie du cyberspace*, Paris, La Découverte, 2004 ; Amanda Mitra, « Marginal voices in cyberspace », *New Media & Society*, 3(1), 2000, p. 29-48.

²¹⁸ Paul Di Maggio, Eszter Hargittai, W. Russel Neuman et John P. Robinson, « Social Implication of the Internet », *Annual Review of Sociology*, 27, 2001, p. 307-336.

socialement déterminée²¹⁹. Les « *cybers réalistes* » s’opposent également à l’argument d’une égalisation des échanges politiques en ligne en soulignant que « *les arènes virtuelles sont tout autant que les autres soumises à des formes d’imposition de pouvoir, accentuées par les inégalités d’accès et/ou de maîtrise de l’outil informatique* »²²⁰.

Dès le milieu des années 1990, des recherches empiriques sont venues compléter et étayer les débats purement théoriques sur les potentialités de l’outil. Ces travaux empiriques ont alors mobilisé des données de différentes natures : statistiques d’usage, études comparées de différentes formes de participation politique *offline* mais aussi analyses qualitatives ou quantitatives d’autopublications numériques.

Ces travaux appréhendent en effet les autopublications comme des indices ou des « traces » de participation politique en ligne. La définition, plus ou moins extensive, donnée par ces recherches à la participation politique détermine la nature des terrains observés (espaces d’autopublication institutionnels, espaces non-institutionnels mais spécifiquement politiques, espaces non spécifiquement politiques) et les catégories d’autopublication considérées comme relevant ou non du domaine de la participation (autopublication, pétition en ligne, retweet, « like », etc.). Deux éléments entrent en compte dans la définition de la participation en ligne : la place plus ou moins grande accordée à la dimension expressive et symbolique des comportements politiques²²¹ et le degré de spécificité de la participation en ligne par rapport à la participation hors-ligne²²².

Le premier élément oppose l’approche « classique » de la participation politique proposée par Almond et Verba – définissant la participation comme l’ensemble des activités menées en vue d’influencer directement ou indirectement l’action gouvernementale – à des définitions plus extensives, revalorisant « *la dimension expressive et symbolique des comportements* »²²³, telles que les discussions politiques ordinaires par exemple. Des travaux adoptant une définition restrictive de la participation auront alors davantage tendance à choisir des terrains numériques spécifiquement politiques – comme des forums mis en place par des

²¹⁹ Viviane Le Hay, Thierry Vedel et Flora Chanvriil, « Usages des médias et politique : une écologie des pratiques informationnelles », *Réseaux*, n° 170, octobre-novembre 2011, p. 45-74 ; Fabien Granjon et Aurélien Le Foulgoc, « Les usages sociaux de l’actualité », *Réseaux*, 2010/2, n° 160-161, p. 225-253.

²²⁰ Laurence Monnoyer-Smith, « La participation en ligne, révélateur d’une évolution des pratiques politiques ? », art. cit.

²²¹ Nona Mayer, *Sociologie des comportements politiques*, Paris, Armand Colin, 2010.

²²² Gersende Blanchard, Simon Gadras et Stéphanie Wojcik, « Analyser la participation politique en ligne : des traces numériques aux pratiques sociales » in Christine Barats (dir.), *Manuel d’analyse du Web*, Paris, Armand Colin, 2012.

²²³ Nona Mayer, *Sociologie des comportements politiques*, op. cit.

sites de média généralistes ou par des partis politiques²²⁴ – voire institutionnels, comme les arènes de délibération en ligne mises en place par la *Commission Nationale du Débat Public* (CNDP) ou d'autres institutions publiques²²⁵. *A contrario*, des travaux privilégiant une définition plus extensive de la participation en ligne pourront étudier des espaces où les discussions politiques ne se déroulent que de manière ponctuelle, comme des forums thématiques non politiques, des blogs non spécifiquement politiques ou des réseaux sociaux numériques²²⁶. Le caractère plus ou moins extensif de la définition adoptée par ces travaux détermine également les catégories d'autopublication considérées comme relevant ou non de la participation politique. Ainsi, pour Schlozman *et alii*, les discussions politiques échangées entre amis sur Facebook ne peuvent être considérées comme des formes de participation, dans la mesure où il n'existerait pas d'efforts politiques organisés en vue d'influencer les autorités publiques²²⁷. De la même manière, Morozov qualifie de « *slacktivism* » l'activité de publication de messages politiques sur les réseaux sociaux numériques, signifiant par-là que cette activité n'aurait aucune conséquence politique réelle et ne servirait en réalité qu'à accroître le sentiment d'utilité des participants²²⁸. En revanche, pour Dominique Cardon, « *émettre un vote sur article ou une vidéo, appuyer sur le bouton « I like » de Facebook [...] rédiger un lien sur Twitter constituent des formes de participation, même minimales* »²²⁹. Pour l'auteur, cette participation qualifiée de « *presse-bouton* » permettrait de réduire les inégalités sociales de participation politique sur internet en élargissant la notion de participation à des formes d'expression beaucoup moins exigeantes socialement et culturellement.

Le second élément entrant en ligne de compte dans la définition de la participation en ligne tient à la spécificité de la participation en ligne par rapport à la participation « classique », hors-ligne. Ce second élément oppose des approches appréhendant les pratiques numériques de participation comme de simples transpositions ou des extensions de celles

²²⁴ Nicolas Desquinabo, « Interactions et argumentations dans les webforums partisans », *Communication au colloque Les usages partisans de l'Internet organisé par l'équipe IRÉNÉE – Université de Nancy 2*, Nancy, 21-22 juin 2007.

²²⁵ Nicolas Benvegno et Mathieu Brugidou, « Prendre la parole sur Internet. Des dispositifs sociotechniques aux grammaires de la discussion », *Réseaux*, 2008/4 (n° 150), p. 51-82

²²⁶ Fabienne Greffet et Stéphanie Wojcik, « Parler politique en ligne. Une revue des travaux français et anglo-saxons », art. cit. ; Scott Wright, « From “third place” to “third space”: everyday political in non-political online spaces », *Javnost- the public*, Vol. 19, n°3, 2012, p. 5-50.

²²⁷ Kay L. Schlozman, Sydney Verba et Henry E. Brady, « Weapon of the Strong ? Participatory Inequality and the Internet », *Perspectives on Politics*, 8(2), 2010, p. 487-459.

²²⁸ Evgeny Morozov, « The brave new world of slacktivism », cité par Gersende Blanchard, Simon Gadras et Stéphanie Wojcik, « Analyser la participation politique en ligne : des traces numériques aux pratiques sociales » in Christine Barats (dir.), *Manuel d'analyse du Web*, op. cit.

²²⁹ Dominique Cardon, *La démocratie internet*, Paris, Éditions du Seuil, 2010, p. 80.

observées hors-ligne²³⁰ à des approches considérant que les notions ou concepts issus du champ de la participation classiques ne permettent pas d'appréhender finement la participation en ligne et ses transformations²³¹.

La comparaison des interactions politiques en ligne et hors ligne et les spécificités de ce que serait un « espace public numérique » sont ainsi les deux questions qui ont animé le débat académique des années 1990 et 2000. Fabienne Greffet et Stéphanie Wojcik, dans leur revue de la littérature des travaux étudiant les discussions politiques en ligne distinguent deux périodes : une première marquée par la prégnance du cadre théorique – et souvent normatif – de la délibération, inspirée des travaux d'Habermas, précédant une seconde période caractérisée par l'abandon progressif du cadre délibératif et d'une ouverture à d'autres approches²³².

Les premières recherches sur les discussions politiques en ligne, parues dans les années 1990, s'inscrivaient ainsi dans le cadre théorique habermassien de la délibération. Ces travaux comparaient les caractéristiques des discussions politiques observées sur internet avec celles d'un espace public idéal « *pour arriver à la conclusion jugée décevante que les NTIC et internet en particulier ne permettraient pas la réalisation de l'idéal délibératif* »²³³. Ces recherches, pour parvenir à ce résultat, prennent pour objet des messages échangés le plus souvent sur des forums thématiques consacrés à des sujets politiques ou sociétaux. Il s'agit alors d'évaluer les discussions observées à l'aune de différents critères d'évaluation, ces critères pouvant être de nature quantitative²³⁴ ou qualitative²³⁵. Michaël Dumoulin, afin d'évaluer la nature délibérative de discussions observées sur trois forums québécois²³⁶ et ainsi vérifier, en suivant la définition de Fishkin, s'il existe « *une réflexion et une constante remise*

²³⁰ Christine Barats, *Manuel d'analyse du Web*, op. cit.

²³¹ Andrew Chadwick, « Recent Shifts in the Relationship Between the Internet and Democratic Engagement in Britain and the United States: Granularity, Informational Exuberance, and Political Learning », in Michael Jensen, Laia Jobra (dir.), *Digital Media and Political Engagement Worldwide: A Comparative Study*, Cambridge, Cambridge University Press, 2012.

²³² Fabienne Greffet et Stéphanie Wojcik, « Parler politique en ligne. Une revue des travaux français et anglo-saxons », art. cit.

²³³ *Ibidem*.

²³⁴ Michaël Dumoulin, « Les forums électroniques : délibératifs et démocratiques ? », in Denis Monière (dir.), *Internet et la Démocratie*, Québec, Monière et Wollank éditeurs, 2002, p. 141-157 ; William H. Dutton « Network rules of order: regulating speech in public electronic fora », *Media, Culture & Society*, vol. 18, n° 2, 1996, p. 269-290 ; Steven M. Schneider, « Creating a Democratic Public Sphere Through Political Discussion: A Case Study of Abortion Conversation on the Internet », *Social Science Computer Review*, vol. 14, n° 4, 1996, p. 373-393.

²³⁵ Tamara Witschge, « Online Deliberation: Possibilities of the Internet for Deliberative Democracy », Peter M. Shane (dir.), *Democracy Online. The Prospects for Political Renewal through the Internet*, New York, Routledge, 2004, p. 109-122.

²³⁶ Deux traitants de politique intérieure au Canada et au Québec, un sur les droits des minorités sexuelles.

en question d'opinions personnelles confrontées à d'autres arguments », a recensé de façon exhaustive 33 009 messages publiés sur ces forums, puis a codé manuellement un échantillon de 201 messages par forum, à partir d'une grille d'évaluation comprenant six critères qualifiant le registre du discours et la fonction des messages dans la discussion (« introduit », « réponse », « sollicite », « affiliation », « impertinence », « négativisme »). À partir de ces données chiffrées, l'auteur conclut que « les discussions, même lorsqu'elles sont inscrites en réponse à d'autres messages, ne font que réitérer ou reformuler leurs opinions. Le contenu représente donc des monologues interactifs du type " argument - contre argument " où les participants recherchent avant tout à s'exprimer, souvent de façon injurieuse et arrogante, en rejetant fréquemment, presque systématiquement, les points de vue des autres ». Si l'auteur mobilise bien ici la notion d' « opinion », il ne s'agit pas de s'interroger sur les modes de production des opinions, ni même sur la relation entre leur contenu et la position sociale de leurs auteurs. Les opinions sont ici questionnées selon la problématique de leur ajustement ou au contraire de leur invariance en situation de débats contradictoires. Conformément à la théorie délibérative, les opinions sont donc davantage appréhendées comme les produits d'un processus conversationnel que comme des attributs individuels. En constatant que sur internet, les échanges des internautes ressemblent davantage à des « monologues interactifs », cet auteur conclut avec regret que les forums électroniques ne seraient donc pas des lieux délibératifs. D'autres auteurs sont arrivés au même constat en tenant compte d'autres critères d'évaluation. Steven Schneider identifie ainsi quatre critères – égalité, diversité, réciprocité, qualité des échanges – tandis que Jennifer Stromer-Galley en retient huit : expression d'opinions raisonnées, référence à des sources extérieures, absence d'attaques *ad hominem*, possibilité d'avoir des opinions ou des perspectives opposées, égalité de participation, réflexivité et narrativité²³⁷. La méthodologie est alors la même : à partir d'un échantillon de messages et d'une grille d'évaluation, le chercheur code manuellement les autopublications puis en tire des éléments statistiques qui valident ou invalident empiriquement le caractère délibératif de la discussion. Tamara Witschge a quant à elle remis en cause ces mesures quantifiées de la délibération qui réduiraient selon elle la portée heuristique du matériau étudié et plaide au contraire pour des analyses qualitatives des discussions politiques, afin de mieux appréhender la complexité des débats²³⁸. Fabienne Greffet et Stéphanie Wojcik ont ensuite démontré comment dès la fin des années 1990 le modèle délibératif a été remis en

²³⁷ Jennifer Stromer-Galley, « Decoding deliberation online », Paper presented at the *Second Conference on Online Deliberation*, Palo Alto, California, mai 2005.

²³⁸ Tamara Witschge, « Online Deliberation: Possibilities of the Internet for Deliberative Democracy », art. cit.

cause, notamment par des auteures féministes soulignant l'inégale capacité des groupes à peser sur la délibération²³⁹ et reprochant au modèle délibératif de délégitimer les formes d'expression ouvertement conflictuelles.

La seconde « ère » des travaux sur la discussion politique en ligne est ainsi marquée par l'abandon progressif du cadre théorique délibératif, une diversification des approches et une plus grande place accordée aux spécificités de l'interaction en ligne²⁴⁰. Le cadre théorique développé par Goffman est ainsi souvent mobilisé dans les travaux étudiant les interactions sociales en ligne. Ces travaux se sont davantage interrogés sur les spécificités des interactions en ligne et leurs effets – l'anonymat, la mise à distance physique des participants, le manque d'indices sociaux sur les participants, la modération *a priori* ou *a posteriori* – pour en arriver à des conclusions souvent contradictoires. Ainsi, pour William Dutton ou Sara Kiesler, l'anonymat et l'absence de marqueurs sociaux dans les discussions en ligne permettraient d'exprimer ses opinions plus librement et favoriseraient ainsi une participation plus large que dans les interactions menées en face-à-face. Pour d'autres auteurs, ces éléments provoqueraient une plus grande charge conflictuelle, l'anonymat encourageant des comportements exutoires ou fantasmatiques²⁴¹. *A contrario*, ces mêmes spécificités de l'échange en ligne – anonymat, mise à distance, etc. – pourraient avoir pour conséquence de décourager l'expression de points de vue minoritaires ou dissidents et rendre plus homogène la discussion en ligne²⁴². Bien qu'ils arrivent à des conclusions contradictoires, ces nouveaux travaux sur la discussion politique ont pour points communs de plaider d'une part pour une prise en considération d'autres registres légitimes d'intervention dans l'espace public comme le témoignage, la narration, ou le récit personnel et de privilégier d'autre part une définition plus extensive de la politique²⁴³.

Cette première tradition d'étude des autopublications numériques conçoit donc les messages publiés sur internet comme des indices de participation politique, ou, de façon plus restrictive, comme des opinions échangées dans le cadre de discussions politiques obéissant à certaines règles qu'il s'agit d'objectiver. Ces travaux appréhendent donc les autopublications et les opinions numériques à partir d'un cadre théorique délibératif ou interactionniste. Dans

²³⁹ Fabienne Greffet et Stéphanie Wojcik, « Parler politique en ligne. Une revue des travaux français et anglo-saxons », art. cit.

²⁴⁰ *Ibidem*.

²⁴¹ Sherry Turkle, *Life on the Screen. Identity in the Age of Internet*, New York, Simon & Shuster, 1995.

²⁴² Fabienne Greffet et Stéphanie Wojcik, « Parler politique en ligne. Une revue des travaux français et anglo-saxons », art. cit.

²⁴³ *Ibidem*.

ces recherches, les opinions étudiées sur internet ne sont pas comparées à des opinions recueillies par d'autres technologies d'étude. Les méthodologies employées ne sont pas standardisées et les terrains étudiés relativement localisés. On ne s'interroge ni sur le mode de production des opinions, ni sur leur mode de diffusion.

Les mécanismes de diffusion des autopublications ont cependant été étudiés par d'autres courants de recherche.

3.2.2 Autopublications, « viralité », réseaux sociaux et influence

Une seconde tradition d'étude des autopublications numériques – issue d'horizons disciplinaires, théoriques et méthodologiques totalement différents des travaux sur la participation politique en ligne – appréhende les messages publiés sur internet à partir d'une toute autre perspective de recherche. Il ne s'agit pas de savoir ce qu'internet « fait » à un phénomène social – la participation politique – mais d'objectiver ce qu'internet peut nous apprendre des mécanismes de diffusion et de circulation des messages et des opinions au sein de réseaux sociaux, numériques et non-numériques. Cette seconde tradition d'étude, que Kévin Mellet qualifie de « marketing viral », réunit trois approches disciplinaires : la sociologie des relations interpersonnelles, le « *word-of-mouth* » marketing et l'analyse mathématique et statistique des réseaux sociaux²⁴⁴.

3.2.2.1. Les origines du « marketing viral »

La sociologie des relations interpersonnelles, initiée par les travaux de Paul Lazarsfeld et du *BASR* dans les années 1940 aux États-Unis a très largement « influencé » le « *word-of-mouth* » marketing – qui n'est rien d'autre, selon Katz, que l'application du *two step flow model* au marketing²⁴⁵ – et, plus récemment, le marketing viral. La quatrième de couverture de la traduction française de *Personal Influence* note d'ailleurs que « *le buzz sur Internet ou*

²⁴⁴ Nous mobiliserons à de nombreuses reprises dans ce paragraphe l'article de Kévin Mellet présentant un état de l'art assez complet de cette tradition d'étude : Kévin Mellet, « Aux sources du marketing viral », *Réseaux*, 2009/5-6 (n° 157-158), p. 267-292.

²⁴⁵ Kévin Mellet, « Aux sources du marketing viral », art. cit.

bien encore le marketing viral ont remis au goût du jour »²⁴⁶ les conclusions de l'ouvrage de Lazarsfeld et Katz.

Les recherches successives de l'école de Columbia, de *The People's Choice* jusqu'à *Medical Innovation*, n'ont eu de cesse d'adapter la méthodologie de l'enquête par sondage – légitimée et importée dans le champ académique par Lazarsfeld²⁴⁷ – à l'étude des influences interpersonnelles, dont l'importance dans la décision de vote n'était pas un résultat attendu de *The People's Choice*²⁴⁸. La problématique initiale de l'étude n'était pas de saisir le rôle des influences interpersonnelles dans le processus de décision de vote, mais de déterminer dans quelle mesure les campagnes électorales médiatisées influençaient les choix électoraux des individus. Parmi le questionnaire administré au panel d'électeurs, seules deux questions faisaient référence à la notion d'influence interpersonnelle²⁴⁹. Près de dix ans après la publication de l'ouvrage, Katz lui-même a reconnu que le dispositif d'enquête de *The People's Choice* ne permettait pas d'affirmer l'existence du *leadership* d'opinion. Comme le résume bien Michel Grumbach et Nicolas Herpin : « dans leur recherche permanente d'une méthodologie adéquate à leur projet théorique, les lazarsfeldiens n'ont cessé d'être confrontés aux canons de l'enquête par sondage qu'ils avaient eux-mêmes contribué à rationaliser et à imposer. Le principal obstacle qu'oppose la philosophie de la survey research à la saisie d'un objet dont l'essence relationnelle fait toute la réalité, c'est l'individualisme atomiste qui guide la totalité de ses actes. Il inspire, en effet, tant les choix d'échantillonnage – l'échantillon aléatoire et représentatif d'individus – que la sélection des données bonnes à recueillir, et donc celle des variables pertinentes pour l'interprétation – les propriétés (structurales) « individuelles » plutôt que relationnelles – ou encore la constitution des unités d'analyse – les classes d'individus plutôt que les classes de relations »²⁵⁰.

Ainsi, *Voting*, adopte la même méthodologie que *The People's Choice* (échantillon aléatoire et représentatif d'individus et *leaders* auto-désignés), mais interroge cette fois systématiquement les enquêtés sur les caractéristiques et les opinions des membres de leur environnement. Il est alors possible de mettre en relation les comportements des enquêtés

²⁴⁶ Elihu Katz et Paul Lazarsfeld, *Influence personnelle*, op. cit.

²⁴⁷ Loïc Blondiaux, *La fabrique de l'opinion*, op. cit.

²⁴⁸ Elihu Katz, « The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis Author », *The Public Opinion Quarterly*, Vol. 21, n°1, 1957, p. 61-78. ; Michel Grumbach et Nicolas Herpin, « À propos de quelques travaux de Lazarsfeld et de son école », *Enquête* [En ligne], 4 | 1988, mis en ligne le 27 juin 2013, consulté le 15 octobre 2013. URL : <http://enquete.revues.org/65>.

²⁴⁹ Paul F. Lazarsfeld, Bernard R. Berelson et Hazel Gaudet, *The people's choice: How the voter makes up his mind in a presidential campaign*, New York, Duell, Sloan, and Pearce, 1944, p. 50.

²⁵⁰ Michel Grumbach et Nicolas Herpin, « À propos de quelques travaux de Lazarsfeld et de son école », art. cit.

avec ceux de leur entourage²⁵¹. Avec *Personal Influence*, l'échantillonnage du questionnaire se fait en « boule de neige » : on demande aux individus (femmes) d'administrer elles-mêmes le questionnaire à leur entourage proche : « *la démarche permet à la fois d'identifier les leaders d'opinion sur la base d'une relation réelle et confronter les paires de partenaires liés par des transactions effectives* »²⁵². Enfin, *Médical Innovation* remplace les différents types d'échantillonnage par un recensement de quelques groupes réels et détermine les leaders à partir d'une cartographie complète de la structure sociométrique de la communauté professionnelle²⁵³.

L'ouvrage de Coleman, Katz et Menzel paru en 1966 constitue une référence de la sociologie des réseaux. L'objectif de l'enquête est d'étudier les mécanismes de diffusion d'une innovation médicale auprès de l'ensemble des médecins de quatre villes américaines. Les auteurs identifient deux catégories de médecins : les innovateurs et les retardataires. Les premiers sont plus souvent choisis par leurs pairs comme informateurs ou conseillers. Mais l'influence des innovateurs n'est pas systématique et dépend d'une autre variable : leur degré d'intégration sociale. À partir d'un sociogramme, les auteurs observent que deux catégories de médecins se distinguent : les premiers, qualifiés d'intégrés, ont des interactions fréquentes avec leurs pairs, les seconds sont au contraire relativement isolés : « *finalement, Coleman et alii révèlent deux courbes d'adoption différentes. La première, celle des médecins intégrés, est exponentielle et prend la forme d'une courbe en S. La seconde, celle des médecins isolés, est linéaire et s'étend sur une temporalité bien plus longue* »²⁵⁴. L'étude conclut donc que la structure du réseau détermine la circulation de l'information, « *il est alors possible de déterminer le nombre d'adopteurs initiaux nécessaires pour garantir l'auto-entretien de la diffusion* ». Cette conclusion, ainsi que bien des résultats de la sociologie des relations interpersonnelles ont intéressé au premier plan le marketing, et, nous le verrons plus loin, la nouvelle science des réseaux.

Les travaux de l'école de Columbia ont en effet joué un rôle important dans le développement des sciences du marketing²⁵⁵. À la suite des travaux du *BASR*, des recherches en marketing ont en effet entrepris d'une part de tester empiriquement les thèses de Katz et

²⁵¹ Bernard Berelson, Paul Lazarsfeld et William McPhee, *Voting. A study of Opinion Formation in a Presidential Campaign*, Chicago, The University of Chicago Press, 1954.

²⁵² Elihu Katz et Paul Lazarsfeld, *Influence personnelle*, *op. cit.*

²⁵³ James Coleman, Elihu Katz et Herbert Menzel, *Medical Innovation. A Diffusion Study*, The Bobbs-Merrill Co, 1966.

²⁵⁴ Kévin Mellet, « Aux sources du marketing viral », art. cit.

²⁵⁵ Frank Cochoy, *Une Histoire du marketing. Discipliner l'économie de marché*, Paris, La Découverte, 1999.

Lazarsfeld à partir de différents produits de consommation et de définir d'autre part les moyens les plus efficaces pour identifier et associer les leaders d'opinion aux stratégies de promotion des firmes²⁵⁶. Robert Brooks démontre ainsi que la stratégie commerciale de la firme doit tenir compte de différents paramètres, comme le degré d'intégration sociale des consommateurs, le type de réseau dans lesquels ils sont intégrés et le type de produit vendu. Dans certaines configurations, lorsque pour certains produits le bouche-à-oreille ne fonctionne pas, les firmes doivent alors privilégier le recours à la publicité traditionnelle. Au contraire, quand les réseaux interpersonnels jouent un rôle déterminant, l'auteur recommande de cibler les leaders d'opinion et de les tenir informés de façon personnalisée²⁵⁷.

Si la thèse du rôle déterminant de l'influence interpersonnelle dans la décision d'achat a depuis fait relativement consensus dans les sciences du marketing, la définition des leaders d'opinion est plus problématique²⁵⁸. Dans les travaux de Lazarsfeld, Katz ou Coleman, les leaders n'influencent qu'une poignée de personnes (quatre ou cinq tout au plus), toutes issues de leur entourage immédiat. Des travaux de marketing, oubliant ou méconnaissant la dimension contextuelle et microsociologique de l'influence, ont attribué à certains individus le rôle de « *super influentials* », à savoir des individus avant-gardistes capables, grâce à l'étendue de leurs réseaux de relations, de propulser à eux seuls une marque ou un produit²⁵⁹. Des auteurs se sont alors opposés à cette conception de l'influence, avançant l'idée que la taille des réseaux dans lesquels évoluent les consommateurs n'a pas une incidence directe sur la propension à échanger des recommandations avec autrui²⁶⁰. Dit autrement, avoir beaucoup d'amis ne signifie pas que l'on est plus influent. D'autres auteurs affirment de plus que le leadership d'opinion n'est pas unilatéral, les leaders d'opinion pouvant aussi être des demandeurs d'informations²⁶¹. Garcia-Bardidia conclut ainsi que « *d'un leadership unidimensionnel, unilatéral et conscient, les recherches se sont progressivement focalisées sur la transmission d'informations* »²⁶².

²⁵⁶ Kévin Mellet, « Aux sources du marketing viral », art. cit.

²⁵⁷ Rod Brooks, « Word-of-Mouth in Selling New Products », *Journal of Marketing*, 22 (2), 1957, p. 154-161.

²⁵⁸ Kévin Mellet, « Aux sources du marketing viral », art. cit.

²⁵⁹ Malcolm Gladwell, *The Tipping Point: How little things can make a big difference*, New York, Little, Brown and Company, 2002 ; Ed Keller et Jon Berry, *The Influentials: one American in ten tells the other nine how to vote, where to eat and what to buy*, New York, Free Press, 2003.

²⁶⁰ Ted Smith, James R. Coyle, Elizabeth Lightfoot et Amy Scott, « Reconsidering models of influence: the relationship between consumer social networks and word-of-mouth effectiveness », *Journal of Advertising Research*, décembre 2007, p. 387-397.

²⁶¹ Renaud Garcia-Bardidia, « Des communications interpersonnelles au marketing viral: état des lieux », Document de travail, 2006.

²⁶² *Ibidem*.

La controverse sur la définition des leaders d'opinion et la question de la circulation de l'information a également été au centre d'une autre discipline, « la nouvelle science des réseaux »²⁶³.

En rendant accessibles et modélisables de très grandes bases de données, Internet a donné une nouvelle actualité à l'analyse statistique et mathématique des réseaux sociaux. La « nouvelle » discipline a dans un premier temps testé les thèses « classiques » de la sociologie des réseaux. Duncan Watts et Steven Strogatz ont ainsi cherché à expérimenter la thèse des « petits mondes » de Stanley Milgram à partir des liens hypertextes contenus dans les pages internet²⁶⁴. Les deux chercheurs ont alors observé que sur internet également, un faible nombre de liens permettait de relier deux nœuds choisis aléatoirement au sein du graphe. Peu de temps après la parution de cette recherche, une première grande controverse opposa, par articles interposés, Barabási et Albert à Huberman et Adamic. L'article de Barabási et Albert, paru en 1999 dans la revue *Science*²⁶⁵, montre que la distribution des liens hypertextes sur internet obéit à une loi de puissance : un petit nombre de sites concentre un très grand nombre de liens et inversement, un très grand nombre de sites ne concentre qu'un très faible nombre de liens. Les deux auteurs expliquent alors ce phénomène par le concept d'« *attachement préférentiel* » : les nouveaux sites internet préféreraient se « connecter » à des sites possédant déjà beaucoup de liens plutôt qu'à des sites en possédant peu. L'article de Huberman et Adamic, paru la même année dans la revue *Nature*²⁶⁶, observe le même phénomène d'hyperconcentration mais explique cette loi de puissance par une hypothèse différente de celle de Barabási et Albert. Chaque site aurait un taux de croissance intrinsèquement différent, et ces différences intrinsèques s'expliqueraient par différentes variables comme le niveau de la qualité et d'intérêt du site ou encore par le niveau d'investissement financier et marketing du site nouvellement créé. Les deux articles, même s'ils n'avancent pas les mêmes hypothèses, ont néanmoins contribué à affirmer le caractère profondément inégalitaire et concentré des réseaux sur internet.

Le caractère inégalitaire et concentré des réseaux numériques a précisément constitué une source de légitimation scientifique à la thèse des « *super influentials* » défendue en

²⁶³ Duncan Watts, « The “new” science of networks », *Annual Review of Sociology*, n° 30, 2004, p. 243-70.

²⁶⁴ Duncan Watts et Steven Strogatz, « Collective Dynamics of “Small-World” Networks », *Nature*, n° 393, 1998, p. 440-442.

²⁶⁵ Albert-Lászlo Barabási et Réka Albert, « Emergence of Scaling in Random Networks », *Science*, n° 286, 1999, p. 509-512.

²⁶⁶ Bernardo Huberman et Lada Adamic, « Growth Dynamics of the Word Wide Web », *Nature*, vol. 401, n° 6749.

marketing par Gladwell : « *s'il est possible de cibler les quelques individus qui sont très connectés, alors la diffusion sera extrêmement rapide, alors que des individus choisis aléatoirement provoqueront une diffusion lente* »²⁶⁷. La controverse sur les leaders d'opinion observée en science du marketing s'est également diffusée dans les travaux de la nouvelle science des réseaux. En opposition à Barabási et Gladwell, Watts et Dodds ont ainsi démontré que la diffusion d'un contenu dans un réseau social dépend moins des caractéristiques individuelles des leaders d'opinion ou des « influenceurs » que de la structure même du réseau²⁶⁸. Pour ce faire, les auteurs procèdent par modélisation mathématique en testant différentes hypothèses à partir de différents modèles de graphes. Pour chaque modèle, ils étudient la capacité d'un individu à générer de larges « cascades d'adoption », c'est-à-dire à diffuser un contenu à de nombreux autres individus. L'analyse statistique de ces modélisations produit deux résultats :

*« D'une part, Watts et Dodds montrent que l'ampleur des cascades dépend très fortement de la densité du réseau dessiné : elle ne se produit que si le réseau n'est ni trop lâche, ni trop dense. Si le réseau est éparé, il est plus facile pour un individu d'influencer ses voisins, car ceux-ci sont soumis à des influences moins variées, donc plus efficaces ; mais il est plus difficile d'atteindre un grand nombre d'individus, et les cascades peuvent s'essouffler facilement. À l'inverse, si le réseau est dense, l'innovateur peut toucher facilement un grand nombre d'individus ; mais les individus sont plus difficiles à convaincre, car l'influenceur n'est ici qu'un voisin parmi de nombreux autres. D'autre part, les auteurs montrent que les individus les plus influents ne le sont que très modérément par rapport aux autres. Quel que soit le réseau dessiné, l'action des individus les plus influents, c'est-à-dire des individus les mieux connectés, n'est que rarement supérieure de manière significative à celle des autres »*²⁶⁹.

En contestant le rôle des leaders d'opinion dans les mécanismes d'influence, Watts et Dodds rejoignent ainsi des auteurs en science du marketing qui s'étaient, quelques années auparavant, opposés aux thèses des « *super influentials* ».

²⁶⁷ Kévin Mellet, « Aux sources du marketing viral », art. cit.

²⁶⁸ Duncan Watts et Peter Sheridan Dodds, « Influentials, Networks and Public Opinion Formation », *Journal of Consumer Research*, 34(4), 2007, p. 441-458.

²⁶⁹ Thomas Beauvisage, Jean-Samuel Beuscart, Thomas Couronné et Kevin Mellet, « Le succès sur Internet repose-t-il sur la contagion ? Une analyse des recherches sur la viralité », *Tracés*, 21/2011, p. 151-166.

3.2.2.2 Les trois approches de la contagion

Le « marketing viral » se situe donc à l'interstice de la sociologie des relations interpersonnelles, du « *word-of-mouth* » marketing et de la « nouvelle science des réseaux ». Les travaux de marketing viral prennent pour objet les mécanismes de « viralité » ou de diffusion des contenus sur internet et les réseaux socionumériques. Thomas Beauvisage *et alii* recensent dans leur revue de la littérature trois principales approches en marketing viral : une première appréhende la viralité comme un phénomène de concentration temporelle de l'attention des internautes sur un contenu, la deuxième cherche à dissocier les logiques de diffusion interpersonnelle, purement « virale », des dispositifs éditoriaux mis en place par les sites pour orienter l'attention des internautes, et la troisième part des caractéristiques du réseau social et des « leaders d'opinion » pour étudier comment s'y propage un contenu. Cette troisième approche renvoie donc directement au débat sur les « *influenceurs* » évoqué à l'instant.

La première approche étudie la dimension temporelle de la diffusion de contenus et d'information sur internet : « *dans ces travaux, la « viralité » est définie de façon large, comme un phénomène de concentration temporelle de l'attention collective des internautes sur certains contenus. L'enjeu est avant tout descriptif : il s'agit de mettre en évidence des régularités dans la dynamique de l'audience sur le Web, des formes typiques et répétées de focalisation de l'attention collective sur différents objets* »²⁷⁰. Huberman et Szabo ont ainsi analysé le cycle de vie de contenus publiés sur deux réseaux sociaux, *Digg* et *Youtube*. Les auteurs constatent que les dynamiques d'audience ne sont pas les mêmes sur les deux sites. Sur le premier, l'audience connaît vite une saturation et augmente peu par la suite tandis que sur *Youtube* l'audience des vidéos se construit dans la durée. Si ces connaissances statistiques sur les trajectoires d'audience permettent de prévoir la popularité et la durée de vie d'un contenu, elles n'expliquent pas en revanche les causes des succès, et ne dissocient donc pas les différents facteurs qui peuvent en être à l'origine : bouche-à-oreille, inscription à l'agenda de grands médias, systèmes de recommandations ou compteurs mis en place par les plateformes d'autopublication, etc.

Une autre approche cherche précisément à expliquer de manière plus fine les raisons du succès d'audience d'un contenu sur internet. L'enjeu est donc de dissocier ce qui relève

²⁷⁰ *Ibidem.*

directement de la communication interindividuelle (de la « viralité ») d'une part et des dispositifs éditoriaux mis en place par les sites pour orienter l'attention des internautes (recommandations, compteurs, etc.) d'autre part. Watts et Salganik ont dans ce but mis en place un dispositif expérimental de téléchargement de musique en ligne pour mesurer l'importance des systèmes d'affichage d'information sur la popularité des titres proposés aux usagers/enquêtés du site. Les 12 207 enquêtés sont divisés en trois groupes. Chaque groupe est invité à choisir des titres parmi 50 morceaux de musique inconnus. Au premier groupe, on affiche le nombre réel de téléchargement du titre, au second groupe on affiche l'indicateur de téléchargement inversé (le titre le moins téléchargé apparaît comme le titre le plus téléchargé), enfin au troisième groupe, aucune indication n'est communiquée. Les résultats de l'étude montre que le succès d'un titre est en grande partie liés aux informations fournies sur sa popularité : en dehors des chansons les mieux classées, le succès d'un titre dépend des indications (vraies ou fausses) fournies sur son passé. Cette expérimentation, stimulante méthodologiquement, permet ainsi de démontrer empiriquement l'importance des dispositifs éditoriaux de cadrage de l'attention des internautes dans le processus de diffusion des contenus sur internet.

Enfin, une troisième approche consiste « à partir du réseau social et à examiner comment s'y propagent les contenus et les comportements. [...] La focale se déplace, depuis la trajectoire du contenu vers le mécanisme de transmission lui-même ; on s'efforce de comprendre ce qui fait, dans la structure du réseau et dans les caractéristiques des individus, qu'un contenu va être ou non transmis entre les points du réseau »²⁷¹. Parmi ces travaux, certains s'intéressent davantage à identifier les leaders d'opinion et à « mesurer » leur influence. D'autres privilégient au contraire l'analyse structurale du réseau pour déterminer les facteurs qui favorisent la circulation d'un contenu. On retrouve donc ici la controverse théorique sur le rôle des « influenceurs » rencontrée en science du marketing et dans la nouvelle science des réseaux.

Les travaux académiques sur les mesures d'influence en ligne ont pris un nouvel essor avec l'émergence du réseau socionumérique Twitter. L'architecture même du site, qui repose sur une hiérarchisation des messages publiés (tweets) en fonction de leur nombre de reprises par d'autres utilisateurs (retweets) est un élément d'explication de cet attrait de Twitter

²⁷¹ Thomas Beauvisage *et alii*, « Le succès sur Internet repose-t-il sur la contagion ? Une analyse des recherches sur la viralité », art. cit.

comme « *laboratoire d'étude des processus de diffusion* »²⁷². Twitter matérialise en effet des données qui peuvent s'apparenter, selon les définitions et concepts mobilisés, à des indices de « diffusion », de « visibilité », d'« audience », d'« autorité » ou d'« influence » : nombre d'abonnés, nombre de tweets publiés, nombre de retweets, etc. De plus, Twitter rend facilement accessible l'intégralité de ces données, *via* son *API* payante²⁷³. L'accès à ces données et l'usage de logiciels de *crawling* permettent donc aux chercheurs de travailler sur des corpus exhaustifs, dépassant souvent les millions de messages.

Si la majorité de ces travaux se réfèrent aux travaux de Lazarsfeld, Katz ou Coleman, la définition donnée à la notion d'influence est relativement lâche et ne rend compte ni du caractère contextuel et socialement situé de l'influence, ni de son aspect microsociologique – tout au contraire, ces travaux valorisent des approches quantitatives à partir de corpus exhaustifs de millions de données. L'enjeu est de savoir qui est influent sur Twitter et quelles sont les données qui mesurent au mieux cette influence. Ces travaux reposent donc sur une démarche totalement empiriste. Par analogie à la définition tautologique de Gallup, si l'opinion publique est ce qui est mesurée par les sondages, l'influence serait, selon ces travaux, ce qui est mesurée par Twitter. Le rôle du chercheur est donc d'identifier quelles données ou algorithmes rendent le mieux compte de cette influence : pour certains il s'agira du nombre de *followers*²⁷⁴, pour d'autres du nombre de retweets²⁷⁵, etc. Ainsi, pour Dominique Boullier et Audrey Lohard « *plutôt que de se demander ce qu'est l'influence, question ontologique qui ne débouche pas empiriquement, il est plus efficace de se demander, dans une visée pragmatiste, ce que c'est qu'influencer (l'action) et à travers quelles médiations cela passe (comment ?). De ce point de vue, la coupure entre théorie et dispositif opérationnel est nettement moins tranchée* »²⁷⁶.

Ainsi, Cha *et alii*, proposent de comparer trois critères de mesure d'influence disponibles sur Twitter – le nombre d'abonnés, le nombre de retweets et le nombre de mentions – afin de déterminer quel serait le plus pertinent pour mesurer l'influence, définie ici

²⁷² Eytan Bakshy, Jake M. Hofman, Winter A. Mason et Duncan J. Watts, « Identifying 'Influencers' on Twitter », WSDM 2011, Hong Kong.

²⁷³ Pour une définition des API, *cf.* glossaire p. 464.

²⁷⁴ Eytan Bakshy, Jake Hofman, Winter Mason et Duncan Watts, « Identifying 'Influencers' on Twitter », art. cit.

²⁷⁵ Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto et Krishna P. Gummadi, « Measuring User Influence on Twitter: the Million Follower Fallacy », *ICWSM'10: Proceedings of international AAAI Conference on Weblogs and Social*, mai 2010.

²⁷⁶ Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis*, *op. cit.*

simplement comme « *la capacité à produire un effet de façon directe ou indirecte* »²⁷⁷. À partir d'un corpus de tweets émis par une population de plus de 6 millions d'utilisateurs, les auteurs ont démontré qu'il n'existait pas de corrélations statistiques entre le nombre d'abonnés d'une part, et le nombre de retweets ou de mentions d'autre part. Les tweets publiés par les utilisateurs ayant le plus grand nombre d'abonnés ne sont pas nécessairement les plus retweetés : le nombre d'abonnés serait donc un mauvais indicateur d'influence. Les auteurs avancent également deux autres résultats : les individus les plus influents sur Twitter ne le seraient que dans un petit nombre de domaines (les auteurs ont isolé trois sujets de discussion : les élections iraniennes de 2009, le virus H1N1 et la mort de Michael Jackson) et les individus les plus influents le seraient rarement spontanément ou accidentellement : l'influence se construirait au contraire dans la durée, en publiant régulièrement des tweets relatifs à un domaine spécifique. Les auteurs redécouvrent donc, à partir d'une démarche totalement empiriste et quantitative, un résultat connu de longue date de la sociologie des relations interpersonnelles. Dès *Personal Influence*, Katz et Lazarsfeld avaient en effet démontré que les leaders d'opinion ne l'étaient pas dans tous les domaines²⁷⁸.

Pour Bakshy *et alii*, le nombre d'abonnés serait au contraire un indicateur pertinent pour prédire l'influence d'un individu sur Twitter – l'influence étant ici définie comme la capacité pour un individu à être le premier à publier une URL qui sera ensuite retweetée par ses abonnés. Mais pour être totalement opérant, l'indicateur du nombre de *followers* doit être également associé à un indicateur mesurant l'« influence locale passée » : c'est-à-dire « *le nombre moyen de fois où une de ses URL a été retweetée par ses followers directs comparé au nombre moyen de fois où cette URL a été retweetée par les followers de ses followers jusqu'à disparition de l'URL dans les timelines* »²⁷⁹. Les individus capables de produire de « larges cascades » de diffusion d'une URL (allant de l'individu émetteur vers ses abonnés, puis de ses abonnés vers les abonnés de ses abonnés, etc.) sont relativement rares. Les auteurs concluent donc, dans une optique opérationnelle pour le marketing, qu'il est plus rentable de cibler un nombre important d'individus moyennement influents (avec des cascades de petite ou moyenne portée) plutôt qu'un petit nombre d'individus très influents (avec des cascades de longue portée).

²⁷⁷ Meeyoung Cha *et alii*, « Measuring User Influence on Twitter: the Million Follower Fallacy », art. cit.

²⁷⁸ Elihu Katz et Paul Lazarsfeld, *Influence personnelle*, *op. cit.*

²⁷⁹ Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis*, *op. cit.*

Dominique Boullier et Audrey Lohard proposent en résumé sept « propositions programmatiques », sept conditions pour définir et mesurer l'influence en ligne : être influent nécessite d'attirer l'attention, l'influence s'exerce domaine par domaine, l'influence s'exerce de façon durable, l'influence se définit par la capacité à susciter la discussion et enfin par la capacité à « orienter » la discussion²⁸⁰. Cette dernière proposition, qui se rapproche pourtant le plus de la définition de l'influence proposée par Lazarsfeld et Katz, et plus généralement par les auteurs du « *word-of-mouth* » marketing, n'est pourtant pas mobilisée dans les travaux de marketing viral. Comme le note Beauvisage *et alii*, « *la question de la nature des « contenus » diffusés est un point aveugle d'une grande partie de ces travaux* »²⁸¹. Les recherches de marketing viral et de la nouvelle science des réseaux prennent en effet pour objet les mécanismes de circulation et de diffusion des autopublications sans s'intéresser à leurs contenus, sans tenir compte des opinions possiblement véhiculées par ces messages. Watts et Dodds, dans leur article intitulé « *Influentials, Networks, and Public Opinion Formation* » n'abordent pas la question du contenu des opinions ou de leurs modes de production. Seules comptent les modalités de circulation des opinions et les mécanismes d'influence. L'opinion publique est ainsi définie comme étant le résultat de processus d'influences croisées entre médias, leaders d'opinion et individus. Si les sondages réifient la notion d'opinion publique, le marketing viral ne la réduit qu'à un processus, à un mécanisme de diffusion et de circulation.

Si la question de la nature des contenus est un point aveugle des travaux du marketing viral et de la nouvelle science des réseaux, des chercheurs ont en revanche cherché à étudier, dans le champ des *cultural studies*, les contextes de réception et d'appropriation des contenus culturels diffusés sur internet : « *les idées qui circulent et survivent sont celles qui peuvent être aisément transformées, car elles pourront alors être mises à l'agenda de plusieurs groupes ou sous-cultures, pour lesquels elles seront porteuses de sens, mais d'un sens bien souvent différent. La plupart du temps, les messages ne circulent pas selon un strict mécanisme de réplique, contrairement à ce que la métaphore virale laisse entendre : ils sont retravaillés, remixés, distordus voire détournés* »²⁸². Beauvisage *et alii* estiment ainsi que l'étude des contenus des autopublications doit être un véritable axe de recherche des futurs travaux de marketing viral.

²⁸⁰ Nous ne mentionnerons pas deux autres propositions, qui ont moins d'intérêt dans le fil de notre démonstration.

²⁸¹ Thomas Beauvisage *et alii*, « Le succès sur Internet repose-t-il sur la contagion ? Une analyse des recherches sur la viralité », art. cit.

²⁸² *Ibidem*.

Cette seconde tradition d'étude des autopublications s'intéresse donc aux logiques et mécanismes de diffusion et de circulation des messages publiés sur internet. L'objet d'étude de ces travaux est moins le contenu des autopublications que leur potentiel de « viralité ». Issus des disciplines du marketing viral et de la nouvelle science des réseaux, ces recherches, tout en mobilisant des références théoriques relatives aux travaux sur le leadership d'opinion et les mécanismes d'influence, adoptent, comme Lazarsfeld et des chercheurs du *B.A.S.R.* en leur temps, une démarche totalement empiriste. En reposant sur une approche purement quantitative et sociométrique des réseaux, ces travaux ont définitivement émancipé les analyses d'influence des enquêtes déclaratives par sondage. Cependant, précisément parce qu'ils reposent exclusivement sur des analyses quantitatives des réseaux, ces travaux ont dans le même temps oblitérer le caractère contextuel et microsociologique de l'influence interpersonnelle. Le contenu même des messages et l'identité sociale de leurs auteurs ne sont ainsi pas étudiés.

Lorsque ces travaux s'intéressent à l'« opinion publique », c'est uniquement en interrogeant les logiques macrosociologiques de circulation et de diffusion des opinions qui sous-tendent sa formation. Ces recherches appréhendent en effet l'objet « opinion » comme un « virus » qui se propage d'individus à individus, et non comme le produit d'une socialisation politique ou d'un habitus. À la différence des travaux sur la participation politique en ligne qui appréhendent les opinions sans questionner leur agrégation ou leur rapport avec une « opinion publique », les travaux de marketing viral n'interrogent pas les modes de production individuels des opinions publiées sur internet.

Une troisième tradition d'étude des autopublications prend pour objet à la fois le contenu des opinions, questionne méthodologiquement leur agrégation et les compare à d'autres modes de recueil et de mesure.

3.2.3 *Opinion mining, sentiment analysis et prédictivité de Twitter*

Issus du traitement automatique du langage (TAL), l'*opinion mining* – ou « fouille d'opinion » en français²⁸³ – et le *sentiment analysis* cherchent précisément à détecter, isoler et associer une polarité (négative ou positive) aux opinions publiées sur internet. Ces opinions peuvent porter sur un objet (dénommé « cible » ou « entité ») ou sur une particularité de cet objet (« sous-cible » ou « *feature* »). Le principal enjeu de cette discipline repose sur un travail de classification des matériaux recueillis en différentes catégories. Ces recherches opèrent ainsi une distinction entre « faits » et « opinions » (distinguer, au sein des autopublications, les opinions des « non-opinions »), entre « cible » et « *feature* » (distinguer si l'opinion détectée porte sur un objet ou une particularité de cet objet) et entre polarité positive et négative (et éventuellement neutre).

Les recherches d'*opinion mining* et de *sentiment analysis* revêtent souvent une finalité opérationnelle. Nous l'avons vu, de nombreuses offres commerciales se sont développées autour de la détection automatique des opinions, notamment pour analyser les avis de consommateurs en ligne. Mais les outils développés par l'*opinion mining* et le *sentiment analysis* sont également mobilisés au sein de recherches académiques cherchant à établir des corrélations statistiques entre les opinions exprimées sur internet et les opinions mesurées par sondages. Ces recherches prétendent en effet disputer aux sondages d'opinion leur prétention prédictive.

3.2.3.1. Les enjeux scientifiques de l'*opinion mining* et du *sentiment analysis*

Les recherches de détection automatique des opinions sont donc confrontées à trois principaux enjeux : dissocier les documents comprenant des « opinions » de ceux relatant uniquement des « faits », puis, au sein des documents porteurs d'opinion, dissocier les opinions portant sur les « cibles » de celles portant sur les « *feature* », et enfin qualifier les opinions détectées en leur associant une polarité positive, négative ou neutre.

Dissocier *a priori* des textes relatant des faits et des textes véhiculant des opinions est une démarche de recherche commune à l'ensemble des travaux d'*opinion mining*, voire du

²⁸³ Egle Eensoo-Ramdani, Evelyne Bourion, Monique Slodzian et Mathieu Valette, « De la fouille de donnée à la fabrique de l'opinion », *Les Cahiers du numérique*, 2 (2011), p. 15-39.

TAL en général²⁸⁴. Selon cette approche, les textes porteurs d'opinion se caractérisent par la présence d'éléments ou d'indices de subjectivité. Une fouille d'opinions débute donc par une détection de ces indices de subjectivité afin de classer les textes en deux catégories, les opinions et les faits.

Cette opposition entre opinions et faits, entre subjectivité et objectivité renvoie au dualisme faits/valeurs défendu par le positivisme logique²⁸⁵. Pour Eensoo-Ramdani, « *considérer la tâche de détection de la subjectivité comme autonome par rapport à la détection d'opinion elle-même revient à assumer une option logiciste puisqu'il s'agit de séparer les faits des opinions. Autrement dit, dans maints travaux, on suppose que les segments linguistiques de longueur et de structure différentes peuvent être classés a priori dans l'une ou l'autre des catégories – fait ou valeur – et que seuls les segments véhiculant des valeurs sont pertinents pour l'analyse* »²⁸⁶. Différentes recherches ont ainsi démontré qu'une détection automatique de tonalité à partir d'un corpus dans lequel ont été préalablement filtrés les éléments textuels ne comprenant pas d'indice de subjectivité arrive à des meilleurs résultats qu'une détection de tonalité réalisée à partir d'un corpus où faits et valeurs n'ont pas été dissociés *a priori*²⁸⁷. Plusieurs indices peuvent être retenus pour témoigner de la présence d'éléments de subjectivité dans un texte ou dans une phrase. Certains auteurs privilégient la présence d'adjectifs dans la phrase²⁸⁸, d'autres l'emplacement des mots²⁸⁹ ou encore des « indices co-occurents de subjectivité »²⁹⁰.

Cette opération de classification *a priori* entre opinions et faits est néanmoins discutabile épistémologiquement et a été dénoncée par certains chercheurs²⁹¹. Boullier et Lohard en résument bien la critique :

²⁸⁴ Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis, op. cit.*, p. 148.

²⁸⁵ Egle Eensoo-Ramdani *et alii*, « De la fouille de donnée à la fabrique de l'opinion », art. cit.

²⁸⁶ *Ibidem*.

²⁸⁷ Rada Mihalcea, Carmen Banea et Janyce Wiebe, « Learning multilingual subjective language via crosslingual projections », *Proceedings of the Association for Computational Linguistics (ACL)*, Prague, Czech Republic, 2007.

²⁸⁸ Vasileios Hatzivassiloglou et Janyce M. Wiebe, « Effects of Adjective Orientation and Gradability on Sentence Subjectivity », *Proceedings of the 18th International Conference on Computational Linguistics*, vol. 1, n°11-12, 2000, p. 299-305

²⁸⁹ Janyce M. Wiebe, Rebecca, F. Bruce et Thomas P. O'Hara, « Development and Use of a Gold-Standard Data Set for Subjectivity Classifications », *Proceedings of the Association for Computational Linguistics (ACL)*, vol. 37, 1999, p. 246-253.

²⁹⁰ Janyce Wiebe, Theresa Wilson et Matthew Bell « Identifying collocations for recognizing opinions », *Proceedings of the ACL/EACL Workshop on Collocation*, 2001.

²⁹¹ Bo Pang et Lillian Lee, « Opinion Mining and Sentiment Analysis », *Foundations and Trends in Information Retrieval*, vol. 2, n° 1-2, 2008, p. 1-135 ; Bing Liu, « Sentiment Analysis and Subjectivity », in N. Indurkha et

« Tout le problème vient bien de cette tentative de séparer les opinions des faits et de prétendre pouvoir le faire à partir d'indicateurs simples, alors même que la posture moderne positiviste qui sous-tend cette approche, la distinction des faits et des valeurs [...] est remise en cause par tous les travaux innovants de philosophie et de sciences sociales. Tout se passe comme si l'on tentait d'appliquer au langage des principes dont les sciences de la nature ont compris les limites [...]. Le subjectif, les valeurs, l'observateur ne peuvent jamais être évacués de l'objectif, des faits, de l'observé. [...] Attribuer une qualité substantielle subjective ou objective a priori à un matériau linguistique, c'est à nouveau tenter de réduire le langage à un étiquetage. Les catégories objectif/subjectif sont des résultats des interactions, des échanges, et demeurent incertaines : elles ne sont jamais données »²⁹².

Afin de démontrer qu'une distinction *a priori* entre subjectif et objectif n'est pas forcément pertinente, Pang et Lee prennent ainsi l'exemple de l'expression « *long batterie life* »²⁹³. Considérée *a priori* comme factuelle, cette expression peut, dans un certain contexte d'énonciation, assigner en réalité une valeur positive à l'ordinateur, puisque la durée de vie de la batterie peut être un élément important dans la décision d'achat du consommateur qui a publié cette expression. Une phrase *a priori* objective peut donc être porteuse d'opinion. Inversement, une phrase contenant des éléments de subjectivité, comme « *Je pense qu'elle me l'a dit* », peut n'exprimer aucune opinion²⁹⁴. En résumé, opérer une distinction opinions/faits *a priori* peut engendrer de nombreuses pertes heuristiques, en écartant des éléments textuels qui sont bien porteurs d'opinion. Inversement, n'opérer aucune opération de classement peut conduire à travailler sur un corpus comprenant des autopublications qui ne sont pas des opinions²⁹⁵. Il s'agit précisément, pour Boullier et Lohard, d'une limite de l'automatisation du sentiment analysis²⁹⁶.

Après avoir isolés – ou non – les textes porteurs d'opinion, une seconde opération consiste à déterminer sur quel objet (« entité ») ou sur quel(s) caractéristique(s) de l'objet (« *feature* ») porte l'opinion détectée. Il s'agit donc d'opérer une nouvelle classification pour dissocier « entité » et « *feature* ». L'opinion peut donc porter sur un objet et/ou sur des caractéristiques de cet objet. L'enjeu est alors de savoir s'il existe une relation entre la polarité

F. J. Damerau (dir.), *Handbook of Natural Language Processing*, Boca Raton, Chapman & Hall/CRC Pres, 2010.

²⁹² Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis*, op. cit., p. 151.

²⁹³ Bo Pang et Lillian Lee, « Opinion Mining and Sentiment Analysis », art. cit.

²⁹⁴ Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis*, op. cit., p. 152.

²⁹⁵ Dans de nombreuses offres commerciales, des phrases factuelles, qui ne véhiculent pas d'opinion se voient alors attribuer une tonalité « neutre ».

²⁹⁶ Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis*, op. cit., p. 152.

de l'opinion exprimée sur l'objet et les polarités des opinions exprimées sur les caractéristiques de l'objet. L'importance accordée à cette distinction entre « entité » et « *feature* » se comprend d'autant mieux que, nous l'avons dit, les travaux d'*opinion mining* intéressent au premier chef le marketing ; les travaux d'*opinion mining* prennent donc souvent comme matériau d'étude des avis de consommateurs sur des biens de consommation. Il ne s'agit donc pas seulement de savoir si les internautes/consommateurs apprécient ou non un nouveau téléphone, mais de déterminer quelles caractéristiques de ce téléphone leur plaît ou leur déplaît.

Enfin, la dernière opération de classement consiste à attribuer une tonalité, généralement négative, neutre ou positive, aux opinions détectées. Il existe deux principales méthodes d'attribution automatique de tonalité²⁹⁷. La première repose sur l'utilisation de dictionnaires de mots, constitués manuellement, à partir de corpus et/ou à partir de dictionnaires déjà existants. Dans ces dictionnaires, une polarité est associée *a priori* à chacun des mots (listes de mots positifs, liste de mots négatifs, etc.). Quel que soit le contexte dans lequel le mot sera inséré, il sera alors toujours associé à la même polarité. La seconde approche du codage automatique de la tonalité repose sur l'apprentissage. Des algorithmes s'« entraînent » à détecter des marqueurs d'opinion dans un premier corpus-test (éventuellement annoté manuellement), puis détectent ensuite « seuls » ces marqueurs dans un corpus plus vaste. Nous l'avons vu, l'analyse automatique de tonalité connaît de nombreuses lacunes techniques et méthodologiques. Les plus graves concernent leur incapacité à repérer de nombreuses figures rhétoriques comme l'humour, le sarcasme, les métaphores, etc.²⁹⁸. Or, nous le verrons, nombre de tweets ont recours à des figures rhétoriques pour exprimer des opinions. Cette incapacité renvoie à la critique linguistique selon laquelle le lexicométrie ne s'arrêterait qu'à la matérialité graphique des textes : « *de fait, le « mot », pris dans sa définition la plus restrictive, ne recouvre pas une réalité linguistique opérante pour permettre la compréhension des textes. Ainsi la lexicométrie peut être soupçonnée de permettre, au mieux, une description du contenu matériel « de surface » des textes, et aucunement d'en recouvrer le sens* »²⁹⁹.

Bing Liu résume dans sa définition de l'opinion l'ensemble de ces opérations de classement : « *a direct opinion is a quintuple (oj, fjk, ooiykl, hi, tl), where oj is an object, fjk is*

²⁹⁷ *Ibidem*.

²⁹⁸ Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis, op. cit.*

²⁹⁹ Damon Mayaffre, « De la lexicométrie à la logométrie », art. cit.

a feature of the object oj , $ooijkl$ is the orientation or polarity of the opinion on feature fjk of object oj , hi is the opinion holder and tl is the time when the opinion is expressed by hi . The opinion orientation $ooijkl$ can be positive, negative or neutral (or measured based on a more granular scale to express different strengths of opinions). For feature fjk that opinion holder hi comments on, he/she chooses a word or phrase from the corresponding synonym set Wjk , or a word or phrase from the corresponding feature indicator set Ijk to describe the feature, and then expresses a positive, negative or neutral opinion on the feature»³⁰⁰. Cette formule, relativement simple et descriptive, rend compte de la manière dont les chercheurs en *opinion mining* conçoivent l'opinion. L'opinion n'est pas le produit d'une interaction, d'un contexte, d'une socialisation ou encore d'un habitus. L'opinion est une tonalité associée à un objet et/ou à une ou plusieurs caractéristiques de cet objet, par un porteur d'opinion à un temps t . Bethard *et alii*, définissent encore plus simplement l'opinion comme « une phrase ou une partie de phrase qui pourrait répondre à la question « Qu'est-ce que X pense de Y ? »³⁰¹. Et si l'auteur ou l'« opinion holder » est une variable centrale de l'équation, Boullier et Lohard notent cependant que « la capacité à utiliser les références de l'auteur de l'opinion est souvent plus difficile lorsque l'on fait du traitement de masse des données d'opinion sur le web »³⁰². Or, précisément, les outils d'*opinion mining* sont de plus en plus souvent mobilisés dans des travaux analysant d'importantes masses de données, sur Twitter notamment.

3.2.3.2 Les analyses prédictives de *sentiment analysis*

Parus en 2010, une série de travaux, américains³⁰³ et allemands³⁰⁴, ont mobilisé des analyses automatisées de sentiment – souvent en complément d'analyses lexicométriques –

³⁰⁰ Bing Liu, « Sentiment Analysis and Subjectivity », in N. Indurkha et F. J. Damerau (dir.), *Handbook of Natural Language Processing* (2ème édition), *op. cit.*

³⁰¹ Egle Eensoo-Ramdani *et alii*, « De la fouille de donnée à la fabrication de l'opinion », art. cit.

³⁰² Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis*, *op. cit.*

³⁰³ Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge et Noah A. Smith, « From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series », *Proceedings of International AAAI Conference on Weblogs and Social Media*, Washington, DC, 2010, p. 1-8. ; Sitaram Asur et Bernardo A. Huberman, « Predicting the Future with Social Media », *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Vol. 01, p. 492-499.

³⁰⁴ Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner et Isabell M. Welp, « Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment », *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 2010, Menlo Park, CA: The AAAI Press, p. 178-185 ; Andreas Jungherr, Pascal Jürgens et Harald Schoen, « Why the Pirate Party Won the German Election of 2009 or the Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welp, I. M. "Predicting Elections With 140 Characters Reveal About Political Sentiment" », *Social Science Computer Review*, p. 1-6.

pour répondre à la question suivante : Twitter est-il prédictif de futurs résultats électoraux, et de l'« opinion publique »³⁰⁵ en général ? Et partant, Twitter est-il davantage « prédictif » que les sondages d'opinion ? Dit autrement, observe-t-on une corrélation statistique entre les opinions publiées et mesurées sur Twitter, les opinions mesurées par sondage et des phénomènes sociaux réels (comme les résultats d'une élection, les entrées ou les recettes générées par un film, etc.) ? En questionnant d'une part l'articulation entre les opinions individuelles et l'« opinion publique » et en interrogeant d'autre part la « prédictivité » des instruments de mesure, ces travaux posent des questions de recherche relativement proches de celles développées par la sociologie électorale. Néanmoins, tant du point de vue des cadres et références théoriques mobilisés que des méthodes employées, ces recherches diffèrent radicalement des travaux sociologiques.

« *Can we analyze publicly available data to infer population attitudes in the same manner that public opinion pollsters query a population?* »³⁰⁶. Afin de répondre à cette question de recherche, Brendan O'Connor *et alii* ont constitué deux corpus de données, l'un composé de tweets dont la tonalité a été annotée, l'autre de données de sondages, puis ont mené des tests de corrélations et de régression statistiques afin de valider l'existence d'un lien entre analyse de sentiment sur Twitter et données de sondages. Les auteurs ont retenu trois sujets de comparaison : un d'ordre économique – la confiance des consommateurs américains – et deux sujets d'ordre politique, la cote de popularité du Président Obama et les intentions de vote en faveur de Barak Obama et John McCain lors de l'élection présidentielle américaine de 2008. Les auteurs ont mobilisé des données de sondages relatives à chaque sujet, et les ont comparées à des tweets codés positivement, neutres, ou négativement. Les corpus de tweets ont été collectés à partir d'une sélection par mots-clés, sur l'ensemble des tweets publiés durant la séquence étudiée. Concernant la confiance des consommateurs, ont été sélectionnés les tweets comprenant les occurrences « *economy* », « *job* » et « *jobs* », pour la popularité du président l'occurrence « *Obama* » et pour les élections, les occurrences « *Obama* » et « *McCain* ». Les tweets sélectionnés ont été ensuite codés automatiquement, à partir d'un dictionnaire comprenant une liste de 1 600 termes prédéfinis comme positifs et 1 200 termes prédéfinis comme négatifs. Conscients des limites heuristiques d'une approche par dictionnaire, les auteurs soulignent simplement que les possibles erreurs de codage au niveau

³⁰⁵ Nous nous en tenons ici à la définition sondagière de l'« opinion publique ». Nous poursuivrons une réflexion critique de la notion dans le chapitre 2.

³⁰⁶ Brendan O'Connor *et alii*, « From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series », art. cit.

individuel peuvent se neutraliser au niveau agrégé, le seul niveau qui compte pour l'analyse. Ensuite, les auteurs ont calculé un ratio quotidien de tweets positifs sur tweets négatifs à propos de chaque sujet. Ces données ont été ensuite comparées, *via* des analyses de corrélations, aux données de sondages. Trois principaux résultats se dégagent de l'analyse. Le premier est que les sondages reflèteraient mieux des tendances de long terme. Lorsque l'on compare les tweets et les sondages uniquement le dernier jour de réalisation du sondage, le taux de corrélation est faible. Plus l'on fait une moyenne des tweets sur une séquence temporelle longue – sur 15, 30, voire 60 jours – plus les résultats se rapprochent de ceux des sondages. Les auteurs en concluent que « *these facts imply that the text sentiment in formation is volatile, and if polls are believed to be a gold standard, then it is best used to detect long-term trends* »³⁰⁷. Le second résultat avancé est que la période historique jouerait sur le taux de corrélation entre tweets et sondages. Le contexte historique serait une variable déterminante du *sentiment analysis* sur Twitter. Pour en arriver à cette conclusion, les auteurs ont observé que durant un an et demi, le taux de corrélation concernant la confiance des consommateurs américains était presque nul. Puis, à partir de la mi-2009, le *sentiment analysis* devient un bien meilleur indicateur et recoupe les tendances mesurées par les sondages. Les auteurs concluent que « *this suggests qualitatively different phenomena are being captured by the text sentiment measure at different times. From the perspective of time series modeling, future work should investigate techniques for deciding the importance of different historical signals and time periods, such as vector autoregressions* ». Enfin, le dernier résultat avancé, sans doute le plus important, est que dans le cas de l'élection présidentielle, la fréquence de citation des candidats (nombre de tweets contenant l'occurrence « Obama » *versus* le nombre de tweets contenant l'occurrence « McCain ») était un meilleur indicateur (c'est-à-dire témoigne d'un taux de corrélation plus fort avec les sondages) que l'analyse de la tonalité des tweets.

Ce dernier résultat a été constaté par plusieurs autres études. Asur et Huberman, dans une démarche et un questionnement relativement proches de ceux d'O'Connor *et alii* constatent également que le simple volume de citation est un meilleur indicateur que la tonalité des tweets³⁰⁸. Les deux auteurs cherchent à valider l'existence d'une corrélation entre le nombre et la tonalité des tweets publiés à propos d'un film d'une part, et le montant des recettes que va générer le film après sa sortie en salles d'autre part. La méthodologie adoptée

³⁰⁷ *Ibidem.*

³⁰⁸ Sitaram Asur et Bernardo A. Huberman, « Predicting the Future with Social Media », art. cit.

est relativement proche de celle d'O'Connor *et alii* : trois corpus de données sont analysés : l'un est composé de tweets annotés, l'autre de données issues du *Hollywood Stock Exchange index* (HSX) – un simulateur économique qui est, selon les auteurs, un bon indicateur des futures recettes d'un film – et enfin un troisième corpus recensant les revenus effectifs générés par chaque film lors de leur premier week-end d'exploitation en salles. Une différence notable avec la recherche d'O'Connor *et alii* est qu'une partie des tweets ont été annotés manuellement : les auteurs ont eu recours au service *Amazon Mechanical Turk*, un site proposant de confier des tâches informatiques répétitives et nécessitant peu de qualification à des internautes, quel que soit leur localisation de par le monde, en échange d'une rémunération très faible. Pour éviter un biais « annotateur », chaque tweet a été codé par trois annotateurs différents. Les auteurs ont ensuite, à partir de ce corpus de tweets annoté, construit un ratio de tweets positifs sur tweets négatifs qu'ils ont comparé avec les deux autres jeux de données. À partir d'une série de tests de corrélations et de régressions logistiques, les auteurs en arrivent à la conclusion que la tonalité des tweets comme leur volume de citations sont deux très bons indicateurs des revenus futurs d'un film, mais notent cependant que le taux de corrélation est plus élevé concernant le volume de citations des films.

Tumasjan *et alii* n'accordent pas quant à eux la même fonction à l'analyse de sentiment et au volume de citations des candidats³⁰⁹. Dans le cadre des élections législatives allemandes de 2009, les auteurs ont eux aussi interrogé la valeur prédictive de Twitter en comparant des échantillons de tweets avec les résultats électoraux définitifs de chaque parti. Les tweets ont été sélectionnés à partir d'une recherche par mots-clés : 70000 messages mentionnant un des six principaux partis en lice, 35000 se référant à des leaders politiques. Pour ces auteurs, l'analyse de sentiment des tweets constituerait un bon indicateur des proximités politiques entre partis et permettrait d'augurer ainsi des possibles coalitions à venir. Pour en arriver à cette conclusion, les auteurs ont codé automatiquement leur corpus de tweets à partir d'un logiciel d'*opinion mining*³¹⁰. Les tweets n'ont pas été codés à partir de l'échelle négatif/neutre/positif mais d'un classement en 12 catégories de « sentiment » : émotions positives, émotions négatives, tristesse, anxiété, colère, hésitation, certitude, travail, accomplissement, argent, orientation future et orientation passée. Les auteurs ont ensuite calculé un score pour chaque parti et candidat. Par comparaison, les auteurs ont constaté qu'il y avait par exemple bien plus de proximité « de sentiment » entre les tweets émis sur le SPD

³⁰⁹ Andranik Tumasjan *et alii*, « Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment », art. cit.

³¹⁰ Le logiciel LIWX2007.

et la CSU, les deux futurs partenaires de la Grande coalition de 2009, qu'entre ceux sur le SPD et Die Link, le parti de gauche radicale. Si l'analyse de sentiment a donc donné lieu à une interprétation probante mais davantage qualitative du jeu électoral allemand, c'est avec l'analyse du volume de citation que la recherche en arrive à affirmer clairement le caractère prédictif de Twitter. Les auteurs ont en effet constaté, pour le dire simplement, que l'ordre de citation des candidats et des partis sur Twitter reflétait très fidèlement l'ordre d'arrivée des candidats dans les urnes. Encore une fois à partir de tests de corrélations statistiques, les auteurs en arrivaient donc à affirmer le caractère prédictif de ce réseau social.

L'ensemble de ces recherches attribuent donc à Twitter et à l'analyse de sentiment une capacité de prédiction des phénomènes sociaux et des mouvements futurs d'opinion. Cependant, si les auteurs constatent une corrélation entre données *online* et données *offline*, ils ne l'expliquent pas. Ces recherches sont donc confrontées à une sorte de « boîte noire » sociologique. D'autant plus qu'un constat est partagé par la majorité de ces travaux : Twitter n'est pas représentatif de la population électorale³¹¹. À l'inverse des sondages d'opinion, l'« impératif de représentativité »³¹² ne peut donc être mobilisé pour expliquer ce caractère prédictif. Comment dès lors comprendre que Twitter soit prédictif de l'« opinion publique » si les internautes qui y expriment leurs opinions ne sont pas représentatifs de l'ensemble de la population ? À cette question, ces recherches n'apportent aucune réponse.

En présentant ces différents travaux, issus de ce qui nous semble être les trois grandes traditions d'étude des autopublications numériques dans le champ académique, nous avons donc souhaité démontrer d'une part l'extrême hétérogénéité de ces recherches, tant du point de vue de leurs problématiques, de leurs hypothèses, de leurs méthodes, de leurs approches, etc. ainsi que l'absence d'autre part de toute référence aux théories de Pierre Bourdieu et de la sociologie critique relatives à la déconstruction de la notion d'opinion publique, aux modes de productions des opinions politiques et à la problématique de la compétence politique.

³¹¹ Nous le démontrerons dans notre chapitre 3.

³¹² Alain Desrosières, *La politique des grands nombres, op. cit.*

Conclusion du chapitre 1

La démocratisation de l'accès à internet et la création de plateformes simplifiées d'autopublication ont favorisé le développement d'une nouvelle pratique sociale, l'autopublication numérique. De manière inédite, des milliers d'individus commentent ainsi quotidiennement leur vie privée mais aussi la vie de la cité. Des millions de messages et de discussions se trouvent dorénavant archivés et accessibles à tout un chacun sur internet. La numérisation de l'écrit a en effet rendu possible cet archivage géant des conversations.

Ces millions de messages peuvent être appréhendés comme un nouveau matériau d'étude du monde social. Un matériau qui présente certaines spécificités heuristiques. La première d'entre elles, et sans doute la plus stimulante, est que les autopublications ne sont pas des données produites en situation de laboratoire. Les messages publiés sur internet ne présentent donc aucun biais déclaratif. Le développement d'outils informatiques dédiés a également permis d'opérer un traitement quantitatif mais aussi qualitatif des autopublications, de manière statique mais aussi dynamique. Ce nouveau matériau permet donc de dépasser les frontières conceptuelles « classiques » entre qualitatif et quantitatif, micro et macro, statistique et dynamique.

La richesse de ce nouveau matériau et ses plus-values heuristiques ont en premier lieu intéressé le champ professionnel de l'opinion. Un nouveau secteur d'activité s'est créé autour de l'étude des autopublications. C'est au sein de ce secteur qu'ont été inventés et développés des outils et des instruments de mesure des messages publiés sur internet. Pour diverses raisons, les instituts de sondage, acteurs traditionnels de l'opinion, ont, sur internet, perdu leur monopole. Cette ouverture du champ à des acteurs venus des secteurs de la veille, de la communication ou de la publicité a eu pour conséquence d'introduire de nouvelles approches, de nouveaux concepts et de nouvelles méthodes d'étude de l'opinion. Mais comme le note Boullier et Lohard « *toutes [ces] mesures actuelles manquent cruellement d'arrière-plan théorique. On comprend ainsi l'impérieuse nécessité d'un investissement par la théorie des sciences sociales de toutes ces questions de mesures sur le web, pour ne pas les laisser aux web sciences, pour qui tout est bon, du moment que l'on trouve une explication ad hoc* »³¹³.

À l'intérieur du champ scientifique ce sont en effet avant tout des disciplines qui connaissent un rapport étroit avec l'informatique qui ont investi ce nouveau matériau et leurs instruments de mesure. Les sciences de l'information et de la communication, les sciences de

³¹³ Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis*, op. cit.

gestion et de marketing ou la « nouvelle science des réseaux » ont chacune développé des cadres conceptuels spécifiques pour étudier les autopublications et les opinions numériques. Si certaines font références à des théories de sociologie politique – l'influence interpersonnelle en science du marketing ou l'espace public habermassien en sciences de l'information – très peu de ces travaux mobilisent le cadre théorique des travaux de sociologie critique sur l'opinion publique. La science politique et la sociologie politique n'ont d'ailleurs accordé un intérêt que très marginal aux autopublications numériques.

Nous défendrons dans le prochain chapitre l'idée selon laquelle les autopublications numériques et leurs instruments de mesure permettent d'objectiver empiriquement et de questionner sous un jour nouveau les théories et hypothèses de la sociologie bourdieusienne relatives à l'opinion publique et à la compétence politique.

CHAPITRE 2 - PANÉLISER TWITTER, UNE MÉTHODE HYBRIDE DE MESURE DES OPINIONS

Dans la majorité des travaux académiques, les autopublications numériques sont étudiées à partir de problématiques de recherche relativement éloignées des préoccupations de la sociologie politique. Cette discipline n'a d'ailleurs accordé qu'un faible intérêt à ce nouveau matériau d'enquête. Pourtant, l'autopublication numérique constitue bien un nouveau mode d'expression des opinions.

Dans ce deuxième chapitre, nous soutiendrons que l'appareillage conceptuel et théorique de la sociologie critique est à même de proposer une analyse opérante et stimulante de ce nouveau matériau et que, réciproquement, l'étude des opinions numériques permet de valider empiriquement certaines hypothèses développées par la critique universitaire des sondages.

La critique universitaire des sondages se développe en France à partir des années 1970. Cette critique est triple. Elle concerne des aspects purement méthodologiques (on questionne les logiques d'échantillonnage, de représentativité, de redressements statistiques, etc.)³¹⁴, soulève des questions épistémologiques (que mesure réellement l'instrument ?³¹⁵, n'existe-t-il pas un biais d'imposition de problématique ?³¹⁶, l'opinion mesurée existe-t-elle ?³¹⁷) et interroge enfin les effets sociaux des sondages (quel est l'impact de la diffusion de l'instrument sur le champ politique ?³¹⁸).

« L'opinion publique n'existe pas », l'exposé de Pierre Bourdieu réalisé à Noroit en janvier 1972 et publié un an plus tard dans la revue *Les Temps modernes*³¹⁹, est sans doute la contribution scientifique qui reste aujourd'hui la plus associée à la critique universitaire des sondages. L'auteur y affirme que les sondages d'opinion reposent sur trois contre-vérités sociologiques : tout le monde ne serait pas en mesure de produire une opinion sur tous les

³¹⁴ Patrick Lehingue, *Subunda. Coups de sonde dans l'océan des sondages*, Éditions du croquant, 2007, p. 57-124.

³¹⁵ Patrick Lehingue, *Subunda, op. cit.*, p. 125-194.

³¹⁶ Daniel Gaxie, *Le cens caché. Inégalités culturelles et ségrégation politique*, Paris, Éditions du Seuil, 1978.

³¹⁷ Pierre Bourdieu, « L'opinion publique n'existe pas », *Les Temps modernes*, n°318, janvier 1973, p. 1292-1309.

³¹⁸ Patrick Champagne, *Faire l'opinion. Le nouveau jeu politique*, Paris, Les Éditions de Minuit, 1990 ; Patrick Lehingue, *Subunda, op. cit.*, p. 193-258.

³¹⁹ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

sujets, toutes les opinions n'auraient pas le même poids, la même valeur, et il n'existerait pas de consensus sur les questions qui méritent d'être posées. Pierre Bourdieu affirme donc que la fonction sociale des sondages est de dissimuler que « *l'état de l'opinion à un moment donné du temps est un système de forces, de tensions et qu'il n'est rien de plus inadéquat pour représenter l'état de l'opinion qu'un pourcentage* »³²⁰. Si l'auteur a par la suite prolongé cette réflexion dans différents travaux³²¹, en détaillant notamment les modes de production des opinions politiques, il ne proposa pas un dispositif empirique plus « adéquat » pour représenter l'état de l'opinion, un dispositif d'enquête alternatif aux sondages.

Nous soutenons ici que les autopublications numériques et leurs outils de collecte et de mesure offrent au sociologue des éléments de réflexion relatifs à l'élaboration d'un dispositif alternatif de représentation et de mesure de l'opinion. Mis au service d'une démarche de recherche rigoureuse, ces nouveaux outils permettent en effet de dépasser les limites épistémologiques des sondages et de l'enquête par questionnaire.

Une posture passive d'enregistrement des opinions numériques neutralise en premier lieu le biais d'imposition de problématique et ses effets éventuels. Sur internet, les opinions publiées ne sont pas suscitées par les questions du chercheur, les enquêtés ne sont donc pas « *mis en demeure de répondre à des questions qu'ils ne se sont pas posées* »³²². Pour autant, peut-on affirmer que les opinions mesurées sur internet sont des opinions spontanées, mobilisées ? Pour Bourdieu, « *l'opinion qui s'affirme comme telle, spontanément, c'est l'opinion des gens dont l'opinion a du poids* »³²³. Cette hypothèse peut être vérifiée empiriquement sur internet. Les données chiffrées relatives à l'audience ou à la diffusion d'un contenu (le nombre d'abonnés d'un compte sur Twitter ou le nombre de retweets d'un message) permettent de mesurer le « poids » d'une opinion et objectiver ainsi l'inégale valeur des opinions. De plus, en rendant partiellement objectivable le contexte dans lequel sont produites et énoncées les opinions numériques, une posture passive d'enregistrement permet de valider certaines hypothèses développées par les approches cognitivistes³²⁴ et ethnographiques des opinions³²⁵. Nous esquisserons alors un modèle théorique qui tente

³²⁰ *Ibidem*.

³²¹ Pierre Bourdieu, « Question de politique », *Actes de la recherche en science sociales*, Vol. 16, septembre 1977, p. 55-89. Et Pierre Bourdieu, *La Distinction, Critique sociale du jugement*, Paris, Les Éditions de Minuit, 1979.

³²² Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

³²³ *Ibidem*.

³²⁴ John Zaller, *The Nature and Origins of Mass Opinion*, Cambridge, Cambridge University Press, 1992.

³²⁵ Nina Eliasoph, *L'évitement du politique. Comment les Américains produisent l'apathie dans la vie quotidienne*, Paris, Economica, 2010 ; Katherine Cramer Walsh, *Talking About Politics. Informal Groups and*

d'articuler ces approches avec la théorie bourdieusienne des modes de production des opinions politiques.

L'usage exclusif d'une posture passive d'enregistrement des opinions comporte néanmoins certaines limites heuristiques. La plus importante est l'impossibilité d'identifier de manière systématique l'identité sociodémographique et politique des auteurs des opinions numériques. Ces variables sont pourtant indispensables à l'analyse. Elles permettent en premier lieu de situer les auteurs des opinions numériques dans l'espace social. L'« internaute » n'est pas une entité abstraite qui évoluerait dans un monde parallèle libéré de toute pesanteur sociale. Les « internautes » sont bien des agents sociaux, inégalement pourvus en capitaux économiques et culturels, inégalement compétents politiquement. Une recherche sociologique sur les opinions numériques ne peut donc faire l'économie de ces données sociodémographiques.

C'est afin de résoudre cette tension entre la plus-value heuristique d'un enregistrement passif des opinions, qui permet de neutraliser le biais d'imposition de problématique, et la nécessité absolue de situer socialement les producteurs de ces opinions, que nous avons imaginé notre méthodologie d'enquête. Le dispositif consiste à suivre durant plusieurs mois un panel d'internautes inscrits sur le réseau social Twitter. Afin de pouvoir situer socialement les membres du panel et leur future production (les tweets), un court questionnaire (comprenant une vingtaine de questions socio-démographiques et de préférence politique) a été administré à un échantillon représentatif d'individus inscrits sur le réseau social. Après l'administration du questionnaire, aucune autre interaction n'a eu lieu entre le chercheur et les enquêtés. À l'aide d'un logiciel de veille et de *tracking*, le logiciel *Ami Opinion Tracker*, nous avons ensuite enregistré et analysé l'exhaustivité des tweets publiés par les membres du panel durant plusieurs mois.

Avant de revenir plus en détail sur la présentation de notre méthodologie et le choix de notre terrain d'enquête – le réseau social Twitter – (4) nous évoquerons dans une première section les différentes traditions de mesure et de représentation des opinions en sciences sociales (1) puis aborderons plus en détail les apports heuristiques d'une posture passive d'enregistrement des opinions numériques (2) en en soulignant néanmoins les principales limites (3).

Social Identity in American Life, Chicago, The University of Chicago Press, 2004 ; Melissa Victoria Harris-Lacewell, *Barbershops, Bibles and BET. Every-Day Talk and Black Political Thought*, Princeton, Princeton University Press, 2004.

1. Représenter, définir et mesurer les opinions en sciences sociales

En sciences sociales comme dans le champ politique, le concept d'opinion publique est fortement associé à un instrument : le sondage d'opinion. Pourtant, l'équivalence entre résultats de sondage et opinion publique n'est pas une évidence.

Trois postures de recherche permettent de dénaturiser cette équivalence. Une première est de recourir à l'histoire. Loïc Blondiaux³²⁶ a ainsi rappelé que le succès commercial, politique et scientifique des sondages n'avait rien d'une évidence et que ce succès est parvenu à faire oublier les définitions et représentations antérieures et concurrentes de l'opinion publique « sondagière ».

La critique épistémologique est une seconde posture permettant de dénaturiser l'équivalence entre résultats de sondages et opinion publique. Daniel Gaxie³²⁷ et Pierre Bourdieu³²⁸ notamment, à l'appui d'un important travail empirique, ont ainsi déconstruit la définition sondagière de l'opinion publique, en affirmant que cette opinion revêt fréquemment une dimension artefactuelle, comme nous l'avons évoqué plus haut.

À la fin des années 1990, des auteurs ont expérimenté de nouveaux dispositifs d'étude de l'opinion publique, comme les questionnaires de choix³²⁹ ou les sondages délibératifs³³⁰. En démontrant que d'autres représentations et mesures de l'opinion étaient possibles, la création de ces dispositifs alternatifs a constitué une troisième manière de dénaturiser l'équivalence entre sondages et opinion publique. Dans cette première section nous reviendrons de manière chronologique sur les différents modes de représentation et de mesure de l'opinion publique. Notre objectif est double. Il s'agit d'une part de mettre en lumière une pluralité de définitions et de dispositifs d'étude de l'opinion au moment où l'émergence des opinions numériques appelle précisément une réflexion méthodologique afin de concevoir des dispositifs opérants pour étudier ce nouveau matériau. L'étude de ces dispositifs alternatifs permet donc d'élargir la réflexion et de ne pas réduire le champ des possibles méthodologiques au seul horizon des sondages. Le deuxième objectif est de mettre en

³²⁶ Loïc Blondiaux, *La Fabrique de l'opinion. Une histoire sociale des sondages*, Paris, Le Seuil, 1998.

³²⁷ Daniel Gaxie, *Le cens caché*, *op. cit.*

³²⁸ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit. ; Pierre Bourdieu, *La Distinction*, *op. cit.*

³²⁹ Peter Neijens, Jan de Ridder et Willem Saris, « An Instrument for Collecting Informed », *Opinions Quality and Quantity*, 26, 1992, p. 245-258.

³³⁰ James Fishkin, *Democracy and Deliberation*, Yale, Yale University Press, 1991 ; Loïc Blondiaux, « Sondages et délibération. Une épistémologie alternative de l'opinion publique », *Politix*, Volume 15, Numéro 57, 2002, p.167-180.

perspective les propositions de dispositifs empiriques alternatifs imaginés par les théoriciens de la démocratie délibérative avec la critique universitaire des sondages. Dit autrement, nous nous demanderons si les sondages délibératifs constituent un dispositif de représentation et de mesure de l'opinion qui répond à la définition bourdieusienne de l'opinion, à savoir celle d' « *un système de forces [et de] de tensions* »³³¹.

Nous aborderons donc dans un premier temps les définitions et les dispositifs d'étude de l'opinion avant l'avènement des sondages d'opinion (I.1). Puis nous verrons comment les sondages ont transformé l'opinion en « construit mesurable » (I.2). Nous reviendrons ensuite plus en détail sur les critiques universitaires des sondages (I.3) et les dispositifs empiriques alternatifs proposés notamment par les théories délibératives (I.4).

1.1 Les définitions et dispositifs d'étude de l'opinion antérieurs aux sondages

Étudier les définitions et représentations de l'opinion publique *avant* l'avènement des sondages peut être compris de deux manières. On peut d'abord s'intéresser à la façon dont les auteurs de sciences sociales et de philosophie politique concevaient l'objet « opinion publique » avant la création des sondages. On mobilisera à cette fin l'ouvrage de Loïc Blondiaux et son histoire sociale des sondages³³². Comme l'a démontré l'auteur, avant la création des sondages, différentes définitions et représentations de l'opinion coexistaient. L'opinion publique était alors conçue comme une catégorie de philosophie politique (I.1.2) dont la réalité était difficilement saisissable par les sciences sociales (I.1.3). Mais étudier les représentations de l'opinion publique *avant* l'avènement des sondages peut aussi renvoyer à toute une série de travaux historiographiques, majoritairement parus après la création des sondages, qui ont entrepris une généalogie du concept (I.1.1).

³³¹ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

³³² Loïc Blondiaux, *La fabrique de l'opinion*, op. cit.

1.1.1 L'opinion publique comme objet historiographique

L'ouvrage de Jürgen Habermas, *L'espace public. Archéologie de la publicité comme dimension constitutive de la société bourgeoise*³³³, paru pour la première fois en 1962, a ainsi ouvert un vaste champ de recherche, pluridisciplinaire³³⁴, relatif à la formation d'un espace public bourgeois et à l'apparition de la notion d'opinion publique sous l'Ancien Régime. L'auteur y défend une thèse maintenant bien connue : c'est au cours du dix-huitième siècle que s'est constituée en Europe, face à un État absolu agissant en secret, une sphère publique bourgeoise, composée d'individus cultivés, lettrés et capables de raisonner, entendant soumettre le pouvoir absolutiste au « tribunal de la raison », afin que les affaires publiques deviennent réellement « publiques », c'est-à-dire vues et connues de tous. L'adjectif « public » s'oppose ici moins à « privé » ou « intime » qu'à « secret » et « dissimulé »³³⁵. C'est notamment par l'essor de la presse d'opinion, des pamphlets et des libelles que se constitue cette sphère publique, assurant une médiation entre la société et l'État. L'opinion publique est alors conçue comme une force sociale et politique, comme le produit des délibérations rationnelles de cette sphère bourgeoise. L'auteur démontre par la suite comment la sphère publique a été assez vite instrumentalisée par la bourgeoisie pour faire valoir ses intérêts particuliers. Comme le résume bien Patrick Champagne, au dix-huitième siècle, l'opinion publique « *c'est l'opinion que portent sur la politique des groupes sociaux restreints dont la profession est de produire des opinions et qui cherchent à entrer dans le jeu politique en le modifiant et en transfigurant leurs opinions d'élites lettrées en opinion universelle, intemporelle et anonyme, ayant valeur en politique* »³³⁶.

L'ouvrage d'Habermas a par la suite donné lieu à de nombreux travaux et a suscité plusieurs controverses. Il a été notamment reproché à l'auteur de passer sous silence l'existence d'autres sphères publiques, composées de catégories sociales inférieures, exclues de la sphère bourgeoise hégémonique mais mobilisées politiquement et culturellement³³⁷. Des auteurs féministes ont également reproché à Habermas de ne pas spécifier et de ne pas tenir

³³³ Jürgen Habermas, *L'espace public. Archéologie de la publicité comme dimension constitutive de la société bourgeoise*, Payot, 1998.

³³⁴ Regroupant des travaux d'historiens, de philosophes, de sociologues, etc.

³³⁵ Pierre Bourdieu, *La noblesse d'État*, Paris, Éditions de Minuit, 1989, p. 545-548.

³³⁶ Patrick Champagne, *Faire l'opinion*, op. cit., p. 47.

³³⁷ Gunter Lottes, *Politische Aufklärung und Plebejisches Publikum*, Munich, 1979, p. 110. Et Oskar Negt, Alexander Kluge, *Erfahrung und Öffentlichkeit. Zur Organisationsanalyse bürgerlicher und proletarischer Öffentlichkeit*, Francfort, 1972. Cités dans Jürgen Habermas, « "L'espace public" 30 ans après », *Quaderni*, Vol. 18, n°18, 1992, p. 161-191.

compte du caractère patriarcal de la sphère publique³³⁸. À l'occasion du trentième anniversaire de l'ouvrage, l'auteur reconnut d'ailleurs le bien-fondé de ces critiques³³⁹.

En synthèse de ces travaux, l'historien Keith Baker remarque que *« l'aspect le plus intéressant de la notion d'«opinion publique» telle qu'elle apparaît au XVIIIe siècle, c'est précisément de n'avoir jamais été conceptualisée comme plurielle. Définie comme autorité ultime, c'est-à-dire rationnelle, objective et stable, l'opinion publique devait être conçue comme unitaire, comme l'était la monarchie absolue, dont elle endosse les attributs idéaux. Cela ne signifie pas, bien sûr, qu'une opinion publique unitaire existait comme un phénomène social ou politique empirique. En termes sociologiques, il y avait manifestement beaucoup d'«opinions» différentes sous l'Ancien Régime »*³⁴⁰.

L'auteur identifie trois types d'opinion : les « bruits publics », les « voies institutionnelles de production de l'opinion » et les « productions extra-institutionnelles d'opinion ». Les bruits publics font référence aux rumeurs, aux « mauvais discours » formulés par le peuple à propos du roi. Si ces bruits publics ont toujours existé, leur prise en considération par le pouvoir politique est en revanche, au dix-huitième siècle, une nouveauté : *« ces bruits en sont venus à être constitués comme fait politique par la monarchie elle-même, par l'appareil de surveillance et de répression, qui a commencé à chercher à contrôler les «mauvais discours», les examiner, à en prendre connaissance, à les rapporter et... à les craindre ! »*³⁴¹. Les « voies institutionnelles de production de l'opinion » renvoient aux lieux contrôlés par la monarchie, comme les Parlements, les États généraux, les Académies provinciales, etc. Ces lieux concentrent les élites dominantes au service de la monarchie. Enfin l'auteur identifie les « productions extra-institutionnelles d'opinion ». Elles se manifestent sous forme de pamphlets et de libelles et échappent au pouvoir royal. Keith Baker observe qu'au dix-huitième siècle, le pouvoir s'engage également dans ce processus en faisant circuler à son tour des pamphlets pour essayer de produire une « contre-opinion »³⁴².

Il est intéressant de noter que cette opposition conceptuelle théorisée par les historiens entre une opinion publique informelle émanant des masses et une opinion formelle, lettrée et

³³⁸ Catherine Hall, « Private Persons versus Public Someones : Class, Gender and Politics in England, 1780-1850 », in Carolyn Steedman, Cathy Urwin et Valerie Walkerdine (dir.), *Language, Gender and Childhood*, Londres, 1985, p. 10. Et Joan Lands, *Women and the Public Sphere in the Age of the French Revolution*, Ithaca, 1988.

³³⁹ Jürgen Habermas, « "L'espace public" 30 ans après », art. cit.

³⁴⁰ Keith Michael Baker et Roger Chartier, « Dialogue sur l'espace public », *Politix*, Vol. 7, n°26, 1994, p. 5-22.

³⁴¹ *Ibidem*.

³⁴² Cette idée de contre-opinion trouve aujourd'hui un nouvel écho avec les phénomènes de « e-réputation » ou « contre-buzz ». Cf. chapitre 1.

élitaire, est posée en des termes relativement proches dans le débat qui opposa les théoriciens américains de l'opinion publique au début du vingtième siècle.

1.1.2 L'opinion comme catégorie de la philosophie politique avant les sondages

Le titre de ce paragraphe paraphrase le sous-titre du premier chapitre de l'ouvrage de Loïc Blondiaux³⁴³. L'auteur démontre dans son ouvrage qu'une pluralité de définitions concurrentes de l'opinion publique coexistaient avant l'avènement des sondages. Les sciences sociales n'avaient d'ailleurs pas le monopole de la définition et de la représentation de l'opinion publique. Avant d'être un concept sociologique, l'opinion publique était avant tout une catégorie de philosophie politique.

Deux modèles d'analyse concurrents de l'opinion publique coexistent aux États-Unis au tournant du vingtième siècle. Un premier modèle conçoit l'opinion publique comme une force politique exerçant un contrepoids à l'action des gouvernants. Ce modèle a été théorisé par le publiciste anglais James Bryce³⁴⁴. Ignorée en France, Loïc Blondiaux note pourtant que sa théorie de l'opinion a servi de matrice « à la presque totalité de la réflexion anglo-saxonne en ce domaine dans la première partie du siècle »³⁴⁵. Pour le publiciste anglais, c'est lorsque la démocratie atteint son stade le plus avancé que l'opinion publique prend toute sa place. Ce « quatrième stade de la démocratie » correspond à un état de « démocratie continue », où l'opinion du peuple, l'opinion publique, guide et oriente en permanence l'action des gouvernants : « un quatrième stade sera atteint lorsque la volonté de la majorité des citoyens pourra être déterminée à tout moment, et sans la nécessité de passer par un corps de représentants, et si possible même sans la nécessité de la machine du vote. Dans un tel état des choses, l'emprise de l'opinion publique pourrait devenir plus complète, parce que plus continue [...] »³⁴⁶. Deux remarques peuvent être formulées. La première est que l'opinion publique occupe dans la théorie de Bryce une fonction relativement similaire à celle décrite par Habermas : le rôle de « tribunal de la raison » qui juge l'action du gouvernement. La seconde est que ce souhait de prendre en compte la volonté du peuple de façon continue a précisément constitué le discours légitimateur des sondages d'opinion. Si la technologie du

³⁴³ Loïc Blondiaux, *La fabrique de l'opinion*, op. cit., p. 71.

³⁴⁴ James Bryce, *The American Commonwealth*, Londres, McMillan, 1911.

³⁴⁵ Loïc Blondiaux, *La fabrique de l'opinion*, op. cit., p.71.

³⁴⁶ James Bryce, *The American Commonwealth*, op. cit., p. 262-263 cité par Loïc Blondiaux, *La fabrique de l'opinion*, op. cit., p. 75.

sondage n'existait pas à l'époque où Bryce écrit ces lignes, et s'il n'avait pas une idée précise de comment pourrait être « sondée » l'opinion publique, on voit bien se dessiner en creux, de manière presque prophétique, la fonction qui a été attribuée aux sondages : celle d'être un équivalent fonctionnel à la « *machine du vote* » en période non-électorale.

Bryce détaille également les processus de formation de l'opinion publique. Il est l'un des premiers à proposer une théorisation de l'articulation entre opinions individuelles et opinion publique. Le publiciste distingue trois séries d'acteurs qui participent au processus de formation de l'opinion. La première est composée de producteurs d'opinion, qui « font » l'opinion, à savoir les hommes politiques et les journalistes. La deuxième est composée de ceux qui diffusent l'opinion : des citoyens qui témoignent d'un intérêt pour les affaires publiques et qui se regroupent en association. Enfin le troisième groupe est composé de ceux qui n'ont pas d'opinion, des individus qui se désintéressent des affaires publiques mais sur lesquels peut agir le second groupe³⁴⁷. La théorie de Bryce préfigure ainsi le modèle du *two-step-flow of communication*, qui sera conceptualisé et empiriquement validé des décennies plus tard. On retrouvera en effet dans l'analyse de Katz et Lazarsfeld de nombreux éléments développés par Bryce : l'importance accordée aux notions d'intérêt et de niveau d'information du public, le rôle exercé par la presse dans la diffusion de l'information, mais aussi le rôle déterminant joué par la discussion politique, qui permet d'échanger et de diffuser des opinions.

Parallèlement aux théories de Bryce, un deuxième modèle théorique de l'opinion publique se développe aux États-Unis. Ce modèle, théorisé par Graham Wallas et Walter Lippmann notamment, s'inspire des théories de la psychologie des foules de Gustave Le Bon. Lippmann, dans *Public Opinion* puis *The Phantom Public* déconstruit le mythe d'une opinion publique informée et rationnelle, conforme à l'idéal démocratique. L'auteur affirme au contraire que les citoyens, devant la complexité croissante du monde, se trouvent impuissants, étrangers, incompetents : « *comment se convaincre que les affaires publiques sont aussi [celles du citoyen]? L'essentiel lui en demeure invisible. Les lieux où tout se passe sont des centres lointains d'où des puissances anonymes tirent les ficelles derrière les grandes scènes publiques. [...] Aucun des journaux qu'il lit ne décrypte ce monde de manière à le lui rendre intelligible ; [...] et ce n'est pas d'écouter des discours, d'énoncer des opinions et de voter qui le rendent capable pour autant de tenir les commandes, il s'en aperçoit bien. Il vit dans un*

³⁴⁷ Loïc Blondiaux, *La fabrique de l'opinion*, op. cit., p. 79.

monde qu'il ne peut voir, qu'il ne comprend pas et qu'il est incapable de diriger »³⁴⁸. Faute de pouvoir produire un avis raisonné sur le monde, les citoyens s'en remettent alors à des « images préfabriquées », des « stéréotypes » : « *des déformations simplificatrices de la réalité qui permettent à l'individu de penser à moindre coût et de faire l'économie d'une expérimentation de la réalité* »³⁴⁹. Il est intéressant de remarquer à quel point cette définition des stéréotypes rappelle les notions d'« heuristiques » ou de « raccourcis cognitifs » développées des décennies plus tard par la psychologie politique puis la sociologie politique. Pour Lippman, ces images et stéréotypes seraient aisément manipulables. L'auteur insiste alors sur la capacité des médias à influencer l'opinion en déformant et en instrumentalisant ces images préconstruites. L'opinion publique serait donc potentiellement dangereuse, car manipulable.

Ces deux modèles opposent donc deux conceptions de l'opinion publique. Cette opposition s'inscrit dans la problématique plus large de la compétence et de la rationalité des masses, question posée à chaque étape du processus d'ouverture du champ politique. Comme l'a démontré Loïc Blondiaux, l'introduction des sondages d'opinion transformera radicalement les termes du débat en fusionnant ces deux conceptions, en mêlant l'opinion et la masse³⁵⁰.

Avant l'avènement des sondages, la question de l'opinion publique était donc principalement traitée de manière théorique par la philosophie et la théorie politique. Cependant, dès le début du vingtième siècle aux États-Unis des travaux empiriques de sciences sociales entreprirent de mesurer de façon très expérimentale l'opinion publique.

1.1.3 Les dispositifs expérimentaux de mesure de l'opinion avant les sondages

Même si elles n'en avaient pas le monopole, l'opinion publique était, avant l'avènement des sondages, un objet déjà étudié par les sciences sociales. Ces travaux de sociologie, de psychologie et de science politique étaient caractérisés par deux phénomènes : une grande variété de définitions du concept d'opinion et une réelle difficulté à saisir

³⁴⁸ Walter Lippman, *Le public fantôme*, Paris, Demopolis, 2008.

³⁴⁹ Loïc Blondiaux, *La fabrique de l'opinion*, op. cit., p. 95.

³⁵⁰ *Ibid.*, p.44.

empiriquement l'objet³⁵¹. Cependant, des recherches expérimentales, dès les années 1920, entreprirent de mesurer empiriquement l'opinion.

Deux types de dispositifs peuvent être évoqués. Le premier interroge des sources indirectes, des données de seconde main, comme des résultats électoraux ou des articles de presse. Le second, issu de travaux de psychologie sociale, repose sur des données empiriques de première main, récoltées par questionnaire.

Au début du vingtième siècle, la majorité des recherches de sociologie et de science politique mobilisent en effet des données de seconde main pour mesurer l'opinion. Nous n'évoquerons pas ici les tentatives d'analyse de l'opinion à partir de résultats électoraux agrégés³⁵² afin de nous consacrer à l'analyse d'un autre indicateur de mesure de l'opinion : la presse. Ces recherches, pour la plupart aujourd'hui oubliées, trouvent pourtant, avec l'émergence des opinions numériques, un nouvel écho. Nous le verrons, lorsque que l'on cherche à mesurer les opinions politiques publiées sur le réseau social Twitter, de nombreux messages sont en fait des liens vers des articles de presse. Comment traiter ces messages ? Peut-on faire de la mesure de ces articles de presse un indicateur d'opinion ? Cette question trouve donc aujourd'hui une pertinence nouvelle et c'est à ce titre que ces recherches méritent d'être évoquées ici.

Le sociologue américain Hornell Hart entendait ainsi mesurer l'opinion publique en étudiant les articles de presse³⁵³. Le chercheur, afin de saisir l'évolution de l'opinion publique, a recensé un échantillon d'articles de presse relatifs à des questions importantes de société (religion, prohibition, désarmement, etc.), parus entre 1905 et 1931. Il a ensuite, pour chaque question, établi un ratio du nombre d'articles parus sur une question sur le nombre total d'articles parus, pondéré par le volume de chaque publication. Loïc Blondiaux note que « *mieux encore, à partir du contenu des journaux, il a construit un indicateur du degré d'approbation ou de désapprobation de l'opinion publique sur ces différentes questions. Il a pu ainsi fournir un tableau de l'évolution de l'opinion publique au sujet des contrôles des naissances de la façon suivante : 86 % des articles y étaient favorables de 1905 à 1914, 50 % de 1915 à 1918, [...] etc.* »³⁵⁴. Ces dispositifs reposent donc sur l'hypothèse selon laquelle les

³⁵¹ *Ibid.*, p.121.

³⁵² *Ibid.*, p.79.

³⁵³ Hornell Hart, « Changing social attitudes and interests », President's Research Committee on Social Trends, *Recent Social Trends*, New York, McGraw Hill (2 vol.), 1933, vol. 1, p. 382-443 cité par Loïc Blondiaux, *La fabrique de l'opinion*, op. cit., p. 136.

³⁵⁴ Loïc Blondiaux, *La fabrique de l'opinion*, op. cit., p.137.

opinions exprimées par les journaux reflètent celles de leurs lecteurs. Cette hypothèse a été depuis testée empiriquement et a été invalidée par certains travaux³⁵⁵. Néanmoins, cela n'a pas empêché la multiplication des études de presse aux États-Unis dans les années 1920 et 1930. Il est encore une fois saisissant de constater que les problèmes méthodologiques rencontrés par ces études se posent aujourd'hui en des termes identiques pour l'étude des opinions numériques : représentativité des échantillons, équivalence entre la taille des articles et leur importance, correspondance entre les attitudes d'un journal et celles de son lectorat, mesure de l'intensité des opinions exprimées, etc.³⁵⁶

Les questionnaires de psychologie sociale constituent le second type de dispositif de mesure empirique des opinions avant l'avènement des sondages. Ces questionnaires sont apparus dans les années 1920 aux États-Unis, concomitamment à un nouveau courant de recherche consacré à l'étude du concept d'attitude. Les attitudes sont définies comme « *des dispositions acquises et durables qui jouent comme matrices de comportement envers certains objets ou certaines situations* »³⁵⁷. L'opinion est alors définie comme l'expression verbale d'une attitude. Comme le note Loïc Blondiaux, les préoccupations des psychologues sont alors relativement éloignées des problématiques de l'opinion publique. L'objectif de ces recherches est d'étudier le processus de formation des opinions individuelles. Pour ce faire, les chercheurs administrent à des enquêtés, très souvent issus d'un public étudiant, un questionnaire. Les questions portent sur des sujets sociétaux, économiques, sociaux, souvent très clivants (la prohibition, la peine de mort, etc.). Pour chaque question les chercheurs retiennent une série d'énoncés, c'est-à-dire des prises de position courantes, des opinions déjà formulées, à propos du thème abordé. Ces énoncés peuvent être ensuite rangés, ordonnés sur un *continuum*, une échelle. Les étudiants obtiennent donc une note d'échelle, censée correspondre à leur attitude sur le sujet. Les résultats obtenus permettent de « *cartographier l'espace des prises de position sur le monde social* »³⁵⁸, et, pour la première fois, les attitudes comme les opinions peuvent être mesurées. L'opinion publique, lorsqu'elle est définie, est alors présentée comme la sommation des opinions individuelles mesurées. Cette définition préfigure donc la conception sondagière de l'opinion publique. Cependant, jusqu'en 1935, la question de la représentativité des opinions n'est pas encore posée en des termes mathématiques.

³⁵⁵ George Lundberg, « The newspaper and public opinion », *Social Forces*, vol. 4, 1926, p. 709-715.

³⁵⁶ Loïc Blondiaux, *La fabrique de l'opinion*, op. cit., p.137.

³⁵⁷ *Ibid.*, p.131.

³⁵⁸ *Ibid.*, p.132.

Étudier l'objet « opinion publique » avant la création des sondages laisse donc entrevoir une pluralité de définitions et de représentations concurrentes du concept. Nous le verrons, certaines de ces définitions, de ces approches et de ces dispositifs apparaissent parfois davantage adaptés à l'étude des opinions numériques que l'appareillage intellectuel des sondages. L'avènement des sondages a pourtant occulté ces approches, et le fait même qu'une pluralité d'approches puisse exister. Les sondages ont en effet durablement conditionné les représentations des chercheurs.

1.2 L'avènement des sondages : quantification, représentativité et dépolitisation de l'opinion

Nous ne mentionnerons pas ici les conditions ayant contribué au succès commercial, politique et scientifique des sondages ; nous ne reviendrons pas non plus sur les origines de l'instrument³⁵⁹. Nous évoquerons simplement trois changements qui ont durablement modifié les représentations et conceptions de l'opinion publique. Un premier changement est d'avoir fait de l'opinion un objet quantifiable, mesurable. C'est en adoptant une démarche de questionnement, *via* l'administration de questionnaires, que les sondeurs ont pu codifier et donc mesurer les opinions (1.2.1). Un deuxième changement provoqué par l'apparition des sondages a été d'introduire en sciences sociales une problématique nouvelle, celle de la représentativité statistique des opinions étudiées (1.2.2). Enfin, la substitution d'une définition scientifique à une définition jusqu'alors exclusivement politique de l'opinion publique constitue une troisième rupture dans l'étude de l'opinion (1.2.3). C'est précisément cette ambivalence du concept – sa définition à la fois scientifique et politique – qui sera dénoncée par la sociologie critique.

³⁵⁹ Pour ces questions, voir encore une fois l'ouvrage de Loïc Blondiaux, *La fabrique de l'opinion*, *op. cit.*

1.2.1 La transformation de l'opinion en « construit mesurable »

Un des principaux changements provoqués par l'apparition des sondages a été de transformer la notion d'opinion publique de « *concept ambigu* » en « *construit mesurable* »³⁶⁰. Avant la création des sondages, l'opinion publique était avant tout un concept de philosophie politique, dont la réalité empirique échappait aux chercheurs de sciences sociales. Seuls les psychologues sociaux avaient entrepris, avec des questionnaires d'attitude, de mesurer les opinions. Les sondages ont systématisé et rationalisé cette entreprise. Les opinions sont désormais conçues comme des données quantifiables que l'on peut mesurer et additionner. L'opinion publique est définie comme la moyenne des opinions individuelles. L'état de l'opinion peut alors se réduire à un pourcentage.

Cette numérisation – au sens premier du terme – des opinions a constitué un changement fondamental dans la manière de se représenter l'opinion. La présentation d'une opinion sous forme de chiffre, de pourcentage, n'allait en effet pas de soi. Loïc Blondiaux observe ainsi que dans les articles de presse français relatant les résultats des premiers sondages d'opinion, les chiffres étaient systématiquement illustrés par des *verbatim* : « [...] tout se passe dans ces premières années comme s'il était impossible de s'en remettre exclusivement aux chiffres, comme si ceux-ci n'étaient pas suffisamment « parlants » et qu'il était impossible de se confier à ce seul langage, à cette seule rationalité »³⁶¹.

La quantification de l'opinion publique résulte d'une codification des opinions individuelles. Cette codification n'a été rendue possible qu'en interrogeant les enquêtés, en leur administrant un questionnaire. Les « opinions » sont alors réduites à des réponses de questions, ouvertes ou fermées. Avec les questions ouvertes, l'opération de codification se fait *a posteriori*. Les chercheurs n'imposent pas aux enquêtés des modalités de réponse préconstituées. Les réponses se matérialisaient sous forme de *verbatim* bruts qu'il s'agit donc de recoder *a posteriori*. L'usage des questions ouvertes a très vite décliné dans les sondages d'opinion : elles représentaient 12 % des questions en 1936-37, contre seulement 3 % en 1984-85³⁶². En 2006, dans le cas français, Patrick Lehingue ne comptabilisait que 2 % de questions ouvertes et même moins de 1 % dans les sondages politiques³⁶³. Les sondeurs ont

³⁶⁰ Paul Cantrell, « Opinion polling and American democratic culture », *International Journal of Politics, Culture and Society*, vol.5, n°3, 1992, p. 415-416.

³⁶¹ Loïc Blondiaux, *La fabrique de l'opinion*, op. cit., p. 343.

³⁶² *Ibid.*, p. 179.

³⁶³ Patrick Lehingue, *Subunda*, op. cit., p. 149.

donc préféré recourir aux questions fermées et au codage *a priori* des réponses. Les questions fermées sont jugées moins contraignantes et favorisent la standardisation des réponses. Les sondeurs imaginent *a priori*, par anticipation, ce que seront les réponses des enquêtés. Les modalités de réponse proposées aux enquêtés sont alors les réponses que les sondeurs prêtent par anticipation aux enquêtés. Les enquêtés n'ont plus qu'à cocher une case, leur investissement dans la situation d'enquête est réduit au minimum. Comme le résume Loïc Blondiaux, « en *standardisant les réponses, en limitant au strict minimum (un simple acquiescement suffit) l'intervention de la personne interrogée, les instituts de sondage se donnent la possibilité de mettre en équivalence les opinions et ainsi de les quantifier. Qu'ont-ils sacrifié au passage ? La spontanéité, la liberté et l'imagination des personnes interrogées autant que la possibilité de mesurer l'intensité de leurs réponses et de comprendre la signification de leurs propos* »³⁶⁴.

En plus de quantifier et mesurer l'opinion publique, les sondages se sont également attribués la prétention de parler au nom du tout et de prédire les mouvements futurs d'opinion.

1.2.2 Échantillonnage, représentativité et prédictivité de l'opinion

L'apparition de la problématique de la représentativité statistique des opinions mesurées constitue une nouveauté scientifique fondamentale pour les sciences sociales. Cette problématique a durablement impacté les représentations scientifiques de l'opinion. Une des premières interrogations adressées aux chercheurs proposant des dispositifs alternatifs de mesure de l'opinion – et on pense bien sûr ici aux dispositifs de mesure des opinions numériques – est de savoir si le matériau étudié est « représentatif ». La représentativité des opinions ou des faits sociaux étudiés apparaît aujourd'hui comme une contrainte, voire un « impératif »³⁶⁵. Pourtant, avant la création des sondages, cette contrainte n'en était pas une, mieux, elle n'existait pas. Elle n'était du moins pas posée en des termes mathématiques. Or c'est précisément cette contrainte qui a contribué pour partie au succès de l'instrument, en lui attribuant la prétention de parler au nom du tout et surtout de prédire les résultats électoraux.

³⁶⁴ Loïc Blondiaux, *La fabrique de l'opinion*, op. cit., p.179.

³⁶⁵ Alain Desrosières, *La politique des grands nombres. Histoire de la raison statistique*, Paris, Éditions La Découverte, 1993.

Les sondages d'opinion ne se limitent pas en effet à mesurer des opinions individuelles, ils prétendent mesurer l'opinion publique dans son ensemble. Or l'opinion publique n'est pas l'addition des opinions des personnes interrogées par le sondage : elle est l'agrégation des opinions de l'ensemble d'une population. Les sondages prétendent donc « sonder » l'ensemble d'une population en interrogeant seulement un échantillon de cette population. Si cette opération de substitution de la partie au tout semble aujourd'hui aller de soi, il n'en était rien à l'époque de la création de l'instrument : cette opération apparaissait même « hasardeuse » pour beaucoup. Pour des raisons évidentes de coût, les instituts de sondage et les laboratoires de recherche ne pouvaient pas interroger de façon exhaustive l'ensemble de la population, comme c'est le cas lors du recensement national. Les inventeurs des sondages ont donc eu recours à un principe statistique : la loi des grands nombres, dont le premier modèle mathématique fut défini dès le début du dix-huitième siècle par Jacques Bernoulli³⁶⁶.

Selon ce principe statistique, les caractéristiques d'un échantillon aléatoire d'une population se rapprochent d'autant plus des caractéristiques statistiques de la population que la taille de l'échantillon augmente. En revanche, la taille « optimale » à atteindre pour que l'échantillon soit considéré comme représentatif ne dépend pas de la taille de la population totale. Pour sonder une population de 3 000 comme de 600 000 individus, un échantillon de 1000 individus suffit. C'est à la fin du dix-neuvième siècle que s'ouvre un débat dans le champ statistique sur la question de la représentativité et plus précisément sur les méthodes d'échantillonnage. Il oppose les partisans de la méthode de l'« échantillonnage au hasard » aux partisans de la méthode dite « du choix judicieux », que l'on connaît aujourd'hui sous l'appellation de méthode des quotas. Avec la méthode des quotas, les individus qui composent l'échantillon ne sont pas choisis au hasard mais « judicieusement », rationnellement, en fonction de certaines de leurs caractéristiques. Dès la fin des années 1930, la controverse scientifique tourne à l'avantage de la méthode de l'échantillonnage aléatoire et de ses variantes ; la méthode des quotas est alors progressivement abandonnée par les statisticiens³⁶⁷. Pourtant, pour des raisons de coûts et de faisabilité technique, c'est bien la méthode des quotas que les premiers sondeurs ont adopté. Et s'il n'a statistiquement de sens que pour un échantillonnage aléatoire, le calcul des marges d'erreurs sera néanmoins mobilisé comme un

³⁶⁶ Initialement c'est Jacques Bernoulli qui en définit le premier modèle mathématique vers 1690, publié en 1715; Norbert Meusnier, « Argumentation et démonstration de la loi des grands nombres » in Evelyne Barbin (dir.), *La démonstration mathématique dans l'histoire*, Besançon, IREM, 1989, p. 89-97.

³⁶⁷ Loïc Blondiaux, *La fabrique de l'opinion*, op. cit., p. 171.

argument de légitimation scientifique, alors même que la méthode employée est celle des quotas. Les marges d'erreur, permettent tout à la fois de souligner l'aspect perfectible donc scientifique de l'instrument et de légitimer dans le même temps des erreurs éventuelles dans la mesure ou dans l'anticipation des résultats d'une élection³⁶⁸.

Le caractère représentatif des échantillons interrogés autorise en effet les sondages à anticiper ou à « prédire » les mouvements d'opinion et notamment les résultats électoraux. C'est d'ailleurs dans le cadre des élections présidentielles américaines que sont apparus les sondages, et c'est leur capacité à anticiper correctement les résultats des différents candidats qui a conféré à l'instrument sa légitimité scientifique et politique³⁶⁹. Avec les questions d'intention de vote, la technologie du sondage réplique presque à l'identique la technologie du vote. Mais en répliquant la technologie du vote, les sondages en adoptent également tout l'arrière-plan théorique et normatif. Les sondages d'opinion entremêlent donc un discours à la fois scientifique et politique sur l'opinion publique.

1.2.3 Une définition scientifique et politique de l'opinion publique

Avec l'avènement des sondages, pour la première fois, l'opinion publique devient un concept scientifique en plus d'être une notion politique. Le recours aux modèles de représentativité statistique confère à l'instrument un attribut de technicité, de scientificité. Les chercheurs en sciences sociales et les sondeurs professionnels débattent alors de questions statistiques et méthodologiques : quelles sont les meilleures techniques d'échantillonnage ? , faut-il et comment doit-on « redresser » les résultats ? , combien de modalités de réponses doivent être proposées ? , etc.

L'introduction de ce discours scientifique sur l'opinion constitue une nouveauté. Avant la création des sondages, la question de l'opinion publique était uniquement pensée en des termes politiques. Depuis son apparition au dix-huitième siècle, deux problématiques jalonnent l'histoire du concept. La première concerne la compétence politique des masses. La masse des citoyens est-elle en capacité de produire des jugements politiques raisonnés ? Faut-il tenir compte de l'opinion des masses, cette « opinion publique plébéienne » ? Ou doit-on confier les affaires publiques à des groupes sociaux restreints, élitaires, à cette « opinion

³⁶⁸ *Ibidem.*

³⁶⁹ *Ibidem.*

publique capacitaire » ? Avec l'avènement du suffrage universel et l'ouverture du champ politique, la lutte pour parler au nom de l'opinion – l'opinion des masses – vient progressivement se substituer à la question de leur compétence.

L'avènement des sondages d'opinion a parachevé cette substitution de problématique. La question ne se pose plus : l'opinion mesurée par les sondages est celle du plus grand nombre, de l'ensemble des citoyens. Conformément à l'idéal démocratique, tous les sondés sont jugés politiquement égaux et compétents, parfaitement intéressés par les affaires publiques et les questions politiques. Toutes leurs voix, toutes leurs opinions sont donc égales. C'est précisément cette dépolitisation, ou plutôt la relégation du caractère politique de la notion d'opinion publique, que la sociologie critique a cherché à déconstruire.

1.3 La critique universitaire des sondages et de l'opinion publique

C'est dans les années 1970 que se développe en France la critique universitaire des sondages et du concept d'opinion publique. Pierre Bourdieu et Daniel Gaxie notamment ont démontré que l'opinion publique mesurée par les sondages était un artefact. La fonction latente des sondages serait de dissimuler que l'état de l'opinion est un système de forces et de tensions (1.3.1). Cependant, Pierre Bourdieu ne proposa pas un dispositif alternatif plus « adéquat » pour représenter l'opinion afin de rendre compte de façon empirique des tensions qui la parcourent et fut contraint de reconnaître que l'enquête par questionnaire constitue un « *pis-aller* »³⁷⁰ (1.3.2).

1.3.1 L'opinion publique (mesurée par les sondages) n'existe pas

Selon Pierre Bourdieu, les sondages d'opinion reposent sur une philosophie politique implicite qui accorde au peuple une « *connaissance infuse de la politique* »³⁷¹. Tous les citoyens seraient en capacité de produire une opinion politique : « [...] *en sommant toutes les personnes interrogées, sans distinction, de produire une « opinion personnelle » [...] ou de choisir par leurs propres moyens, sans aucune assistance, entre plusieurs opinions toutes*

³⁷⁰ Pierre Bourdieu, *La Distinction*, op. cit, p. 591.

³⁷¹ *Ibid.*, p.463.

préparées, le sondage d'opinion accepte implicitement une philosophie politique [...] qui reconnaît à tous non seulement le droit mais le pouvoir de produire un tel jugement »³⁷². Conformément à l'idéal démocratique, les sondages d'opinion postulent donc que tous les citoyens sont politiquement égaux et compétents.

Pour les sociologues critiques, cette philosophie politique est une pure fiction. Ces auteurs ont démontré empiriquement que tous les individus n'étaient pas compétents politiquement. La compétence politique est une variable inégalement distribuée dans la population. Pour désigner ces inégalités de compétence politique, Daniel Gaxie parlera ainsi de « cens caché »³⁷³, Pierre Bourdieu de « système censitaire méconnu »³⁷⁴. Dans les régimes démocratiques modernes, la compétence politique occuperait une fonction équivalente à celle du cens dans les régimes censitaires. L'auteur de *La Distinction* distingue deux types de compétence politique : une compétence technique et une compétence sociale, statutaire. La compétence technique se caractérise par « *la possession des connaissances savantes et pratiques nécessaires pour produire des actions et des jugements proprement politiques* »³⁷⁵. Le niveau de compétence technique varie principalement en fonction du capital scolaire. Cependant, la compétence technique ne suffit pas à comprendre l'intérêt que les individus peuvent manifester pour la politique. C'est le niveau de compétence statutaire, directement corrélé au niveau de compétence technique, qui détermine la propension à prendre la parole politique, à formuler des avis politiques. La compétence statutaire est ainsi définie comme le « *sentiment (socialement autorisé et encouragé) d'être fondé à s'occuper de politique, d'être autorisé à parler politique, d'avoir autorité pour parler politiquement des choses politiques* »³⁷⁶.

Lors de la situation d'enquête, les inégalités de compétence politique se manifestent en premier lieu par la propension à répondre aux questions politiques. La part des « non-réponses » varie ainsi significativement en fonction du sexe, de l'âge, du niveau d'instruction, de la profession, du lieu de résidence et de la tendance politique ; à tel point que les variations des « non-réponses » constituent pour les sociologues critiques l'information la plus importante fournie par les enquêtes d'opinion. Accepter de répondre à une question politique nécessite donc de se sentir autorisé à formuler une opinion politique et est donc directement

³⁷² *Ibid.*, p.465.

³⁷³ Daniel Gaxie, *Le cens caché*, *op. cit.*

³⁷⁴ Pierre Bourdieu, *La Distinction*, *op. cit.*, p. 464.

³⁷⁵ *Ibid.*, p. 476.

³⁷⁶ *Ibid.*, p. 478.

corrélé au niveau de compétence statutaire des sondés. Cependant, il serait vain de penser que les individus qui acceptent pour leur part de répondre aux questions politiques sont tous compétents politiquement. Beaucoup de réponses enregistrées par les sondages ne sont pas de vraies opinions, à savoir des « *discours constitués prétendant à la cohérence* »³⁷⁷. Ces « non-réponses masquées » peuvent être ainsi de simples réponses de complaisance à la problématique ou encore des discours éthiques non politiquement constitués. Répondre « politiquement » à une question politique nécessite en effet deux conditions préalables. La première est d'être capable de percevoir le caractère politique de la question, ce qui n'est pas évident lorsque la question posée porte par exemple sur des aspects de morale domestique. La seconde est d'être capable de répondre en appliquant des catégories proprement politiques. Toutes les réponses de sondage ne sont donc pas des réponses « politiques ». Les opinions formulées par les enquêtés résultent en réalité de différents modes de production.

Pierre Bourdieu identifie trois modes de production des opinions politiques. Le premier est le recours à l'ethos de classe, à savoir une « *formule génératrice non constituée comme telle qui permet d'engendrer, sur tous les problèmes de l'existence ordinaire, des réponses objectivement cohérentes entre elles et compatibles avec les postulats pratiques d'un rapport pratique au monde* »³⁷⁸. Le deuxième mode de production est le « parti » politique systématique défini comme « *un système de principes explicites et spécifiquement politiques, justiciables du contrôle logique et de l'appréhension réflexive, bref, une sorte d'axiomatique politique [...] qui permet d'engendrer ou de prévoir l'infinité des jugements et des actes politiques inscrits dans l'algorithme* »³⁷⁹. Enfin, le troisième mode de production des opinions est « *un repérage, opéré sur le mode du savoir, des réponses conformes à « la ligne » définie par un parti politique* »³⁸⁰.

Malgré cette diversité des modes de production, les sondages classent dans la même catégorie de réponse et accordent la même valeur « faciale » à des opinions qui expriment en réalité des dispositions très différentes. Et c'est précisément là l'effet le plus « pernicieux » des sondages d'opinion : mettre en demeure les individus de répondre à des questions qu'ils ne se sont pas posées et détourner le sens de leurs réponses en imposant une lecture proprement politique à des « opinions » qui ne sont en fait que des dispositions éthiques.

³⁷⁷ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

³⁷⁸ *Ibid.*, p.490.

³⁷⁹ *Ibid.*, p.491.

³⁸⁰ *Ibid.*, p.491.

Les sondages accordent donc la même valeur à des opinions produites par des individus qui n'avaient aucun avis sur la question avant d'être interrogés et des opinions déjà formulées, défendues et portées politiquement par des groupes sociaux mobilisés. En accordant la même valeur à toutes les opinions, les sondages occultent ainsi que l'opinion publique est « *un système de forces [et] de tensions* » opposant des groupes sociaux en capacité de mobiliser et défendre politiquement des opinions concurrentes.

Comment dès lors mesurer la « force » des opinions, comment objectiver ce système de tensions ? Comment représenter l'état de l'opinion ? Ces questions resteront sans réponse dans la sociologie de Pierre Bourdieu.

1.3.2 Une représentation alternative de l'opinion publique ?

S'il se livra à une critique radicale des sondages et entreprit une déconstruction méthodique du concept d'opinion publique, Pierre Bourdieu ne proposa pas un dispositif alternatif d'étude empirique de l'opinion publique « dissimulée », définie, en reprenant ses termes, comme un système de forces et de tensions.

En suivant la théorie de l'auteur, on aurait pu imaginer un programme de recherche dont l'objectif aurait été d'étudier en période « de crise »³⁸¹ les dynamiques de mobilisation des groupes sociaux et leurs effets de politisation sur les individus peu compétents politiquement, en se focalisant notamment sur le processus de formation des opinions individuelles, lorsque des individus dépolitisés ou peu politisés se trouvent en face d'opinions mobilisées.

Si certains des concepts de Pierre Bourdieu sont mobilisés dans la sociologie des mouvements sociaux³⁸² – comme sa théorie des champs ou celle sur la délégation politique – le sociologue ne formula pas une théorie ou un programme de recherche spécifiquement dédié à l'analyse des mobilisations collectives. Pourtant, on peut supposer que sa conception de

³⁸¹ « Dans les situations où se constitue l'opinion, en particulier les situations de crise, les gens sont devant des opinions constituées, des opinions soutenues par des groupes, en sorte que choisir entre des opinions, c'est très évidemment choisir entre des groupes. Tel est le principe de l'effet de politisation que produit la crise : il faut choisir entre des groupes qui se définissent politiquement et définir de plus en plus de prises de position en fonction de principes explicitement politiques » Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

³⁸² Lilian Mathieu, *Comment lutter ? Sociologie et mouvements sociaux*, Paris, Textuel, 2004

l'opinion publique – définie comme un système de forces et de tensions – aurait pu s'insérer dans une sociologie plus large des mobilisations collectives.

Concernant la mesure des opinions et l'enquête par questionnaire, les écrits de Bourdieu laissent place à une certaine ambiguïté. Cette ambiguïté concerne le statut de l'enquête par questionnaire. L'auteur dissocie les sondages d'opinion – qu'ils soient pratiqués par les instituts commerciaux ou par certains instituts de recherche – et les « enquêtes par questionnaire fermé », mobilisées dans ses propres travaux, et notamment dans *La Distinction*. Si l'on conçoit les sondages comme une certaine catégorie d'enquête par questionnaire, on peut alors trouver surprenant que l'auteur mobilise dans ses recherches un instrument qu'il délégitime scientifiquement. Si l'auteur dissocie bien sondage et enquête par questionnaire, il leur reconnaît cependant des caractéristiques communes, ou plutôt des défauts communs. L'auteur affirme ainsi que l'enquête par questionnaire constitue un « pis-aller », notamment pour étudier les goûts et pratiques culturelles : « *s'agissant de ressaisir des systèmes de goûts, l'enquête par questionnaire fermé n'est qu'un pis-aller, imposé par la nécessité d'obtenir un nombre important d'informations comparables sur une population suffisamment nombreuse pour autoriser le traitement statistique* »³⁸³. Tout comme le sondage d'opinion, l'enquête par questionnaire comporte des risques d'imposition de problématique. C'est d'ailleurs pour cette raison que Pierre Bourdieu « sortit » de son enquête les agriculteurs qu'il avait préalablement interrogés, car jugés beaucoup trop éloignés de la culture dominante. La seule vertu du questionnaire serait donc d'autoriser un traitement statistique des faits sociaux.

Dans « L'opinion publique n'existe pas », Bourdieu reconnaît néanmoins que les sondages peuvent être perfectibles : « *en réalité, l'enquête d'opinion serait sans doute plus proche de ce qui se passe dans la réalité si, transgressant complètement les règles de l'« objectivité », on donnait aux gens les moyens de se situer comme ils se situent réellement dans la pratique réelle, c'est-à-dire par rapport à des opinions déjà formulées ; si [...] on énonçait une série de prises de positions explicites de groupes mandatés pour constituer les opinions et les diffuser, de façon que les gens puissent se situer par rapport à des réponses déjà constituées* »³⁸⁴. En proposant aux enquêtés des énoncés déjà formulés et en spécifiant – différence majeure – quels groupes sociaux les défendent, le sondage ou l'enquête par questionnaire serait d'avantage opérant pour étudier l'opinion.

³⁸³ Pierre Bourdieu, *La Distinction, op. cit.*, p. 591.

³⁸⁴ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

Au final, Pierre Bourdieu, comme d'autres sociologues critiques³⁸⁵, conclue que la meilleure manière de limiter et contrôler les biais méthodologiques et épistémologiques des enquêtes par questionnaire fermé est de compléter et d'enrichir les données récoltées avec d'autres méthodes, comme l'observation ethnographique, les entretiens qualitatifs ou des données secondaires³⁸⁶.

Notons toutefois ici que dans l'un de ses derniers articles, « Le mystère du ministère », Bourdieu esquisse les contours d'un dispositif « politique » – et non pas « scientifique » – idéal de production et de consultation des opinions. Ce dispositif valoriserait une production collective et délibérative des opinions : « *pour échapper à l'agrégation mécanique des opinions atomisées sans tomber dans l'antinomie de la protestation collective – et apporter ainsi une contribution décisive à la construction d'une véritable démocratie –, il faut travailler à créer les conditions sociales de l'instauration d'un mode de fabrication de la « volonté générale » (ou de l'opinion collective) réellement collectif, c'est-à-dire fondé sur les échanges réglés d'une confrontation dialectique [...]*³⁸⁷ ». Loïc Blondiaux a ainsi vu dans ces propos³⁸⁸ une certaine accointance intellectuelle avec les dispositifs de mesure délibérative de l'opinion imaginés presque au même moment, entre autres par James Fishkin.

C'est en plein renouveau théorique des recherches sur l'opinion aux États-Unis que vont naître ces nouveaux dispositifs.

1.4 Les études alternatives de l'opinion

Aux États-Unis, les recherches sur l'opinion connaissent en effet un renouveau à partir des années 1990. Ces recherches s'intéressent dans un premier temps « *à la manière dont les individus bricolent, sélectionnent et agencent des fragments d'information en vue de répondre aux demandes d'opinion qui leur sont adressées* »³⁸⁹. Pour ce faire, ces travaux mobilisent des concepts de psychologie cognitive et privilégient principalement des méthodes qualitatives,

³⁸⁵ Patrick Champagne, *Faire l'opinion*, op. cit.

³⁸⁶ Pierre Bourdieu, *La Distinction*, op. cit., p.594.

³⁸⁷ Pierre Bourdieu, « Le mystère du ministère, Des volontés particulières à la "volonté générale" », *Actes de la recherche en sciences sociales*, Vol. 140, n°140, 2001, p. 7-11.

³⁸⁸ Loïc Blondiaux, « Faut-il se débarrasser de la notion de compétence politique ? Retour critique sur un concept classique de la science politique », *Revue française de science politique*, Vol. 57, n°6, p. 759-774.

³⁸⁹ Loïc Blondiaux, « Les théories contemporaines de l'opinion publique : un retour aux « classiques », *Hermès*, n°31, 2001.

comme les *focus groups*³⁹⁰ ou des approches ethnographiques³⁹¹. L'objectif n'est plus de « mesurer » les opinions, mais seulement d'en étudier la formation (1.4.1).

C'est à la fin des années 1990 que des auteurs vont proposer des méthodes alternatives de *mesure* des opinions. Issus des théories délibératives, ces dispositifs entendent tenir compte, en l'objectivant, du caractère interactif et collectif du processus de formation des opinions. Prenant le contre-pied des sondages traditionnels, ces théories reposent sur le postulat selon lequel le dispositif d'enquête contribue à créer le processus de formation des opinions³⁹² (1.4.2).

Si ces dispositifs semblent faire écho aux vœux « politiques » de Pierre Bourdieu d'une production réellement collective des opinions³⁹³, les sondages délibératifs ne permettent pas en revanche d'objectiver la réalité « empirique » de l'opinion, définie comme un système de forces et de tensions. (1.4.3).

1.4.1. Les approches qualitatives de formation des opinions

Les recherches sur l'opinion connaissent aux États-Unis un renouveau théorique à partir des années 1990. Une nouvelle série de travaux consacrés à la formation des opinions individuelles va en effet voir le jour sous l'influence croisée des recherches de psychologie politique et des nouvelles recherches sur les médias³⁹⁴. Les travaux de Popkin, Sniderman, Zaller ou Gamson vont ainsi renouveler les problématiques relatives aux notions de rationalité et de compétence politique. Après des décennies de controverses, la question posée par ces auteurs n'est plus de savoir si et pourquoi les citoyens sont incompetents, mais de comprendre, étant donné ce manque de connaissance et d'intérêt pour la politique, comment les citoyens arrivent malgré tout à produire des choix, des jugements et des opinions politiques³⁹⁵. Pour répondre à cette question, ces travaux mobilisent des concepts issus de la

³⁹⁰ William Gamson, *Talking politics*, Cambridge University Press, 1992 ; ou dans le cas français Sophie Duchesne et Florence Haegel, « La politisation des conversations au croisement des logiques de spécialisation et de conflictualisation », *Revue Française de Science Politique*, vol. 54, n°6, décembre 2004, p. 877-909.

³⁹¹ Nina Eliasoph, *L'évitement du politique*, *op. cit.* ; Katherine Cramer Walsh, *Talking About Politics*, *op. cit.* ou encore Melissa Victoria Harris-Lacewell, *Barbershops, Bibles and BET*, *op. cit.*

³⁹² Loïc Blondiaux, « Sondages et délibération. Une épistémologie alternative de l'opinion publique ? », art. cit.

³⁹³ Pierre Bourdieu, « Le mystère du ministère, Des volontés particulières à la "volonté générale" », art. cit.

³⁹⁴ Loïc Blondiaux, « Mort et résurrection de l'électeur rationnel. Les métamorphoses d'une problématique incertaine », *Revue française de science politique*, n°5, 1996. p. 753-791.

³⁹⁵ *Ibidem*.

psychologie cognitive des comportements politiques – les « raccourcis cognitifs », les « heuristiques » ou encore les « schémas »³⁹⁶ – et privilégient des approches qualitatives.

Ces travaux rejettent en effet majoritairement l'enquête par questionnaire : elle ne serait pas à même de rendre compte des raccourcis, des heuristiques, en un mot des ressources que mobilisent les enquêtés pour produire une opinion politique. William Gamson privilégie ainsi le recours aux *focus groups*³⁹⁷. L'auteur a formé une trentaine de groupes composés de cinq à six individus issus des classes populaires. Les enquêtés ont été amenés à discuter de différents sujets de société (le nucléaire, l'affirmative action, etc.). À partir de ces matériaux d'enquête, Gamson démontre que les individus qui ne disposent pas d'une connaissance savante de la politique sont quand même capables de « *négozier avec un univers de références qui leur est étranger, celui de la politique, et qu'ils importent, dans ce processus, des ressources qui leur appartiennent en propre* »³⁹⁸. Ces ressources sont de trois ordres : l'expérience personnelle quotidienne, le bon sens populaire et les discours des médias. En France, Sophie Duchesne et Florence Haegel ont également observé, à l'appui de matériaux issus de *focus groups*, que les citoyens les moins compétents étaient en mesure de produire des avis politiques, conflictuels sur l'action publique³⁹⁹.

Depuis le début des années 2000, deux phénomènes caractérisent les études qualitatives sur les opinions politiques. Sous l'influence des travaux pionniers de Robert Huckfield⁴⁰⁰, une première caractéristique est un retour vers une analyse contextuelle et localisée de production des opinions ; là où les travaux évoqués plus haut étaient semble-t-il tombés dans un tropisme psychologique⁴⁰¹. Le deuxième trait caractéristique, conséquence directe du premier, est un recours privilégié à des méthodes ethnographiques. Il ne s'agit plus de savoir comment les citoyens peu compétents produisent des choix et des opinions politiques, mais dans quels contextes sociaux et locaux ils parlent et formulent ces opinions. Nina Eliasoph s'est ainsi demandé « *comment les citoyens créent [...] des contextes propices à la discussion politique dans leur vie quotidienne* »⁴⁰². Pour répondre à cette question, l'auteure a observé durant deux ans et demi trois groupes sociaux : des associations de

³⁹⁶ Paul Sniderman, « The new look in public opinion research », in Ada Finifter (dir.), *Political science: the state of the discipline*, vol. 2, Washington, American Political Science Association, 1993, p. 193-218.

³⁹⁷ William Gamson, *Talking politics*, op. cit.

³⁹⁸ Loïc Blondiaux, « Mort et résurrection de l'électeur rationnel », art. cit.

³⁹⁹ Sophie Duchesne et Florence Haegel, « La politisation des discussions, au croisement des logiques de spécialisation et de conflictualisation », art. cit.

⁴⁰⁰ Robert Huckfeldt, *Politics in Context: Assimilation and Conflict in Urban Neighborhoods*, New York, Agathon Press, 1986.

⁴⁰¹ Loïc Blondiaux, « Mort et résurrection de l'électeur rationnel », art. cit.

⁴⁰² Nina Eliasoph, *L'évitement du politique*, op. cit.

bénévoles, des clubs de loisirs et des groupes d'activistes. À l'issue de son travail de terrain, Eliasoph conclut que les citoyens « *sans être indifférents ni satisfaits, n'expriment des préoccupations politiques que dans les conversations les plus privées. [...] Plus l'arène devient publique, moins le souci du bien commun semble pouvoir s'exprimer légitimement* »⁴⁰³. Katherine Cramer Walsh⁴⁰⁴ et Melissa Victoria Harris-Lacewell⁴⁰⁵ ont adopté une démarche similaire en observant de manière ethnographique des groupes sociaux dans la durée.

Paru avant cette série de recherches ethnographiques, l'ouvrage de John Zaller propose une synthèse des approches cognitives et constructionnistes de l'opinion publique. Le modèle théorique avancé repose sur quatre postulats⁴⁰⁶ : les variables d'attention et d'exposition à l'information politique sont inégalement distribuées dans la population ; la capacité de réaction critique aux arguments développés par les élites et repris par les médias varie en fonction du degré de compétence politique de la personne ; les individus ne possèdent pas d'attitudes stables sur les enjeux sur lesquels ils sont susceptibles d'être interrogés par voie de sondage et développent au contraire une multiplicité de « considérations » à propos de chaque enjeu, lesquels peuvent se révéler contradictoires les uns avec les autres ; ils construisent le plus souvent leurs opinions dans la minute où ils sont confrontés à la question, en mobilisant pour ce faire les considérations qui leur sont pour de multiples raisons les plus immédiatement accessibles à l'esprit. Loïc Blondiaux note ainsi, dans sa présentation de l'ouvrage de Zaller, qu'« *en rappelant par ailleurs que l'opinion, y compris à l'état individuel, n'est nullement un attribut figé et donné une fois pour toutes mais se construit en continu, par la prise en compte d'informations nouvelles, sous la forme d'un échange permanent avec l'extérieur et dans l'interaction, Zaller rejoint, par des voies très différentes, la réflexion sur l'opinion délibérative que mène James Fishkin* »⁴⁰⁷.

Alors que l'analyse de Zaller se base sur une critique des sondages « classiques », en prenant comme matériau empirique des sondages déjà réalisés, comme l'ont fait Converse dans les années 1960 ou Bourdieu et Gaxie dans les années 1970, James Fishkin ou Wilhem Saris et Peter Neijens vont eux proposer des dispositifs empiriques alternatifs d'étude et de mesure des opinions.

⁴⁰³ *Ibidem.*

⁴⁰⁴ Katherine Cramer Walsh, *Talking About Politics*, *op. cit.*

⁴⁰⁵ Melissa Victoria Harris-Lacewell. *Barbershops, Bibles and BET*, *op. cit.*

⁴⁰⁶ Loïc Blondiaux, John Zaller, « Repenser l'opinion. Extrait de *The Nature and Origins of Mass Opinion* », New York, Cambridge University Press, 1192, p. 91-96.

⁴⁰⁷ *Ibidem.*

1.4.2. Les mesures délibératives de l'opinion

À la fin des années 1990, des chercheurs vont en effet entreprendre de mesurer *autrement* l'opinion. Fishkin, Saris et Neijens ne se contentent pas de critiquer les mesures traditionnelles de l'opinion : ils imaginent des dispositifs de mesure *alternatifs*, censés tenir compte et pallier les limites des sondages traditionnels. Ces travaux partagent également certaines des hypothèses développées par les recherches qualitatives de l'opinion. Il est en effet supposé que les individus n'ont pas tous des opinions préconstituées, informées et stables au moment où ils sont interrogés⁴⁰⁸. Au contraire, les individus se forgeraient des opinions de manière continue, au contact d'autres individus, par leur exposition aux médias ou au gré de leur expérience quotidienne. Ces dispositifs alternatifs entendent donc tenir compte de cette réalité. Contrairement aux sondages classiques, qui partent du postulat que la formation des opinions est antérieure au questionnement – le dispositif d'enquête ne doit alors en aucun cas altérer les opinions recueillies – ces dispositifs alternatifs, à l'inverse, reconnaissent implicitement que c'est le dispositif de questionnement lui-même qui contribue à créer le processus de formation de l'opinion⁴⁰⁹. Il s'agit alors d'objectiver ce processus de formation et de cadrage de l'opinion.

Ainsi, l'objectif des sondages délibératifs est de mesurer l'impact d'une délibération de qualité sur la formation de l'opinion publique⁴¹⁰. Le dispositif d'enquête imaginé s'articule en quatre étapes. La première consiste à administrer un sondage « classique » à un échantillon représentatif de la population. Il s'agit là de mesurer l'opinion publique avant le processus de délibération. Il est alors attendu que ce premier sondage enregistre une grande part d'opinions fantômes, non constituées. Les individus interrogés sont ensuite invités, moyennant rétribution, à participer à une expérience démocratique de deux jours. Cette expérience alterne des réunions en petits groupes et des séances plénières avec des « experts ». L'objectif est de sensibiliser, d'informer et d'instaurer un processus de délibération relatif à un enjeu de société. Le compte-rendu de cette expérience est ensuite diffusé à la télévision, afin d'être exposé au plus grand nombre. Enfin, dans un quatrième temps, l'échantillon est sondé une seconde fois, à l'issue du processus de délibération et de sa diffusion télévisée. Au terme de l'enquête, on peut donc mesurer l'évolution de l'opinion avant et après le processus de délibération.

⁴⁰⁸ Loïc Blondiaux, « Sondages et délibération. Une épistémologie alternative de l'opinion publique », art. cit.

⁴⁰⁹ *Ibidem*.

⁴¹⁰ *Ibidem*.

Plusieurs critiques ont été adressées aux sondages délibératifs. Certaines concernent la représentativité des individus qui acceptent de participer à l'expérience et soulignent notamment la sous-représentation des non-diplômés⁴¹¹ ; d'autres questionnent le rôle prééminent accordé aux experts et interrogent leur processus de sélection⁴¹². Une autre critique questionne les causes explicatives des changements d'opinion enregistrés avant et après la délibération. Fishkin infère ces changements à la seule fourniture d'informations supplémentaire aux enquêtés. Or, d'autres facteurs pourraient expliquer ces variations : des effets liés à la dynamique des groupes, des évolutions plus générales de l'opinion, des tendances à aller dans le sens des intentions de l'expérimentateur, etc.⁴¹³.

Mais la principale objection que l'on peut adresser aux sondages délibératifs est de mesurer une opinion publique qui n'existe pas, ou en tout cas pas davantage que celle mesurée par les sondages traditionnels. Cependant, contrairement au cas des sondages, ce constat est parfaitement assumé par les recherches délibératives. Ces dispositifs ne se contentent pas de mesurer l'opinion publique, ils cherchent également à la construire. Les théories délibératives remettent ainsi au premier plan la dimension proprement politique de la notion.

1.4.3 Mesures virtuelles et mesures empiriques de l'opinion publique

Avec les sondages délibératifs, l'objectif de James Fishkin est avant tout politique : il s'agit de pallier l'« indifférence rationnelle »⁴¹⁴ de la plupart des citoyens aux affaires publiques en créant une expérience de démocratie « idéale », en incitant les citoyens-enquêtés à se conduire comme des citoyens « idéaux ». L'expérience délibérative doit être ensuite connue du plus grand nombre, afin que se produise un effet d'émulation : « *si nous diffusions ensuite les résultats de cette expérience à l'intention de tout le pays, cela permettrait peut-être d'intéresser au débat public tous ceux qui ne le sont pas spontanément. Les téléspectateurs pourraient constater que des personnes semblables à eux parviennent à certaines conclusions*

⁴¹¹ David Denver, Gordon Hands et Bill Jones, « Fishkin and the Deliberative Opinion Poll: Lessons from the Study of the Granada 500 TV Program », *Political Communication*, 12, 1995, p. 147-156.

⁴¹² Loïc Blondiaux, « Sondages et délibération. Une épistémologie alternative de l'opinion publique », art. cit.

⁴¹³ *Ibidem*.

⁴¹⁴ James S. Fishkin, « Vers une démocratie délibérative : L'expérimentation d'un idéal » in Stephen L. Elkin et De Karol Edward Soltan (dir.), *Citizen competence and democratic institutions*, Pennsylvania State University Press, 1999, chapitre XII, p. 211

quand elles sont mieux informées »⁴¹⁵. La démocratie « idéale » de Fishkin est inspirée de la démocratie athénienne et des « *town meetings* ». Elle repose sur deux principes fondamentaux : tout le monde peut participer aux réunions publiques et ces réunions doivent se dérouler en face-à-face, afin que de vraies discussions puissent avoir lieu sur les enjeux : « *si la citoyenneté devait être réduite à un vote sans discussion, alors le vote lui-même deviendrait insignifiant* »⁴¹⁶. En sélectionnant aléatoirement un échantillon représentatif de citoyens et en les réunissant dans un même lieu afin qu'ils délibèrent de problèmes publics, les sondages délibératifs contribuent donc à créer cette démocratie et cette opinion publique idéales. Le dispositif de Fishkin repose donc sur un présupposé normatif clairement assumé : l'échange d'informations et la discussion favoriseraient la production de véritables opinions, à savoir des « *opinions informées, fondées sur une évaluation systématique et raisonnée des arguments en jeu* »⁴¹⁷.

Contrairement aux sondages traditionnels, les sondages délibératifs ne prétendent donc pas seulement mesurer l'opinion publique *telle qu'elle est*, mais *telle qu'elle devrait être et telle qu'elle pourrait être*. Le sondage post-délibération mesure ainsi une opinion publique « virtuelle », en puissance : l'opinion publique telle qu'elle pourrait être si les citoyens étaient correctement informés des enjeux. Les sondages délibératifs construisent puis mesurent l'opinion publique là où les sondages traditionnels mesurent l'opinion déjà construite. Comme l'a souligné Loïc Blondiaux, les sondages délibératifs semblent ainsi répondre au vœu « politique » de Pierre Bourdieu de construire un dispositif idéal de production et de consultation des opinions, « *d'un mode de fabrication de la «volonté générale» (ou de l'opinion collective) réellement collectif, c'est-à-dire fondé sur les échanges réglés d'une confrontation dialectique* »⁴¹⁸. Cependant, œuvrer à la production d'un « *mode de fabrication de l'opinion collective* » et à la production d'une opinion publique idéale est un projet très différent de celui de mesurer empiriquement la réalité de l'opinion publique telle qu'elle existe, que Pierre Bourdieu a défini comme « un système de forces [et] de tensions ».

On peut donc, en résumé, distinguer trois « opinions publiques » : une *opinion publique artefactuelle*, mesurée par les sondages d'opinion, une *opinion publique réelle mais dissimulée*, mise au jour par la sociologie critique, à savoir un système de forces et de tensions, et une *opinion publique idéale*, virtuelle, construite et mesurée par des dispositifs

⁴¹⁵ James S. Fishkin, « Vers une démocratie délibérative : L'expérimentation d'un idéal », art. cit.

⁴¹⁶ *Ibidem*.

⁴¹⁷ Loïc Blondiaux, « Sondages et délibération. Une épistémologie alternative de l'opinion publique », art. cit.

⁴¹⁸ Pierre Bourdieu, « Le mystère du ministère », art. cit.

délibératifs expérimentaux. Les dispositifs délibératifs reconnaissent donc, à l'instar de la sociologie critique, le caractère artefactuel de l'opinion mesurée par les sondages traditionnels et « militent » pour la mise en œuvre d'une démocratie délibérative et la production d'une opinion publique « idéale », c'est-à-dire idéalement informée et intéressée par les affaires publiques. Mais les sondages délibératifs ne permettent en rien de représenter et de mesurer l'*opinion publique dissimulée*, ce système de forces et de tensions évoqué par Pierre Bourdieu.

Nous allons maintenant montrer que les autopublications numériques permettent de matérialiser virtuellement – au double sens du terme – cette *opinion publique dissimulée*.

2. Les autopublications numériques, une matérialisation empirique des théories critiques de l'opinion publique

Les autopublications numériques présentent un certain nombre de particularités heuristiques que nous avons détaillées dans le précédent chapitre. Un enregistrement passif de ces messages par des logiciels spécialisés⁴¹⁹ permet de tester empiriquement quatre hypothèses défendues par la sociologie critique. En permettant d'objectiver ces hypothèses, les nouvelles technologies d'étude des autopublications numériques autorisent un possible dépassement des limites épistémologiques des sondages et des enquêtes par questionnaire.

Une première hypothèse développée par la sociologie critique – qui est plus exactement la négation d'une des trois hypothèses sur lesquelles reposent les sondages – est que tous les citoyens ne sont pas en capacité de produire une opinion politique. Nous l'avons vu, la sociologie critique a démontré qu'une grande part des opinions enregistrées par les sondages n'était en réalité pas de véritables opinions, mais des « non-réponses masquées », des réponses de complaisance aux questions des sondeurs. Sur internet, le chercheur collecte uniquement des opinions qui ont une existence matérielle antérieure à l'enquête et qui ne sont donc pas des artefacts générés par le dispositif de recherche. Un enregistrement passif des autopublications permet de plus d'objectiver le mode de production des opinions étudiées et de distinguer ainsi les opinions politiquement constituées et les discours éthiques sur la politique (2.1).

⁴¹⁹ Concernant le logiciel utilisé pour cette recherche, *Ami Opinion Tracker*, Cf. Annexe 6 p. 487.

Corrélativement à cette première hypothèse, Pierre Bourdieu a également défendu que toutes les opinions n'avaient pas la même valeur. La valeur, la « force » ou le « poids » d'une opinion peuvent être mesurés sur internet. Les données chiffrées relatives à l'audience d'un compte ou à la diffusion d'un message peuvent être ainsi mobilisées pour objectiver empiriquement l'inégale valeur des opinions (2.2).

L'absence de consensus entre sondeurs et sondés sur les questions que l'on se pose et qui méritent d'être inscrites dans le débat public constitue une troisième hypothèse défendue par Pierre Bourdieu. En enregistrant les opinions sans jamais les susciter, on étudie sur internet les préoccupations des enquêtés et non celles des enquêteurs ou des chercheurs. (2.3).

Enfin, une posture passive d'enregistrement permet d'objectiver le contexte dans lequel sont produites et énoncées les opinions numériques. Si ces données contextuelles jouent un rôle déterminant dans les approches cognitivistes et ethnographiques des opinions, elles n'apparaissent que secondaires dans l'analyse de Pierre Bourdieu. En suivant les travaux les plus récents de Daniel Gaxie⁴²⁰, nous esquisserons un modèle théorique qui tente d'articuler ces données contextuelles avec l'analyse dispositionnelle des modes de production des opinions politiques. Nous démontrerons alors qu'un enregistrement passif des opinions numériques permet de vérifier empiriquement certaines hypothèses de ce modèle (2.4).

2.1 Mesurer l'inégale capacité à produire une opinion

Le principal postulat des sondages est de considérer que tous les individus possèdent une opinion sur le sujet de l'enquête avant d'être interrogés ou, s'ils n'ont pas réfléchi à la question en amont, qu'ils sont capables de générer une opinion spontanément, « sur le champ ». Pour les sociologues critiques, seule une minorité de la population est capable de produire une opinion politique. Les opinions mesurées par sondage sont donc bien souvent des « non-réponses masquées », des « concessions de complaisance à la problématique » ou des dispositions éthiques naïvement perçues comme des opinions politiques.

En neutralisant le biais d'imposition de problématique, un enregistrement passif des autopublications numériques permet en premier lieu de collecter des « opinions » qui offrent

⁴²⁰ Daniel Gaxie, « Retour sur les modes de production des opinions politiques », in Philippe Coulangeon et Julien Duval (dir.) *Trente ans après La Distinction de Pierre Bourdieu*, Paris, La Découverte, 2013.

la garantie de ne pas être des artefacts générés par la situation d'enquête. Les « non-réponses masquées » n'existent pas sur internet (2.1.1). Cependant, l'absence de questionnement ne signifie pas nécessairement que toutes les opinions collectées sur internet soient de « véritables » opinions politiques. Dit autrement, il peut exister sur internet des discours éthiques publiés *spontanément* sur la politique. En mettant au jour le travail d'énonciation de l'opinion – travail précisément dissimulé par les sondages – les autopublications numériques offrent au chercheur la possibilité d'objectiver le mode de production des opinions numériques et de distinguer ainsi les opinions politiques et les discours éthiques spontanés sur la politique (2.1.2).

2.1.1 Un enregistrement « neutre » des opinions numériques

Les opinions publiées sur les réseaux sociaux, les forums électroniques ou les blogs ne sont pas le produit d'une interaction d'enquête. Plus précisément ces opinions préexistent *matériellement* à l'enquête. Les opinions numériques sont en effet *déjà formalisées* lorsqu'elles se présentent au chercheur. Elles prennent la forme d'un texte écrit, publié. Contrairement aux sondages, la situation d'enquête n'altère donc pas le matériau collecté et ne génère aucun artefact. Sur internet, les technologies d'étude des autopublications numériques n'enregistrent pas de « fausses » opinions ou des « non-réponses masquées » mais du « silence ». L'hypothèse selon laquelle tout le monde n'a pas une opinion sur tous les sujets peut être vérifiée empiriquement de deux manières sur internet : non seulement en mesurant le nombre d'opinions publiées sur un sujet, mais aussi, inversement, en objectivant l'absence d'opinions publiées sur ce sujet⁴²¹.

En enregistrant des opinions qui préexistent *matériellement* à l'enquête, les nouvelles technologies d'étude des autopublications numériques réitèrent donc le « vieux rêve » positiviste d'un enregistrement « neutre » des opinions. Ce « vieux rêve » était – et continue d'être – celui des sondages d'opinion. Près de quatre-vingts ans après sa création, les sondeurs vantent en effet toujours la « neutralité » de leur instrument tandis que les sociologues

⁴²¹ Précisons néanmoins qu'il serait hasardeux de réduire l'absence d'« opinions numériques » sur un sujet à une absence d'opinions « réelles ». En effet, les individus qui ont des opinions politiques – ils ne forment déjà qu'une minorité de la population – ne l'expriment pas tous nécessairement sur internet, loin s'en faut. Les logiques sociales de publication numérique des opinions et les (éventuelles) relations de corrélation entre les opinions exprimées sur internet et les opinions mesurées par sondage ou les opinions exprimées dans le champ de production idéologique ouvrent un vaste chantier de recherche pour les sciences sociales.

critiques s'évertuent – ou s'épuisent – à la déconstruire. On pourrait donc penser que c'est cette même naïveté positiviste qui accompagne l'arrivée de ces nouvelles technologies numériques. Cela serait néanmoins omettre une différence fondamentale avec les sondages : la dissimulation de la situation d'enquête. Comme nous l'avons démontré dans notre précédent chapitre, les individus qui publient une opinion sur internet ne se savent pas « observés » ou étudiés. Par définition, les *verbatim* recueillis par ces nouvelles technologies ne sont donc pas des produits de l'enquête.

Si l'absence de questionnement et la dissimulation de la situation d'enquête garantissent donc que les opinions collectées sur internet ne soient pas des artefacts, il n'est en revanche pas certain que toutes les opinions exprimées sur internet soient politiquement constituées. En mettant au jour le travail d'énonciation de l'opinion, les autopublications numériques permettent de vérifier cette hypothèse, en objectivant le mode de production des opinions collectées.

2.1.2 Objectiver le mode de production des opinions numériques

L'autopublication numérique permet en effet de mettre au jour le travail d'énonciation des opinions – travail précisément dissimulé par les sondages d'opinion : « [en] *faisant disparaître l'exigence d'un travail d'énonciation, [le sondage] présuppose tacitement que la personne interrogée serait capable de produire ou même de reproduire la proposition qui constitue l'énoncé de la question ou même d'adopter spontanément la relation au langage et à la politique que suppose la production d'une telle interrogation [...]* »⁴²². Ce travail d'énonciation constitue l'essence même de la politique en « *transmutant l'expérience en discours, l'ethos informulé en logos constitué et constituant* »⁴²³. En invitant les individus à cocher une modalité de réponse ou à glisser une enveloppe dans une urne, l'enquête par sondage comme les consultations électorales tiennent pour résolue, en la dissimulant, cette opération d'énonciation.

Le travail d'énonciation de l'opinion permet de mettre au jour le langage politique des enquêtés. Pierre Bourdieu distingue trois types de langage politique : le langage dominant, le langage dominé et le langage emprunté. Le langage dominant, « *intrinsèquement euphémisé et*

⁴²² Pierre Bourdieu, *La Distinction, op. cit.*, p. 535.

⁴²³ *Ibid.*, p. 536.

euphémisant, [...] s'impose avec les apparences de l'universalité, déréalise tout ce qu'il nomme [...], met à distance, neutralise [...] produit un discours « d'intérêt général » sur les questions « d'intérêt général »⁴²⁴. Ce discours dominant « détruit, en le discréditant, le discours spontané des dominés : ils ne leur laissent que le silence ou le langage emprunté ». Le langage emprunté est un langage « détraqué » où « les grands mots » ne « sont là que pour marquer la dignité de l'intension expressive, et qui, ne pouvant rien transmettre de vrai, de réel, de « senti », dépossède celui qui le parle de l'expérience même qu'il est censé exprimer »⁴²⁵.

On peut logiquement supposer que chaque catégorie de langage (dominant, dominé, emprunté) est associée à un mode de production des opinions particulier. Un message rédigé dans le langage politique dominant peut être ainsi raisonnablement appréhendé comme le produit d'un « parti politique systématique », cet algorithme générateur de prises de positions savantes sur la politique. À l'inverse, un message formulé dans le langage dominé peut être le produit d'une disposition éthique, un produit de l'ethos de classe.

Contrairement à l'acte de vote⁴²⁶, ou au questionnaire fermé⁴²⁷, les espaces d'autopublication⁴²⁸ offrent aux internautes une véritable capacité expressive. Cette plus grande capacité expressive et l'absence de questionnement permettent de mettre au jour la nature du langage politique et donc le mode de production des opinions publiées. Lorsqu'il entreprend de quantifier et de mesurer les opinions numériques, le chercheur – contrairement au sondeur – a alors connaissance du mode de production des opinions étudiées. Il n'est donc pas contraint d'attribuer une « valeur faciale » identique à des opinions produites selon des logiques sociales très différentes. Plus précisément, l'opération de codification, préalable indispensable à la quantification, peut être réalisée *a posteriori*, en aval de la phase de collecte des opinions. En codant des *verbatim a posteriori*, et non *a priori* – c'est-à-dire par anticipation – le chercheur peut ainsi objectiver empiriquement les différents modes de production des opinions étudiées. En dissociant les opinions politiquement constituées et les discours éthiques, le chercheur peut *in fine* objectiver la capacité de l'internaute à produire une opinion politiquement constituée.

⁴²⁴ *Ibid.*, p. 538-539.

⁴²⁵ *Ibid.*, p. 538.

⁴²⁶ Patrick Lehingue, *Le vote. Approches sociologiques de l'institution et des comportements électoraux*, Paris, La Découverte, 2011, p. 49.

⁴²⁷ Loïc Blondiaux, *La fabrique de l'opinion*, op. cit.

⁴²⁸ Il convient néanmoins de préciser que certains de ces espaces imposent une restriction du nombre de caractères des messages publiés. C'est notamment le cas du réseau social Twitter, notre terrain d'enquête.

Cependant, tout comme pour le traitement des questions ouvertes, qui obéit à la même logique⁴²⁹, le codage *a posteriori* des *verbatim*, notamment lorsqu'il repose sur une analyse lexicométrique, n'est pas une garantie en soi d'objectivation sociologique plus fine des opinions. Dit autrement, comptabiliser abstraitement des mots-clés ne suffit pas à caractériser une intensité expressive ou un degré de sophistication des opinions politiques. C'est précisément ce qu'ont reproché Donald Kinder et David Sears aux chercheurs de l'École de Michigan⁴³⁰. Campbell et ses collègues avaient en effet entrepris de mesurer l'intensité « idéologique » des enquêtés à partir d'une analyse de leurs réponses à des questions ouvertes. Pour ce faire, les auteurs comptabilisaient des références à des termes idéologiquement connotés. La simple comptabilisation de ces termes était ainsi appréhendée comme un indicateur de conceptualisation idéologique. Or, pour Kinder et Sears, « *l'usage d'un vocabulaire idéologique ne garantit en rien que les idées sous-jacentes soient comprises en profondeur ou même que les termes soient correctement utilisés* »⁴³¹.

Cette mise en garde peut être également adressée aux chercheurs souhaitant objectiver le langage politique ou le mode de production des opinions numériques. Le simple décompte d'un nombre de termes politiques jugés « savants » ne permet pas de distinguer le langage politique dominant et le langage emprunté, « détraqué ». Par définition le langage emprunté mobilise des « mots-clés » savants, des bribes de langage dominant (dont la seule fonction est de « *marquer la dignité de l'intension expressive* »⁴³²). Seule une analyse très fine de cooccurrences de termes pourrait sans doute différencier le langage emprunté du langage dominant, en étudiant statistiquement des associations de mots-clés savants avec des marqueurs du langage dominé. Avec les opinions numériques, la qualité de l'enquête ne se joue pas en amont – en formulant de façon la plus « neutre » possible les questions – mais en aval, en objectivant finement les modes de production des opinions collectées.

En résumé, contrairement aux sondages qui « *suppose[nt] que tout le monde peut avoir une opinion ; ou, autrement dit, que la production d'une opinion est à la portée de tous* »⁴³³, les nouvelles technologies d'étude des autopublications numériques permettent non seulement d'objectiver la présence ou l'absence d'opinions publiées sur un sujet politique

⁴²⁹ À la différence près que les réponses à des questions ouvertes restent la réponse à des questions alors que les autopublications n'ont pas été suscitées par les questions du chercheur.

⁴³⁰ Loïc Blondiaux, « Mort et résurrection de l'électeur rationnel », art. cit., p. 769.

⁴³¹ Donald Kinder et David Sears, « Public opinion and political action » in Gardner Lindzey et Eliot Aronson (dir.), *The Handbook of social psychology*, New York, Random House, 1985, p. 666.

⁴³² Pierre Bourdieu, *La Distinction*, op. cit., p. 538.

⁴³³ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

mais également de distinguer les opinions politiquement constituées et les discours éthiques spontanés sur la politique ; en un mot d'objectiver empiriquement la capacité des internautes à produire une opinion politiquement légitime.

2.2 Mesurer l'inégale valeur des opinions numériques

Un deuxième postulat implicite des sondages est de considérer que toutes les opinions se valent⁴³⁴. La technologie sondagière est d'ailleurs calquée sur la technologie du vote. Une réponse égale une opinion ; une opinion égale une voix. Toutes les opinions « sondées » se voient donc attribuer un poids identique. Si cette conception arithmétique des opinions revêt une certaine pertinence dans le cas des sondages d'intention de vote – précisément parce que l'élection repose sur cette même logique de comptabilisation arithmétique des voix – elle pose davantage problème lorsqu'il s'agit de saisir l'état de l'opinion sur des enjeux politiques ou sociétaux.

En dehors de l'isoloir, toutes les opinions n'ont en effet pas la même valeur, la même force réelle. Pour Pierre Bourdieu, une opinion est une idée-force, une idée qui n'a d'existence réelle qu'en étant portée politiquement par un groupe et incarnée par un porte-parole. Or le sondage fait « *exister des opinions qui ne préexistaient pas à l'interrogation et qui ne se seraient pas exprimées autrement ou qui, exprimées autrement, c'est-à-dire par l'intermédiaire de porte-parole attitrés, auraient été toutes différentes ; des opinions qui, en tout cas, avaient peu de chance de se formuler spontanément, de se manifester, la manifestation étant une des manières de donner force aux opinions en manifestant le groupe qu'elle mobilise* »⁴³⁵.

À la fin des années 1970, quand Pierre Bourdieu écrit ces lignes, les agents dominés du champ politique n'avaient en effet aucun moyen de formuler et de *manifester spontanément et publiquement* une opinion – ou même une disposition éthique sur la politique – en première personne⁴³⁶. Ils ne pouvaient alors faire « entendre » leurs opinions – s'ils en avaient – que « par procuration », en s'en remettant à des porte-paroles attitrés. Seuls les

⁴³⁴ *Ibidem*.

⁴³⁵ Pierre Bourdieu, *La Distinction, op. cit.*, p. 484.

⁴³⁶ On pourrait citer comme contre-exemples le courrier des lecteurs ou les participations aux émissions télévisées ou radiophoniques. Cependant, dans ces deux cas, il existe un filtre éditorial. La prise de parole n'est pas spontanée mais autorisée.

agents dominants du champ politique, en ayant accès au champ de production idéologique, pouvaient faire directement et personnellement entendre leurs opinions.

Avec les espaces d'autopublication numérique, tous les individus, quelle que soit leur position sociale, ont la possibilité technique⁴³⁷ d'exprimer une opinion ou un sentiment éthique sur la politique. Ainsi, sur le réseau social Twitter, les individus très peu politisés et les agents du champ de production idéologique – journalistes, éditorialistes, hommes politiques, etc. – disposent de la même capacité technique⁴³⁸ d'expression et de publication. Il ne s'agit pas de sous-entendre que Twitter ait ouvert le champ de production idéologique aux agents dominés ou même que Twitter ait élevé par magie le niveau général de compétence politique de la population. Tout au contraire, nous soutenons ici que les espaces d'autopublication, en réunissant sur un même espace des agents très inégalement politisés et en fournissant aux chercheurs des données relatives à l'audience d'un compte ou d'un message, permettent d'objectiver l'inégale valeur des opinions publiées sur le réseau.

En effet, nous l'avons vu dans le précédent chapitre, les espaces d'autopublication, et Twitter en particulier, génèrent automatiquement et en temps réel des données chiffrées relatives au nombre d'abonnés d'un compte, ou encore au nombre de reprises d'un message (nombre de retweets). Ces données ont été majoritairement investies par des travaux de « marketing viral »⁴³⁹ afin d'objectiver l'« influence » des internautes ou la « viralité » de leurs messages. Nous soutenons ici que ces données peuvent être mobilisées comme des indices permettant d'objectiver la valeur d'une opinion. Une opinion publiée par un éditorialiste reconnu n'a pas le même poids qu'une « opinion » formulée sur un mode éthique par un individu peu politisé. Cette proposition n'est en soi pas très surprenante. Mais sur internet, elle peut être objectivée statistiquement. On peut rapporter chaque opinion publiée à l'audience de son producteur et/ou à celle de ses diffuseurs (les individus qui retweetent le message), mais également au nombre de reprises globales du message (nombre de retweets).

Cette proposition théorique d'une mesure « quantifiée » de la valeur des opinions peut néanmoins apparaître comme un contre-sens si l'on se réfère à la définition que donne Pierre Bourdieu des « idées-forces ». Les idées-forces sont défendues et incarnées par des groupes

⁴³⁷ Ce qui ne signifie pas que tous les individus soient en capacité d'exploiter cette capacité technique. Il faut avoir un minimum de familiarité avec l'outil informatique pour cela. Voir sur ce point Fabien Granjon, *Reconnaissance et usages d'Internet. Une sociologie critique des pratiques de l'informatique connectée*, collection Sciences Sociales, Presses des Mines, Paris, 2012.

⁴³⁸ Il convient de clairement dissocier la capacité technique de la capacité politique (la compétence statutaire) qui est, elle, totalement indépendante de la capacité technique.

⁴³⁹ Cf. chapitre 1 ou Kévin Mellet, « Aux sources du marketing viral », art. cit.

animés par une dynamique proprement politique, collective. Or la proposition d'une mesure quantifiée de la valeur des opinions repose quant à elle sur une logique arithmétique et surtout individuelle de production des opinions. Les données chiffrées disponibles sur les espaces d'autopublication n'indiquent pas nécessairement *la capacité de mobilisation* proprement politique des individus qui partagent et diffusent ces messages. Une opinion retweetée par dix mille internautes n'a pas de force réelle si elle n'est pas portée et défendue politiquement. Autrement dit, sur internet, des opinions *spontanées*, publiées, ne sont pas nécessairement des opinions *mobilisées*⁴⁴⁰. Seule une analyse qualitative, « manuelle », peut permettre d'objectiver cette capacité de mobilisation politique et alors, le cas échéant, donner sens à une analyse quantifiée de la valeur des opinions.

2.3 Mesurer les préoccupations des enquêtés plutôt que celles des enquêteurs

Un troisième et dernier postulat implicite des sondages est de considérer qu'il existe un consensus entre sondeurs et sondés sur les questions qui se posent et qui méritent d'être inscrites dans le débat public. Dit autrement, les questions posées par les sondages d'opinion seraient des questions que se posent personnellement chacun des sondés. Ces questions reflèteraient donc les préoccupations de l'« opinion publique ».

Là encore, Pierre Bourdieu a démontré que cette hypothèse était sociologiquement infondée⁴⁴¹. Les questions posées par les sondages d'opinion reflètent en réalité bien davantage les préoccupations des sondeurs et de leurs commanditaires que celles des sondés⁴⁴². Les sondés, qui sont dans leur majorité des agents dominés, ne partagent pas les préoccupations des agents dominants du champ politique. Plus les questions posées ont trait à des questions proprement politiques, plus l'imposition de problématique est forte. Si les questions relatives au rôle du Président de la République préoccupent les agents du champ politique ou plus vastement les agents du champ de l'opinion, elles n'ont guère de réalité et ne revêtent guère de sens pour des agents peu voire pas du tout compétents politiquement.

⁴⁴⁰ Alors que pour Bourdieu, « *l'opinion qui s'affirme comme telle, spontanément, c'est l'opinion des gens dont l'opinion a du poids* », Cf. Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

⁴⁴¹ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

⁴⁴² Pour Patrick Lehingue, recenser les questions posées par les sondages d'opinion sur un temps long peut ainsi constituer un matériau d'étude intéressant pour les futurs historiens. Ce matériau fournit en effet une indication sur la façon dont on pose et cadre les questions qui animent le débat public, et plus intéressant encore, sur la façon dont évolue le cadrage de ces débats au fil des années. Cf. Patrick Lehingue, *Subunda. Coups de sonde dans l'océan des sondages*, op. cit.

Contrairement aux deux précédentes limites évoquées dans cette section, cette troisième critique ne peut être formulée qu'à l'encontre des sondages « commerciaux ». Si l'on peut légitimement reprocher à des sondages « universitaires » de considérer comme universelle la capacité à produire une opinion politique ou de postuler que toutes les opinions ont le même poids, on peut en revanche plus difficilement leur tenir rigueur de ne pas refléter les préoccupations des sondés. La fonction de la recherche scientifique n'est pas de refléter les préoccupations des agents sociaux. Les questions posées dans un sondage universitaire doivent permettre de tester des hypothèses de recherche et non de traduire au mieux les attentes des enquêtés⁴⁴³. En revanche, si la problématique de recherche consiste précisément à saisir les préoccupations des agents sociaux, un enregistrement passif des autopublications numériques peut alors constituer un dispositif d'enquête plus pertinent que le questionnaire fermé⁴⁴⁴.

En effet, avec le questionnaire fermé, le chercheur recense *a priori* une liste de thématiques sociétales, sociales, économiques et politiques dont il suppose – dont il anticipe – qu'elles correspondent aux préoccupations des sondés. Par définition, cette liste limite le champ des pensables politiquement. Plus encore, elle circonscrit et réduit les préoccupations des sondés aux préoccupations *anticipées* par le chercheur. Cette opération de codification *a priori* est alors très « périlleuse » sociologiquement, car très sensible à un potentiel risque d'ethnocentrisme du chercheur. Les préoccupations anticipées par le chercheur peuvent être avant tout les « préoccupations » mises à l'agenda médiatique par les agents du champ de production idéologique.

En ne reposant pas sur une démarche de questionnement, les technologies d'étude des autopublications numériques neutralisent ce biais potentiel d'ethnocentrisme et enregistrent les préoccupations des enquêtés et non celles des chercheurs. Cette affirmation méthodologique doit être néanmoins nuancée. Lorsqu'il souhaite objectiver et mesurer les préoccupations exprimées par les internautes, le chercheur se trouve face à des millions de messages. Il doit donc effectuer une recherche par mots-clés pour filtrer ce corpus et affiner sa recherche. Le choix des mots-clés est alors déterminant. Le chercheur se trouve à cette étape dans une position similaire à celle du sondeur qui doit recenser par anticipation les préoccupations des sondés. Cela étant dit, alors que la liste des préoccupations arrêtée par le

⁴⁴³ On peut néanmoins reprocher aux sondages universitaires de répondre à une demande sociale extérieure qui subordonne les intérêts proprement scientifiques de la recherche.

⁴⁴⁴ Pour étudier la population qui exprime des opinions sur internet. Cette population est bien sûr beaucoup plus réduite que la population totale.

sondeur est limitée et irréversible⁴⁴⁵, le choix des mots-clés est potentiellement illimité et peut être redéfini et affiné à n'importe quelle étape de la recherche⁴⁴⁶. Le chercheur peut alors « avancer à tâtons », en testant une multitude de mots-clés ou différentes combinaisons de mots-clés pour définir les contours d'une thématique qu'il juge pertinente pour l'analyse.

Le choix des mots-clés renvoie à la question plus générale du langage politique des agents sociaux. Comme nous l'avons évoqué précédemment, le langage politique des dominants n'a pas grand-chose à voir avec celui des dominés : tout oppose en effet le langage euphémisé et abstrait d'« intérêt général » des dominants et le langage très particularisé, « sensible » et fonctionnel des dominés. Pendant longtemps le langage dominé – écrit⁴⁴⁷ – a été réduit au « silence »⁴⁴⁸ : les agents dominés n'avaient aucun moyen de publier des messages écrits pouvant être lus potentiellement par un large public. Avec les espaces d'autopublication numérique, le langage dominé revêt une matérialité qu'il n'avait pas auparavant.

Lorsqu'il entreprend de mesurer les préoccupations des agents sociaux sur internet, le chercheur doit donc commencer par objectiver sa propre distance au langage politique dominant – auquel il a toutes les chances d'appartenir. S'il souhaite « mesurer » la thématique « emploi et chômage », le chercheur doit commencer par définir une liste de mots-clés et d'expressions. Le choix de ces mots-clés n'est pas neutre socialement. À titre d'exemple, les mots-clés et expressions suivantes peuvent apparaître *a priori* comme « neutres » pour le chercheur : « plein emploi », « hausse du chômage », « emplois d'avenir », « emplois aidés », « travail dissimulé », etc. C'est en effet dans ces termes que se pose la question de l'emploi dans le débat public, et c'est sûrement dans ces termes que le chercheur appréhende lui-même cette question. Cependant, il y a de fortes chances que ces mots-clés renvoient à des autopublications « dissertatives », à des réflexions macro-économiques, à des considérations d'« intérêt général » très proches des préoccupations des agents du champ de production idéologique. Ces mots-clés *a priori* « neutres » sont en fait marqués socialement et appartiennent au vocabulaire du langage politique dominant. Au contraire, des expressions du type – « perdu mon taf », « c'est la crise », « job », « missions au black » – renverront très

⁴⁴⁵ À part s'il est prévu une modalité « autres » posée en ouvert et qui fasse ressortir des thématiques saillantes.

⁴⁴⁶ Cf. chapitre 1 sur le nouveau rapport au temps induit par le web.

⁴⁴⁷ Le langage oral pouvait lui s'exprimer sur certains supports médiatiques, comme par exemple la parole donnée au « peuple » à la télévision cf. Vincent Goulet, *Médias et classes populaires. Les usages ordinaires des informations*, Paris, INA Éditions, 2010 ou Peter Livingstone et Sonia Lunt, *Talk on Television: Audience Participation and Public Debate*, Londres, Routledge, 1994.

⁴⁴⁸ Pierre Bourdieu, *La Distinction, op. cit.*, p. 536

probablement à des messages relatant une expérience personnelle, des messages très particularisés – typiques du langage politique des dominés.

Concernant l'étude du langage dominé sur internet, une attention toute particulière doit être accordée à la question de l'orthographe. Si les fautes d'orthographe⁴⁴⁹ apparaissent comme un problème technique à résoudre pour le traitement automatique du langage⁴⁵⁰, elles constituent en revanche pour le sociologue un indice précieux : elles sont en effet révélatrices du niveau du capital scolaire, culturel et linguistique des internautes. Une recherche par mots-clés, si elle se donne pour objectif de « sonder » les messages publiés par des agents dominés, ne peut donc faire l'économie d'un travail de recension des fautes d'orthographe les plus fréquemment associées à certains termes, comme par exemple, nous le verrons, les noms des hommes politiques.

Une fois les opérations de filtrage par mots-clés et de constitution des corpus effectuées, le chercheur peut ensuite, de manière inductive, codifier les messages et ainsi mesurer les préoccupations ou les intérêts des internautes. Cette objectivation permet alors de vérifier empiriquement de nombreuses hypothèses « classiques » de la science politique, dont celle notamment défendue par Philip Converse dans les années 1960 : l'hypothèse des publics spécialisés ou *issue publics*⁴⁵¹. Selon le politiste américain, il existerait différentes sous-populations de citoyens spécialisées dans différents domaines. Les individus ne seraient pas compétents dans tous les domaines, mais seulement dans un domaine particulier, leur domaine de prédilection. Pierre Bourdieu, dans une toute autre perspective théorique, postulait que les agents ont « *d'autant plus d'opinions sur un problème [qu'ils sont] plus intéressé[s] par ce problème* »⁴⁵². Les enseignants ont ainsi par exemple plus d'opinions que la moyenne sur le système d'enseignement, faisant eux-mêmes partie de ce système. On peut donc objectiver cette hypothèse sur internet en rapportant les préoccupations ou les intérêts des internautes à leur position dans l'espace social.

⁴⁴⁹ Rappelons, en nous référant une nouvelle fois à Pierre Bourdieu, que l'orthographe, c'est-à-dire : « *la graphie droite, désignée et garantie comme normale par le droit, c'est-à-dire par l'État* » est un « *artefact social, très imparfaitement fondé en raison logique et même linguistique, qui est le produit d'un travail de normalisation et de codification opéré par l'État* ». cf. Pierre Bourdieu, « Esprit d'État », *Actes de la recherche en sciences sociales*, Vol. 96, n°96-97, 1993, p. 49-62.

⁴⁵⁰ Voir par exemple Olivier Kraif et Claude Ponton, « Du bruit, du silence et des ambiguïtés : que faire du TAL pour l'apprentissage des langues ? », *Actes de TALN* (Toulouse), 2007.

⁴⁵¹ Philip Converse, « The nature of belief systems in mass publics », in David Apter (dir.), *Ideology and discontent*, New York, New York: Free Press, 1964.

⁴⁵² Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

Les autopublications numériques et leur enregistrement par des logiciels spécialisés permettent donc de vérifier empiriquement les contre-hypothèses de Pierre Bourdieu et de dépasser ainsi les trois limites épistémologiques des sondages. Mais les autopublications numériques permettent également de valider certaines hypothèses développées par les approches cognitivistes et ethnographiques.

2.4 Contextualiser la production des opinions numériques

Une posture passive d'enregistrement permet enfin d'objectiver le contexte dans lequel sont produites et énoncées les opinions numériques. Sur internet, les publications des internautes sont en effet systématiquement horodatées. Le chercheur sait donc précisément à quel moment un individu formule et publie son opinion. Il sait de plus si l'opinion publiée s'insère ou non dans un échange collectif et peut, le cas échéant, analyser cet échange, objectiver le nombre de participants, leur niveau de politisation, etc.

Ces éléments de contexte jouent, nous l'avons vu, un rôle déterminant dans les approches cognitivistes. Pour Zaller, « *l'opinion, y compris à l'état individuel, n'est nullement un attribut figé et donné une fois pour toutes mais se construit en continu, par la prise en compte d'informations nouvelles, sous la forme d'un échange permanent avec l'extérieur et dans l'interaction* »⁴⁵³. Dans ce modèle, le contexte – à la fois conjoncturel et interactionnel – interfère donc directement sur la formation des opinions politiques. Dans les approches ethnographiques, le contexte joue également un rôle primordial. Nina Eliasoph a ainsi démontré que les citoyens américains ne dévoilaient leurs opinions politiques que dans des contextes situationnels très particuliers⁴⁵⁴.

En revanche, dans les travaux de Pierre Bourdieu, les éléments *contextuels* apparaissent comme secondaires. L'analyse des modes de production des opinions politiques est avant tout *dispositionnelle*. La capacité à produire une opinion politique dépend de la compétence statutaire des agents. Le contexte dans lequel ils produisent leurs opinions importe donc peu. Dans les travaux de Bourdieu, les opinions politiques n'ont été en fait étudiées que dans un seul contexte : celui de la situation d'enquête par sondage.

⁴⁵³ Loïc Blondiaux et John Zaller, « Repenser l'opinion. Extrait de *The Nature and Origins of Mass Opinion* », New York, Cambridge University Press, 1992, p. 91-96.

⁴⁵⁴ Nina Eliasoph, *L'évitement du politique*, op. cit.

Pierre Bourdieu fait néanmoins allusion dans *La Distinction* à des dimensions contextuelles, extra-dispositionnelles⁴⁵⁵. Dans « L'opinion publique n'existe pas », l'auteur évoque l'effet de politisation des « situations de crise », des situations donc conjoncturelles⁴⁵⁶. Selon Daniel Gaxie, des éléments de nature contextuelle peuvent présider à l'activation des modes de production des opinions politiques⁴⁵⁷.

En suivant la démarche de Daniel Gaxie, nous tenterons ici d'esquisser un modèle théorique qui intègre des éléments de nature contextuelle à l'analyse dispositionnelle des modes de production des opinions politiques (2.4.1). Nous démontrerons ensuite qu'une posture passive d'enregistrement des opinions numériques permet de vérifier empiriquement certaines hypothèses de ce modèle (2.4.2).

2.4.1 Intégrer des éléments contextuels à l'analyse dispositionnelle des modes de production des opinions politiques

Dans l'une de ses plus récentes contributions⁴⁵⁸, Daniel Gaxie questionne l'exhaustivité de la typologie des modes de production des opinions politiques exposée dans *La Distinction* et propose de la compléter. Dans sa démonstration, l'auteur avance et « durcit » deux hypothèses présentes selon lui en filigrane dans les écrits de Bourdieu. La première est que les modes de production ne seraient pas des catégories exclusives : il serait en effet plus juste selon lui de raisonner en termes de continuum de dispositions ou de modes de production mixtes (ou composites)⁴⁵⁹. Les agents sociaux n'activeraient pas nécessairement toujours le même mode de production des opinions et pourraient de plus activer plusieurs modes de production en même temps. La seconde hypothèse, qui complète en partie la première, est qu'il existerait des conditions *contextuelles* – et non uniquement *dispositionnelles* – qui présideraient l'activation d'un mode de production donné.

⁴⁵⁵ Daniel Gaxie, « Retour sur les modes de production des opinions politiques », *op. cit.*

⁴⁵⁶ « Dans les situations où se constitue l'opinion, en particulier les situations de crise, les gens sont devant des opinions constituées, des opinions soutenues par des groupes, en sorte que choisir entre des opinions, c'est très évidemment choisir entre des groupes. Tel est le principe de l'effet de politisation que produit la crise », Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

⁴⁵⁷ Daniel Gaxie, « Retour sur les modes de production des opinions politiques », *op. cit.*

⁴⁵⁸ *Ibidem.*

⁴⁵⁹ L'auteur cite le passage suivant : « tous les jugements politiques comportent une part inévitable de fides implicita », Pierre Bourdieu, *La Distinction*, *op. cit.*, p.497-498.

Comme Daniel Gaxie, nous pensons que des éléments *contextuels* peuvent expliquer les conditions d'activation de la production des opinions politiques. Nous compléterons cette proposition en dissociant les éléments qui relèvent d'un *contexte général* ou conjoncturel et les éléments qui relèvent d'un *contexte situationnel* ou interactionnel (2.4.1.2). Cependant, nous ne proposons pas de compléter la typologie de Bourdieu mais de ne retenir au contraire que deux modes de production des opinions politiques (2.4.1.1).

2.4.1.1 Modes de production et facteurs de production

Daniel Gaxie propose en effet de compléter les trois modes de production exposés par Pierre Bourdieu – l'expression d'un « parti politique systématique », l'activation de l'« ethos de classe » et le choix à deux degrés – en les subdivisant en plusieurs modes distincts. Ainsi, derrière l'ethos de classe se cacheraient en réalité deux modes de production : un mode de production reposant sur une disposition éthique et un mode de production « *privatif à distance* », qui a trait plus directement aux conditions matérielles d'existence. De manière analogue, le « parti politique systématique » pourrait être aussi dissocié en deux modes distincts : la *production* en première personne et la *reproduction* en première personne. Le second repose sur la *reproduction* d'arguments qui circulent dans les milieux politiques et journalistiques, le premier sur la production d'arguments politiques « inédits », produits par des agents propriétaires de leurs petits moyens de production ; des agents majoritairement issus du champ de production idéologique.

Pour Daniel Gaxie, « *tout indique que tous les agents sociaux activent le plus souvent plusieurs modes de production en même temps* »⁴⁶⁰. Pour illustrer cet argument, l'auteur cite l'exemple d'une magistrate « *habile à mobiliser divers principes politiques de jugement* » – qui est une disposition du « parti politique systématique » – mais « *développe aussi longuement ses opinions sur divers sujets relatifs à son activité professionnelle en s'appuyant sur son expérience personnelle* » – disposition relevant davantage de l'activation de l'ethos de classe. L'évocation d'une expérience personnelle, si elle peut *a priori* être appréhendée comme une disposition de l'ethos de classe, n'est pas selon nous un mode de production *en soi*. Mobiliser une expérience personnelle en tant qu'argument mis au service d'une réflexion

⁴⁶⁰ Daniel Gaxie, « Retour sur les modes de production des opinions politiques », *op. cit.*, p. 301 se référant à Pierre Bourdieu, *La Distinction*, *op. cit.*, p. 500.

politique d' « intérêt général » n'est pas la preuve d'une disposition éthique – c'est-à-dire non politique, non savante – sur la politique. Il conviendrait en réalité de distinguer les *modes de production* et les *facteurs de production*.

Si l'on conçoit les modes de production comme des « *formules génératrices* » – terme mobilisé par Pierre Bourdieu pour définir l'ethos de classe⁴⁶¹ – ou comme des « algorithmes » – terme également employé pour qualifier le parti politique systématique⁴⁶² – on peut alors établir une *équivalence fonctionnelle* entre les modes de production et les catégories de langage politique. D'un côté le langage politique dominant, le *logos*, la maîtrise verbale, qui repose sur un rapport « neutralisé » et « quasi-théorique » à la politique et qui est la condition « *de la production et de la réception de discours d' « intérêt général » sur les questions d' « intérêt général »*, et qui suppose « *la capacité non seulement de déchiffrer et de manipuler les termes « spéciaux » du langage politique, mais aussi de se situer au niveau de quasi abstraction où se situe communément le discours politique* »⁴⁶³. De l'autre le langage dominé, l'ethos, la maîtrise pratique qui « *peut orienter la pratique quotidienne dans tout ce qu'elle a de politique (objectivement) sans accéder jamais à l'explicitation, moins encore à la conceptualisation systématique, et la maîtrise symbolique de l'expérience* »⁴⁶⁴.

À la lecture de ces définitions, il nous semble que le concept de langage politique développé par Bourdieu ne peut donc être réduit à sa dimension proprement « linguistique ». L'*ethos* et le *logos* se rapprochent en réalité d'avantage du concept de « langage informatique », c'est-à-dire d'un système de (dé)cryptage et (dé)codage des informations (politiques). Dans cette optique, nous proposons de retenir uniquement deux modes de production : le parti politique systématique (le *logos*, le langage dominant) et l'ethos de classe (l'*ethos*, le langage dominé). Le mode de production de choix à deux degrés, selon nous trop ambigu, ne sera pas retenu. Au-delà de sa focalisation sur un type particulier de délégation

⁴⁶¹ Pierre Bourdieu, *La Distinction*, op. cit., p. 590-591.

⁴⁶² *Ibid.*, p. 590-591.

⁴⁶³ *Ibid.*, p. 539

⁴⁶⁴ *Ibid.*, p. 537.

politique⁴⁶⁵, qui est de plus daté⁴⁶⁶, ce mode de production regroupe des dispositions qui appartiennent en réalité à chacun des deux autres modes de production⁴⁶⁷.

Si l'on accepte de concevoir les modes de production comme des algorithmes, il convient alors de s'interroger sur les *données traitées* par les algorithmes, ou, s'il on veut rester dans un vocabulaire marxiste, sur les *facteurs de production* propres à chaque *mode de production*. On peut questionner alors les *modalités de sélection des données* traitées ou des *facteurs de production mobilisés*, et qualifier également leur *nature*, leur *volume* et leur *origine*. Il convient cependant de préciser que le volume global des différents facteurs, leurs poids relatifs, leurs origines et leurs modalités de sélection sont propres à chaque mode de production. Dit autrement, le mode de production « parti politique systématique » ne repose pas sur la même combinaison de facteurs que le mode de production activant l'éthos de classe.

On peut recenser trois principaux facteurs de production : les expériences personnelles et les récits biographiques, les biens informationnels et médiatiques, et les stocks accumulés de connaissances (capitaux culturels). Le volume – global et relatif – des facteurs de production et leurs modalités de sélection sont propres à chaque mode de production. Un agent très politisé ne mobilise pas des biens informationnels de même nature – et de même valeur symbolique – que des agents peu ou pas politisés. Et si le facteur « expérience personnelle » peut être employé dans chacun des modes de production, le produit transformé – l'opinion produite – ne sera pas de même nature selon que ce facteur est mobilisé par un mode de production activant l'éthos de classe ou un parti politique systématique.

Une attention particulière doit être également portée à l'*origine* des facteurs de production. Le facteur de production a-t-il été produit « en interne » par le producteur de l'opinion lui-même ou a-t-il été *importé*, « acheté » à d'autres producteurs ? On retrouve derrière cette question des *origines* des facteurs de production la distinction proposée par

⁴⁶⁵ Comme l'a très justement observé Daniel Gaxie : « *En caractérisant le « choix à deux degrés », Bourdieu s'est concentré sur un mode de délégation en particulier : la délégation à un parti politique, et en particulier au Parti communiste. Il s'agit là non pas du « choix à deux degrés » en général, mais de l'une de ses variantes. Bourdieu note d'ailleurs qu'il existe aussi des processus analogues au sein des groupes primaires* ». Daniel Gaxie, « Retour sur les modes de production des opinions politiques », *op. cit.*

⁴⁶⁶ Il est sans doute inutile de démontrer ici la perte d'influence du PCF.

⁴⁶⁷ Bourdieu note ainsi que « *le principe selon lequel les électeurs du Parti communiste produisent leurs opinions varie [...] selon qu'ils savent pratiquement ou théoriquement [...] ce qu'il faut penser ou que, au contraire, ils sont livrés aux dispositions de leur ethos* ». Ce mode de production regroupe donc des agents qui produisent des opinions selon leur ethos de classe, et des agents qui « tentent » de produire leurs opinions dans un langage proprement politique.

Daniel Gaxie entre les producteurs d'opinion en première personne et les reproducteurs d'opinions. Mais par analogie, cette notion d'origine des facteurs peut également recouvrir la question de la délégation et du choix à deux degrés. Les opinions mobilisées par les délégataires pour produire – ou reproduire – leur (propre) opinion peuvent être aussi considérées comme des facteurs de production importés.

Après avoir dissocié conceptuellement les modes de production et les facteurs de production, reste posée la question des conditions qui président à l'activation de la production d'une opinion.

2.4.1.2 Les facteurs contextuels d'activation de la production des opinions politiques

En s'interrogeant sur les conditions qui président à l'activation des modes de production des opinions politiques, Daniel Gaxie a ainsi introduit l'idée selon laquelle des éléments *contextuels* pouvaient être intégrés à l'analyse *dispositionnelle* des modes de production⁴⁶⁸. Partant du principe que les agents sociaux peuvent activer plusieurs modes de production simultanément ou de manière asynchrone, l'auteur postule alors que ces éléments contextuels permettent de comprendre comment les agents sociaux activent un mode de production donné.

En suivant notre propre typologie, un agent social ne peut potentiellement activer que deux modes de production : l'ethos de classe ou le parti politique systématique. Or, comme l'a souligné Pierre Bourdieu, entre l'ethos et le logos, la discontinuité est radicale⁴⁶⁹. Un agent social peu ou pas politisé ne peut pas, ne sait pas, produire une opinion relevant d'un parti politique systématique. Il n'est en revanche pas unimaginable qu'un agent social très politisé, capable de produire un discours théorique et abstrait d'« intérêt général », puisse, dans une situation très particulière, adopter le langage politique des dominés. Cette situation paraît tout de même assez peu probable.

Étant donné le caractère exclusif des deux modes de production retenus, les éléments contextuels, dans notre analyse, servent alors plutôt à éclairer les logiques de sélection – conscientes ou inconscientes – des facteurs de production. Plus simplement encore, les

⁴⁶⁸ Daniel Gaxie, « Retour sur les modes de production des opinions politiques », *op. cit.*

⁴⁶⁹ Pierre Bourdieu, *La Distinction*, *op. cit.*

éléments contextuels permettent d'éclairer les conditions qui président à la production d'une opinion politique.

Nous proposons de distinguer les éléments qui relèvent d'un contexte général ou conjoncturel et les éléments qui relèvent d'un contexte situationnel. Daniel Gaxie évoque comme exemple d'élément conjoncturel le degré de politisation des débats publics. Selon l'auteur, « *plus un problème est débattu au sein des espaces de débat politique, plus ces débats sont repercutés au-delà de ces espaces, plus s'accroît le nombre des agents qui se prononcent « politiquement » que ce soit par répétition des opinions de leur délégataire, par reproduction d'une position exprimée dans les espaces de discussion, ou par reproduction d'une opinion propre* »⁴⁷⁰. Cet effet de politisation de la conjoncture sur les discussions ordinaires a été également souligné dans de nombreuses recherches ethnographiques⁴⁷¹. Il constitue un bon exemple de facteur conjoncturel participant à l'activation de la production des opinions politiques.

Au-delà de ces éléments *conjoncturels*, on peut également postuler que des éléments contextuels *situationnels* participent à l'activation de la production des opinions politiques. Le contexte situationnel concerne le cadre immédiat dans lequel interagissent les agents sociaux. Où, quand, comment, et avec qui les agents produisent-ils des opinions politiques ? Si cette question relativement simple a été posée dans de nombreux travaux ethnographiques⁴⁷², elle demeure en revanche un point aveugle de la sociologie bourdieusienne. Le seul contexte étudié dans les travaux de Bourdieu est celui de la situation d'enquête : c'est face à des sondeurs, en répondant à des questions qu'ils ne se sont dans leur majorité jamais posées, que les agents sociaux produisent des « opinions ». La prise en compte d'éléments contextuels par les analyses dispositionnelles apporterait pourtant des éléments de réponse à ces questions et permettrait de lever une certaine ambiguïté relative à la production des discours éthiques, c'est-à-dire non savants, sur la politique. Ces discours éthiques sont, nous l'avons vu, considérés comme des artefacts générés artificiellement par la situation d'enquête. Pourtant, on peut se demander si, en dehors des situations d'enquête, il n'existerait pas des situations où des agents peu compétents politiquement se verraient incités à produire une opinion politique, lorsque par exemple, ils se trouvent engagés dans une discussion qui « dévie » sur des sujets

⁴⁷⁰ Daniel Gaxie, « Retour sur les modes de production des opinions politiques », *op.cit.*

⁴⁷¹ Vincent Goulet, *Médias et classes populaires. Les usages ordinaires des informations*, *op. cit.* ; Céline Braconnier et Jean-Yves Dormagen, *La démocratie de l'abstention: Aux origines de la démobilisation électorale en milieux populaires*, Paris, Gallimard, 2007.

⁴⁷² Comme par exemple l'ouvrage de Nina Eliasoph, *L'évitement du politique*, *op. cit.*

politiques. Ou si, au contraire, dans une situation d'intimité qui échappe à tout contrôle social, les agents peu politisés ne produiraient pas des discours éthiques spontanés sur la politique, en regardant le journal télévisé par exemple. Et le cas échéant, si cette situation n'interférerait pas sur les modalités ordinaires de production des opinions politiques.

Nous postulons donc que des éléments d'ordre contextuel permettent d'enrichir les analyses dispositionnelles des modes de production des opinions politiques. De plus, sur internet, ces données contextuelles peuvent être objectivées et mesurées quantitativement.

2.4.2 Objectiver les éléments de contexte sur internet

Une posture passive d'enregistrement des opinions numériques permet en effet d'objectiver et de mesurer quantitativement le contexte dans lequel sont produites et énoncées les opinions numériques. L'objectivation de ces données contextuelles permet de vérifier empiriquement les hypothèses évoquées plus haut.

Une objectivation des données contextuelles *conjoncturelles* sur internet permet ainsi de vérifier empiriquement l'hypothèse selon laquelle le degré de politisation des débats publics détermine le volume des opinions politiques produites par les agents sociaux. Ainsi, comme nous l'avons vu, pour Daniel Gaxie, « *plus un problème est débattu au sein des espaces de débat politique, plus ces débats sont répercutés au-delà de ces espaces, plus s'accroît le nombre des agents qui se prononcent « politiquement* »⁴⁷³. Cette hypothèse, peut être mesurée *quantitativement* sur internet.

Le chercheur peut d'abord mesurer la place qu'occupe un problème dans les *espaces de débat politique*. L'ensemble de la production du champ médiatique est numérisée sur internet. On peut donc facilement mesurer le volume des articles, des productions et des débats sur un sujet donné. S'il souhaite ainsi mesurer l'intensité d'une campagne électorale, le chercheur peut commencer par comptabiliser, à partir d'une recherche par mots-clés⁴⁷⁴, le nombre d'articles parus sur les sites de journaux d'information. Le chercheur détermine lui-même le périmètre des espaces de débat politique qu'il souhaite étudier. Ces espaces peuvent être de plus subdivisés : on peut alors comparer le nombre d'articles parus dans des journaux

⁴⁷³ Daniel Gaxie, « Retour sur les modes de production des opinions politiques », *op. cit.*

⁴⁷⁴ On peut choisir comme mots-clés les noms des candidats à l'élection, des partis politiques ainsi que d'autres termes systématiquement associés à la campagne.

d'opinion et dans des journaux omnibus, dans des sites de médias audiovisuels et dans des sites de journaux « papiers », dans la presse d'opinion de gauche et dans celle de droite, etc. Le chercheur peut ensuite calculer le ratio du nombre d'articles consacrés à la campagne sur le nombre d'articles total parus dans ces journaux. L'évolution quotidienne de ce ratio peut alors constituer un baromètre d'intensité de la campagne électorale dans les espaces de production idéologique.

Le chercheur peut ensuite comparer cet indicateur avec un indicateur volumétrique de production des opinions politiques publiées par les agents sociaux eux-mêmes. À partir d'un corpus exhaustif d'autopublications numériques publiées sur les espaces d'autopublication, il est possible de mesurer, *via* une recherche par mots-clés, le nombre de messages consacrés à la campagne sur le nombre total de messages publiés. Le chercheur peut alors au final comparer le degré de politisation des espaces de débat public avec le degré de politisation des messages publiés par les agents sociaux sur les espaces d'autopublication. L'hypothèse d'un effet de politisation de la conjoncture sur la production des opinions politiques peut être ainsi vérifiée empiriquement sur internet.

Une posture passive d'enregistrement des opinions numériques permet également d'objectiver des données contextuelles *situationnelles*. Un nombre important de messages publiés en ligne s'insèrent en réalité dans des échanges collectifs, dans des discussions en ligne. Le chercheur peut ainsi dissocier les messages qui s'insèrent dans des échanges en ligne et les productions isolées, qui ne répondent à aucun message ni ne suscitent aucune réaction. Le cas échéant, le chercheur peut alors objectiver la discussion à laquelle prend part l'enquêté en analysant le nombre de participants, les sujets débattus, le niveau de compétence politique des participants, etc. Ces données contextuelles permettent de vérifier de nombreuses hypothèses formulées par les approches cognitives et délibératives. En trouvant une matérialité empirique sur internet, ces hypothèses permettent alors de mieux saisir les conditions qui président à l'activation des modes de production des opinions politiques.

Les nouvelles technologies d'étude des opinions numériques permettent donc de ne pas reproduire les biais des sondages et de l'enquête par questionnaire. Cependant, un

dispositif de recherche qui serait basé uniquement sur une logique de collecte et d'enregistrement des opinions numériques serait confronté à plusieurs difficultés d'envergure.

3. Les limites de l'usage exclusif d'une posture passive d'enregistrement des opinions numériques

La procédure de collecte et d'enregistrement des autopublications numériques offre la garantie au chercheur de mesurer des opinions qui ne sont pas des artefacts d'enquête. Leur codage *a posteriori* permet d'en objectiver le mode de production, la valeur et le contexte d'énonciation.

Cependant, un dispositif de recherche qui reposerait uniquement sur ces opérations d'enregistrement et de codage se heurterait à plusieurs difficultés d'envergure. Comprendre ces difficultés – qui ont été les nôtres au début de cette recherche – nécessite d'explicitier les potentialités – et surtout les limites – des logiciels de veille et de *tracking*⁴⁷⁵.

Une première limite de ces logiciels est de prendre comme unité de mesure l'autopublication et non le producteur de l'autopublication. À partir d'une recherche par mot-clé, les logiciels de *tracking* indiquent *le nombre de messages* correspondant à la recherche, mais ne mentionnent pas *le nombre d'individus* à l'origine de ces messages. Le chercheur ne peut pas alors mesurer l'intensité productive des auteurs des autopublications, en calculant par exemple un ratio de nombre de messages publiés par individu. Un tel indicateur serait pourtant indispensable pour objectiver l'investissement des espaces d'autopublication par des réseaux militants⁴⁷⁶ (3.1).

Une deuxième difficulté est de ne pas pouvoir caractériser socio-démographiquement et politiquement les auteurs des autopublications numériques. Si les opinions collectées délivrent certains indices sociologiques sur leurs producteurs (*via* le niveau de langage, la technicité des termes mobilisés, le niveau d'orthographe ou le contenu même du message⁴⁷⁷) et si certains producteurs dévoilent parfois d'eux-mêmes des informations

⁴⁷⁵ En nous concentrant particulièrement sur le logiciel *Ami Opinion Tracker* mobilisé pour cette thèse.

⁴⁷⁶ Anaïs Théviot, « Twitter en regardant la télévision : une campagne transmédiatique interactive ? Analyse comparée des stratégies numériques au Parti Socialiste et à l'Union pour un Mouvement Populaire lors des ripostes-party », *Télévision*, 5 (2014), p. 95-112.

⁴⁷⁷ Un message soutenant explicitement le Parti socialiste peut raisonnablement être appréhendé comme l'opinion d'un électeur socialiste.

sociodémographiques (leur sexe, leur âge, leur profession, etc.)⁴⁷⁸, ces informations ne sont pas systématiquement disponibles. Ces données sont pourtant indispensables pour situer les internautes dans l'espace social et analyser sociologiquement leurs opinions. (3.2).

Après avoir pris connaissance de ces limites, nous nous sommes donc trouvé face au dilemme suivant : accepter les contraintes imposées par les logiciels de veille et de *tracking* afin de privilégier une étude quantitative – et possiblement « exhaustive » – des opinions numériques ou bien refuser de recourir à ces logiciels et alors s'orienter *de facto* vers une recherche exclusivement qualitative et localisée (3.3).

3.1 Une unité de mesure problématique

Il convient d'emblée de préciser que la limite technique que nous allons évoquer ici a été partiellement « résolue » lors des dernières années de notre travail doctoral. En clair, la limite évoquée ici n'en est aujourd'hui plus une. Pourquoi alors lui consacrer un paragraphe ? D'une part car c'est la découverte de cette limite qui nous a conduit à développer notre propre dispositif d'enquête. D'autre part car cette limite illustre à quel point les logiciels de veille et de *tracking* n'étaient initialement pas conçus pour répondre à des problématiques de sciences sociales.

Cette limite concerne l'unité de mesure des indicateurs volumétriques fournis par les logiciels de veille et de *tracking*. Cette unité de mesure est l'autopublication et non l'auteur de l'autopublication, l'opinion et non le producteur de l'opinion, le message et non l'individu. Si le chercheur peut par exemple facilement et immédiatement connaître le nombre de tweets mentionnant un homme politique sur une plage de temps donnée, il ne peut en revanche connaître le nombre d'individus à l'origine de ces messages.

Afin de mieux comprendre cette limite technique et ses enjeux méthodologiques, nous allons rapidement évoquer les différentes étapes du processus de collecte et d'enregistrement des autopublications numériques. Il existe deux façons différentes de collecter les messages publiés sur les réseaux sociaux. Pour simplifier notre démonstration nous prendrons uniquement l'exemple du réseau social Twitter. Les logiciels de veille et de *tracking* ne permettent pas de collecter et d'archiver l'exhaustivité des tweets publiés sur le réseau

⁴⁷⁸ Cf. chapitre 1.

social⁴⁷⁹. L'utilisateur du logiciel – ici le chercheur – a alors deux possibilités : créer des « sources de type *suivi de compte* » ou créer des « sources de type *recherche par mot-clé* ». Créer une « source de type *suivi de compte* » consiste à collecter et archiver l'intégralité des tweets publiés par un utilisateur donné. Le chercheur entre dans le logiciel l'adresse URL⁴⁸⁰ du compte Twitter dont il souhaite enregistrer les tweets. Le logiciel scanne ensuite de manière continue l'intégralité des tweets publiés par le compte. Idéalement, le chercheur pourrait alors créer autant de sources qu'il existe de comptes Twitter. Cela est dans les faits impossible. Non seulement le chercheur ne dispose pas d'une liste exhaustive de tous les utilisateurs de Twitter, d'autre part entrer « manuellement » une à une ces sources dans le logiciel demanderait sans doute plusieurs années de travail à temps plein. Les « sources de type *suivi de compte* » constituent pourtant le seul moyen de disposer d'une unité de mesure qui soit l'individu et non uniquement le tweet.

Une solution par défaut consiste alors à créer des « sources de type *recherche par mot-clé* ». Ces sources ont pour fonction de collecter sur l'ensemble du réseau social la totalité des tweets contenant un mot-clé donné. Ainsi, en créant « une source de type *recherche* : Hollande », le logiciel collecte et archive l'intégralité des tweets mentionnant l'occurrence « Hollande » sur le réseau social⁴⁸¹. Les « sources de type *recherche par mot-clé* » collectent et archivent de manière exhaustive les tweets mentionnant le mot-clé associé à la source créée. Tous les tweets collectés par ces sources viennent chaque minute enrichir une banque de données stockée dans le logiciel. Plusieurs milliers ou millions de tweets peuvent être ainsi archivés. Dans le cadre d'un projet de recherche mené parallèlement à ce travail doctoral⁴⁸², nous avons créé plus de 140 « sources de type *recherche* » avec comme mots-clés des noms d'hommes politiques, de partis politiques, d'institutions politiques et des termes relatifs à la *politics* ou aux *policies*, etc. Après trois années de collecte, cette base de données s'élève aujourd'hui à une centaine de millions de tweets politiques collectés.

Les logiciels de veille et de *tracking* permettent de visualiser graphiquement et numériquement la volumétrie des tweets enregistrés dans la base de données. Très concrètement, le chercheur peut ainsi générer facilement une courbe retraçant *le nombre de*

⁴⁷⁹ Twitter monnaie en effet l'accès à ses données. Les tiers (entreprises ou développeurs informatiques) souhaitant accéder à la totalité des données doivent souscrire à des API payantes (cf. glossaire, p. 464.).

⁴⁸⁰ Cf. glossaire p. 464.

⁴⁸¹ Là encore, pour atteindre un archivage exhaustif, on peut arguer qu'il suffirait de créer autant de « sources de type recherche par mot-clé » que la langue française ne compte de mots. Cette opération demanderait elle aussi des années de travail.

⁴⁸² L'ANR Imagiweb.

tweets mentionnant un homme politique donné avec comme unité temporelle la minute, le jour, la semaine, le mois ou l'année. Cependant, il est impossible de générer une courbe qui retracerait l'évolution du *nombre d'individus* qui ont publié des tweets mentionnant l'homme politique donné.

Cette impossibilité *technique* entraîne de nombreuses conséquences *méthodologiques*. Une des plus-values heuristiques des technologies d'étude des autopublications numériques est, on l'a vu, de pouvoir mesurer le nombre d'agents sociaux qui publient spontanément une opinion – ou du moins un message – à propos d'un sujet politique. Contrairement aux panels par sondage qui postulent qu'un *nombre* identique de Français forment une opinion sur un sujet donné d'une vague de sondage à l'autre – les variations enregistrées concernent alors uniquement *la nature* des opinions sondées – un enregistrement passif des autopublications numériques objective le fait que le *nombre* d'opinions exprimées sur internet n'est pas le même d'une vague à l'autre. En d'autres termes, le chercheur n'enregistre pas sur Twitter un volume identique d'opinions politiques publiées sur un sujet d'un jour à l'autre ou d'un mois à l'autre.

Un même individu peut publier un nombre presque illimité de tweets par jour. Or tous les individus ne publient pas le même nombre de tweets par jour. L'équation « un individu égale une opinion » n'est donc pas respectée sur Twitter. On pourrait arguer que ce constat n'est pas problématique si l'on considère que *premièrement* toutes les opinions n'ont pas la même valeur, que *deuxièmement* un nombre important d'opinions publiées par un groupe restreint d'individus est un indicateur de l'intensité de la mobilisation de ce groupe, et que *troisièmement*, c'est donc moins le nombre d'individus que l'intensité productive des individus ou des groupes mobilisés sur Twitter qui est révélateur de l'état de l'opinion à un moment donné. Le seul problème de ce raisonnement est qu'il suppose de pouvoir calculer l'intensité productive d'un individu ou d'un groupe. Ce calcul nécessite de connaître le *nombre d'individus* qui se sont exprimés sur un sujet donné et non uniquement le *nombre de tweets* publiés. Or cette information n'est précisément pas délivrée par les logiciels de veille et de *tracking*.

Ne pouvant calculer un ratio de tweets par individu, le chercheur ne peut pas alors objectiver la mobilisation de certains individus ou groupes, et notamment la mobilisation politique de groupes militants. Publier massivement des tweets pour soutenir un candidat lors d'un meeting ou d'un débat télévisé est devenu pourtant une activité militante parmi

d'autres⁴⁸³. S'il souhaite connaître le nombre d'individus qui ont exprimé une opinion sur Twitter durant un débat politique télévisé, le chercheur ne saura alors pas dissocier les *tweets d'opinion* et les *tweets militants* publiés de façon presque « industrielle » par des groupes restreints de militants politiques⁴⁸⁴.

L'unité de mesure retenue par ces logiciels – le tweet, et non l'individu – est donc un véritable handicap méthodologique pour le chercheur. Cette contrainte technique a été néanmoins récemment résolue par les informaticiens collaborant à ce projet de recherche. Le chercheur peut désormais choisir comme unité de mesure le message ou l'individu lorsqu'il étudie une production de tweets publiés sur un temps donné. Si ces logiciels offrent désormais la possibilité de calculer le volume de production de tweets politiques par individu, ils ne permettent pas néanmoins de caractériser socio-démographiquement et politiquement ces individus.

3.2 L'impossible caractérisation sociologique des producteurs d'opinion numérique

Nous avons déjà évoqué dans le précédent chapitre la difficulté méthodologique à caractériser socio-démographiquement et politiquement les auteurs des autopublications numériques. Pourtant, situer les producteurs des opinions numériques dans l'espace social constitue la condition *sine qua non* de la mise en application des potentialités épistémologiques des nouvelles technologies d'étude des opinions numériques décrites plus haut.

Nous avons en effet démontré qu'une posture passive d'enregistrement des opinions numériques et une codification *a posteriori* des *verbatim* permettaient de dépasser potentiellement les quatre limites des sondages d'opinion et des enquêtes par questionnaire fermé.

Dépasser ces limites nécessite de pouvoir situer socialement les producteurs des opinions au moins aussi finement que le font les sondages d'opinion. Tout sondage comporte en effet des questions « signalétiques », socio-démographiques (sexe, âge, profession, niveau

⁴⁸³ Anaïs Théviot, « Twitter en regardant la télévision : une campagne transmédiatique interactive ? Analyse comparée des stratégies numériques au Parti Socialiste et à l'Union pour un Mouvement Populaire lors des ripostes-party », art. cit.

⁴⁸⁴ Nous ne sous-entendons pas que les tweets militants ne soient pas des tweets d'opinion. Au contraire ils expriment une opinion mobilisée.

de diplôme, etc.) et de préférence politique qui permettent de caractériser sociologiquement les répondants et donc de vérifier que les quotas sont respectés. Les sondeurs et les chercheurs peuvent ensuite croiser ces variables signalétiques avec les réponses des sondés aux questions d' « opinion »⁴⁸⁵ et ventiler ainsi les résultats obtenus en fonction de ces variables sociologiques. Ces informations sociodémographiques sont indispensables à l'analyse sociologique des opinions numériques, et ce à chaque étape du programme de recherche esquissé plus haut.

Ces informations socio-démographiques permettent d'abord de déterminer précisément quelles catégories de la population publient des opinions numériques, dans quelles proportions et à quelle fréquence. La sociologie politique a de longue date démontré⁴⁸⁶ que l'intérêt déclaré pour la politique, la propension à répondre aux enquêtes d'opinion ou la propension à participer aux élections étaient fonction du sexe, de l'âge, de la profession, du niveau de diplôme, etc. Définir précisément dans quelle mesure ces inégalités socioculturelles déterminent de manière analogue la propension à produire et publier des opinions politiques sur internet nécessite donc de pouvoir situer socialement les producteurs de ces opinions.

Mais ces informations socio-démographiques permettent également de mesurer à quel point l'activation d'un mode de production donné est elle aussi déterminée socialement⁴⁸⁷. Nous l'avons vu, en codifiant les opinions politiques *a posteriori*, le chercheur peut, à partir du *verbatim* étudié, déterminer le niveau de langage politique, ou, selon notre définition, son mode de production. S'en tenir à cette codification, à cette information *induite*, sans connaître précisément l'identité sociale de son producteur, contraint le chercheur à définir le producteur de l'opinion à partir d'une typologie binaire dominant/dominé, ou logos/ethos. Obtenir des informations sociodémographiques sur les producteurs des messages permet d'une part d'affiner cette typologie binaire et d'autre part de déterminer dans quelle mesure les différentes catégories sociales s'orientent vers un mode de production plutôt qu'un autre.

Les données quantitatives générées de manière automatique et systématique par les espaces d'autopublication, telle que le nombre de reprises d'un message (nombre de retweets) ou le nombre d'abonnés à un compte producteur d'opinion constituent, nous l'avons vu, des indices intéressants si l'on souhaite objectiver l'inégale valeur des opinions. Ces données

⁴⁸⁵ Patrick Lehingue dissocie quatre catégories de questions : les questions signalétiques, les questions se référant aux habitudes, les questions cognitives et enfin les questions d'opinion. Cf. Patrick Lehingue, *Subunda*, op. cit. p. 150.

⁴⁸⁶ Daniel Gaxie, *Le cens caché*, op. cit.

⁴⁸⁷ Daniel Gaxie, « Retour sur les modes de production des opinions politiques », op. cit.

chiffrées ne suffisent pas cependant à déterminer le poids ou la force sociale d'une opinion. Deux autres variables doivent selon nous être prises en compte pour déterminer la force sociale d'une opinion : le degré de proximité au champ de production idéologique (déterminer la profession du producteur de l'opinion et sa position sociale dans le champ) et son activité politique ou son degré d'engagement politique (identifier les militants politiques, syndicaux, associatifs, et leur fonction). Ces données sont effet indispensables à la détermination du potentiel politique mobilisateur de l'idée-force étudiée.

Enfin, si le chercheur souhaite étudier le contexte situationnel d'énonciation des opinions numériques, lorsque les opinions étudiées s'insèrent dans un échange collectif en ligne par exemple, il est alors nécessaire de pouvoir qualifier socio-démographiquement et politiquement les individus qui prennent part à cet échange. Là encore, ces données ne sont pas détectables et qualifiables automatiquement par les logiciels de veille et de *tracking*.

La prise en compte de cette limite – qui, contrairement à la précédente, n'a pas été résolue informatiquement et n'est pas prête de l'être – pose donc au sociologue un dilemme de taille.

3.3 « Accepter d'abandonner la référence sociodémographique »⁴⁸⁸ ou abandonner l'analyse statistique ?

Nous nous sommes donc trouvé confronté au choix suivant :

- Privilégier un dispositif d'étude rationalisé et une analyse quantitative de grande ampleur voire une analyse exhaustive des opinions numériques, mais, pour ce faire, recourir aux logiciels de veille et de *tracking* et accepter de perdre la référence socio-démographique, c'est-à-dire accepter de ne pas pouvoir situer socialement les producteurs des opinions numériques ;
- Refuser d'abandonner la référence socio-démographique, ne pas recourir aux logiciels de veille et de *tracking* et privilégier une analyse « manuelle » qualitative sur un terrain très localisé.

⁴⁸⁸ Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis, op. cit.*, p. 189.

La première option est défendue par différents auteurs, dont notamment Dominique Boullier et Audrey Lohard⁴⁸⁹. Pour ces deux sociologues, les limites des logiciels de veille et de *tracking*, et notamment l'impossibilité de caractériser socio-démographiquement les producteurs des opinions numériques, devraient inciter les chercheurs en sciences humaines et sociales à changer de paradigme :

« Ce qui est alors étudié [sur internet] ne fait plus référence à des émetteurs bien identifiés dont les chercheurs posséderaient les caractéristiques sociodémographiques classiques et considérées comme définitives de la personne (en tant que sujet statistique tout au moins). Ici les énoncés sont nivelés et c'est leur agrégation qui fait sens. Ce qui ne devrait pas effrayer toute une partie des sciences sociales qui n'a pas de culte particulier des individus et qui est prête à considérer que ces flux d'opinions sont des objets sociologiques tout à fait légitimes. Mais il faut alors changer de paradigme et ne plus chercher des explications par des causes qui seraient logées dans des propriétés socio-démographiques considérées comme des attributs d'un autre statut que toutes celles auxquelles on peut accéder sur le web. [...] Il faut donc faire son deuil et pour longtemps des repères socio-démographiques sur le web actuel et pour cela changer de paradigme d'explication » dans les sciences humaines et sociales, tellement fondées sur une théorie trop implicite de l'individu comme individu et non comme sociation »⁴⁹⁰.

Nous refusons très clairement de nous résoudre à abandonner la référence socio-démographique, à faire « notre deuil de l'individu » et à changer de paradigme. Tout au contraire, nous soutenons que l'appareillage conceptuel et théorique de la sociologie dispositionnelle est particulièrement opérant pour étudier les opinions numériques. Tout comme Pierre Bourdieu et Jean-Claude Passeron, qui, à propos des recherches des années 1960 sur la télévision, déploraient que « tous les gens qui ont la télévision cessent d'être des hommes réels, donc divers, pour devenir des « télévionnaires » sans visage »⁴⁹¹, nous déplorons que les recherches académiques actuelles considèrent les « internautes » comme des êtres irréels et sans ancrage social. Nous défendons au contraire l'idée selon laquelle les « internautes » sont des agents sociaux bien réels, inégalement pourvus en capitaux économiques et culturels, inégalement compétents politiquement. Les internautes n'évoluent

⁴⁸⁹ *Ibid.*, p. 190.

⁴⁹⁰ *Ibid.*, p.190.

⁴⁹¹ Pierre Bourdieu et Jean-Claude Passeron, « Sociologues des mythologies et mythologies des sociologues », *Les temps modernes*, n°211, 1963, p. 998-1021.

pas dans un monde parallèle qui aurait ses logiques propres, un monde « virtuel » libéré de toute pesanteur sociale.

Cependant, refuser d'abandonner la référence socio-démographique implique de refuser dans le même temps de recourir à un traitement statistique et quantitatif de grande ampleur des opinions numériques. Sans l'aide d'un logiciel de veille et de *tracking*, le chercheur ne peut pas en effet collecter et archiver seul, de manière exhaustive, les autopublications numériques publiées en permanence sur les réseaux sociaux. Une seconde option consisterait donc à privilégier une recherche qualitative, « manuelle », sans doute un peu « artisanale », et travailler à partir d'un échantillon raisonné de données. Une telle approche n'est pas moins légitime scientifiquement qu'une recherche quantitative. Seulement, une approche qualitative ne permettrait pas de vérifier l'ensemble des hypothèses de recherche que nous avons développées dans ce chapitre.

En définitive, nous avons refusé de choisir entre ces deux options et avons construit une méthodologie d'enquête alternative, relativement inédite, qui entend concilier approche quantitative et qualification sociologique des producteurs d'opinions numériques.

4. Panéliser Twitter : concilier une posture passive d'observation et une logique active de qualification sociologique des producteurs de tweets politiques

Le dispositif méthodologique imaginé dans le cadre de cette recherche doctorale repose sur trois exigences : obtenir un nombre important d'informations comparables sur une population suffisamment nombreuse pour autoriser le traitement statistique⁴⁹² ; qualifier sociologiquement très finement cette population ; enregistrer et analyser la production de cette population sans interagir avec elle.

Échantillonner aléatoirement une population d'individus inscrits sur un réseau social – ici, Twitter – et lui administrer un court questionnaire comprenant une vingtaine de questions socio-démographiques permet de concilier ces trois exigences. Une fois le questionnaire

⁴⁹² Pierre Bourdieu, *La Distinction, op. cit.*, p. 591.

complété par les individus tirés au sort, plus aucune autre interaction⁴⁹³ n'a lieu entre le chercheur et les enquêtés. S'ils l'acceptent, les répondants deviennent membres d'un panel. Le chercheur peut alors archiver et collecter de façon exhaustive les tweets qui seront publiés par les enquêtés. Ces messages pourront être systématiquement rattachés aux attributs sociologiques de leurs producteurs. De plus, les individus qui n'ont pas répondu au questionnaire peuvent être partiellement qualifiés, *via* les informations parcellaires qu'ils indiquent d'eux-mêmes sur le profil⁴⁹⁴ de leur compte.

Avant de détailler plus précisément le principe et les avantages de cette méthode (4.2) et les phases de constitution du panel (4.3), il convient d'expliciter les raisons qui nous ont conduit à étudier un réseau social particulier (4.1).

4.1 Panéliser un réseau social particulier : le choix de Twitter

Jusqu'alors, au cours de ces deux premiers chapitres, nous avons parlé d'« autopublications » et d'« opinions numériques » sans spécifier ni distinguer les différents espaces d'autopublication sur lesquels elles étaient publiées. Notre terrain d'enquête ne s'est lui porté que sur un seul espace : le réseau social Twitter. Avant d'évoquer les raisons qui nous ont conduit à choisir ce réseau social en particulier et d'expliciter les avantages et les inconvénients de ce choix, il convient de préciser pourquoi nous avons centré notre terrain sur un seul espace en particulier.

Nous l'avons entrevu dans le précédent chapitre, il existe une très grande variété d'espaces d'autopublication numérique : les blogs, les réseaux sociaux, les *forums* électroniques, les sites de presse, les sites d'achat en ligne, etc. Or, comme l'ont souligné Dominique Boullier et Audrey Lohard, chacun de ces espaces peut « *potentiellement accueillir des opinions, mais ces dernières trouveront souvent une expression différente en fonction des sites sur lesquels elles seront énoncées. D'une certaine manière, elles seront formatées différemment. Les manières de traiter les opinions seront très différentes en fonction de l'environnement technique dans lequel on les trouvera. Car l'environnement technique ne formate pas seulement les manières d'exprimer son opinion (ne pas dépasser*

⁴⁹³ Nous avons néanmoins réalisé un vingtaine d'entretiens semi-directifs quelques mois après le « lancement » du panel. Nous y reviendrons dans le chapitre suivant.

⁴⁹⁴ Cf. glossaire p. 464.

140 caractères, mettre une note entre 1 et 5, etc.), mais induit également un contexte social et cognitif qui façonne les manières de donner son opinion »⁴⁹⁵. Si nous ne partageons pas entièrement l'arrière-plan théorique qui sous-tend cette citation, il est certain que chaque espace d'autopublication constitue en réalité un terrain d'enquête en soi, qui obéit à ses logiques propres. Si, dans une perspective comparatiste, il peut être intéressant d'étudier plusieurs de ces espaces, cette comparaison nécessiterait néanmoins selon nous un investissement (en temps et en moyens techniques) trop important pour une recherche doctorale.

Nous avons donc choisi de ne retenir qu'un seul espace d'autopublication comme terrain d'enquête, le réseau social – ou site de *micro-blogging*⁴⁹⁶ – Twitter. Précisons d'emblée, en toute franchise, que ce fut un choix par défaut. Notre choix initial s'était porté sur le réseau social Facebook, et ce pour une raison simple : en France, la population internautes inscrite sur ce réseau social est numériquement la plus importante (54 % des internautes déclaraient avoir un compte Facebook en 2012⁴⁹⁷) et sociologiquement la plus diversifiée. 63 % des cadres et 53 % des ouvriers disposant d'un accès à internet déclarent être inscrits sur Facebook alors que seuls 17 % des cadres et 10 % des ouvriers déclarent avoir un compte Twitter. C'est en raison de nombreuses difficultés et contraintes techniques, dont certaines imposées par Facebook (notamment en raison d'une politique de confidentialité et d'accès aux données très ambiguë et instable), que nous avons été contraint d'abandonner ce terrain d'étude.

Notre second choix s'est alors porté sur Twitter. Le choix de ce terrain d'étude, bien qu'intéressant – nous allons tenter de le démontrer –, a suscité deux motifs d'insatisfaction. Le premier concerne la population de ce réseau social qui est, nous venons de le voir, insuffisamment diversifiée socialement. Le second a trait à une contrainte technique qui est l'une des spécificités de ce réseau social : la limite des 140 caractères. Tout message publié sur Twitter ne doit pas comporter plus de 140 caractères. Les tweets sont donc des messages très brefs, composés d'une trentaine de mots tout au plus. Cette contrainte limite donc considérablement la capacité expressive des individus ; capacité qui constitue pourtant une des plus-values heuristiques de l'autopublication numérique par rapport à d'autres modes et formats d'expression des opinions politiques, comme par exemple le questionnaire fermé ou

⁴⁹⁵ Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis, op. cit.*, p. 54.

⁴⁹⁶ Cf. glossaire p. 464.

⁴⁹⁷ IFOP, « Observatoire des réseaux sociaux », novembre 2012.

le vote⁴⁹⁸. D'un point de vue théorique et méthodologique, il peut être alors plus ardu de mesurer l'intensité d'une opinion ou d'objectiver son mode de production lorsque cette opinion prend la forme d'un message de moins de 140 caractères.

Cette contrainte expressive et le format standardisé des messages publiés sur Twitter facilitent néanmoins leur traitement informatique et leur codage manuel. La comparabilité formelle et la taille restreinte des tweets favorisent en effet la codification manuelle de très importants volumes de messages. De façon très concrète, alors qu'une analyse qualitative d'un billet de blog peut aisément demander plusieurs dizaines de minutes au chercheur, la codification d'un message de 140 caractères peut être réalisée en moins d'une minute.

Au-delà de ces dimensions techniques et méthodologiques, le choix de Twitter comme terrain d'enquête nous inspire finalement deux motifs de satisfaction. En premier lieu, ce réseau social semble être le terrain d'étude privilégié des recherches académiques qui entendent démontrer le caractère prédictif d'internet. Nous l'avons vu dans le précédent chapitre, la majorité de ces travaux prennent pour terrain d'étude Twitter – en partie d'ailleurs pour les mêmes raisons que nous : une plus grande facilité d'accès aux données⁴⁹⁹, des messages plus aisément codifiables, des volumes gigantesques de tweets publiés. Si nous ne partageons pas le même arrière-plan théorique que ces travaux, il peut être tout de même intéressant de comparer certaines de leurs données aux nôtres, et ainsi de contester et/ou vérifier certains de leurs hypothèses et résultats.

Enfin, notre deuxième motif de satisfaction a trait à la place qu'occupe aujourd'hui Twitter dans les champs médiatique, journalistique, sondagier et politique. Malgré la taille relativement réduite de sa population⁵⁰⁰, le réseau social est particulièrement mobilisé par les médias, les journalistes et les hommes politiques qui font référence à cet espace soit comme un micro-trottoir ou un « baromètre d'opinion »⁵⁰¹, soit comme un indicateur d'« audience sociale » des programmes télévisés, ou encore comme un lieu d'altercations politiques, etc. Le fait que les journalistes et plus largement les agents des champs médiatique et politique soient eux-mêmes en très grande majorité inscrits sur le réseau social n'est sans doute pas étranger à

⁴⁹⁸ Patrick Lehingue, *Le vote*, *op. cit.*

⁴⁹⁹ Contrairement à Facebook, il n'est généralement pas nécessaire d'obtenir l'accord de l'utilisateur pour avoir accès à ses données.

⁵⁰⁰ Seuls 11 % des internautes français déclarent être inscrits sur Twitter contre 54 % pour Facebook. Cf. IFOP, « Observatoire des réseaux sociaux », 2012.

⁵⁰¹ On peut citer entre autres le « Politivox » de libération.fr, le « Twittoscope » du journal gratuit Métro, « Le Scan » du figaro.fr, ou encore le « Twittrack2012 » d'atlantico.fr.

l'intérêt – démesuré ? – qu'ils portent à Twitter. Cette recherche tentera d'en apporter une preuve empirique en objectivant cet « entre soi ».

C'est donc à ce terrain d'enquête singulier que nous avons appliqué notre dispositif méthodologique.

4.2 Le principe et les avantages méthodologiques de la panélisation de Twitter

Avant d'évoquer les principaux attraits méthodologiques de notre dispositif (4.2.2), nous allons en expliciter le principe (4.2.1).

4.2.1 Le principe de la panélisation

Le principe de notre dispositif méthodologique est au final assez simple. Il consiste à collecter et archiver durant plusieurs mois l'intégralité des tweets politiques publiés par un échantillon représentatif d'individus inscrits sur le réseau social Twitter ; individus dont on connaît précisément et systématiquement les attributs sociologiques : le sexe, l'âge, la profession, le niveau d'étude, le lieu d'habitation, la profession des parents, les trajectoires de vote, etc.

Afin de connaître de façon systématique ces attributs sociologiques, un court questionnaire composé d'une vingtaine de questions socio-démographiques a été administré à un échantillon aléatoire – et donc représentatif – d'individus inscrits sur Twitter. En répondant à ce questionnaire, les individus tirés au sort ont également accepté de faire partie d'un panel dont la fonction scientifique a été clairement dévoilée.

Après l'étape d'administration du questionnaire, plus aucune autre interaction n'a eu lieu entre le chercheur et les individus panélisés. L'objectif du chercheur était alors de se faire « oublier » des enquêtés afin que le dispositif d'enquête n'interfère pas sur leur production future de tweets. Il s'agissait donc de conserver le principal attrait des opinions numériques – ne pas être une production générée par une interaction de recherche – tout en disposant d'informations sociologiques sur leurs producteurs, *via* l'administration du questionnaire d'entrée.

Nous avons également sélectionné aléatoirement un échantillon d'individus qui n'ont pas répondu au questionnaire, en cherchant à qualifier de façon la plus détaillée possible leurs attributs sociologiques, en collectant « manuellement » les informations que les internautes publient parfois d'eux-mêmes sur les réseaux sociaux. Contrairement aux sondages d'opinion, nous avons donc pu identifier partiellement le profil social des individus qui ont implicitement ou explicitement refusé de faire partie de l'enquête. Nous avons également pu suivre leur production, en parallèle de celle du « panel répondants ».

4.2.2 Les avantages méthodologiques de la panélisation

L'intérêt de notre dispositif est, on l'a compris, de tirer parti des potentialités méthodologiques et épistémologiques d'un enregistrement passif des opinions numériques tout en s'affranchissant des limites imposées par les logiciels de veille et de *tracking*.

Les avantages d'un enregistrement passif des opinions numériques ont été déjà longuement détaillés plus haut. En neutralisant le biais d'imposition de problématique, le chercheur enregistre uniquement des opinions qui préexistent matériellement à l'enquête, des opinions qui offrent la garantie de ne pas être des artefacts. En mettant au jour le travail d'énonciation de l'opinion – énonciation relativement brève sur Twitter rappelons-le – le chercheur peut également objectiver le mode de production de l'opinion étudiée. De plus, le nombre de « retweets » d'un message et le nombre d'« abonnés » aux comptes des individus producteurs d'opinions permettent de mesurer la valeur potentielle des opinions publiées.

Notre dispositif permet donc de tirer pleinement parti de ces potentialités tout en s'affranchissant des limites imposées par les logiciels de veille et de *tracking*. Une première limite de ces logiciels est de proposer comme unité de mesure volumétrique uniquement le tweet, et non l'individu. Une analyse panélisée de Twitter permet de résoudre ce problème. Si nous avons vu qu'il était humainement impossible de créer autant de « sources de type *suivi de compte* » qu'il existe de comptes Twitter, il est en revanche raisonnablement possible d'en créer 1228 (soit la taille de notre panel). Dès lors, l'unité de mesure disponible n'est plus uniquement le tweet mais le compte producteur de tweet, l'individu. On peut dès lors calculer le nombre de tweets publiés par un individu sur un temps donné et ainsi mesurer son volume et sa fréquence de publication de tweets et de tweets politiques.

De plus, la logique de panel permet de suivre sur un temps long l'évolution – à la fois collective et individuelle – du volume et de la fréquence de production des tweets politiques. Le chercheur peut connaître le nombre de tweets politiques publiés par l'ensemble du panel en t_0 et $t+1$, connaître le nombre total d'individus qui ont produit ces tweets en t_0 et $t+1$, et surtout connaître le nombre d'individus *identiques* qui ont produit ces tweets en t_0 et $t+1$. Avec une mesure non échantillonnée du réseau social, le chercheur peut constater une hausse sensible d'individus ayant produit un tweet politique en $t+1$ par rapport à t_0 . Mais cela ne signifie pas nécessairement que les individus qui ont produit un tweet politique en t_0 l'ont fait en $t+1$ ⁵⁰². Autrement dit, une mesure non échantillonnée du réseau social ne permet pas d'objectiver le *turn-over* de production de tweets politiques. Une analyse panéalisée de Twitter le permet. Au niveau individuel, on peut connaître à quels moments précis un individu a produit une opinion politique sous forme de tweet sur les douze derniers mois.

Notre dispositif méthodologique permet de plus de rapporter les opinions publiées et enregistrées par le logiciel aux caractéristiques sociologiques de leurs auteurs. Il est alors possible d'effectuer des tris croisés à partir des informations socio-démographiques *déclarées* par les enquêtés d'une part et les informations *codifiées* par le chercheur d'autre part. On peut ainsi par exemple codifier la tonalité des tweets mentionnant un homme politique donné puis ventiler ces résultats en fonction des variables sociodémographiques des enquêtés. Le chercheur peut ainsi connaître le pourcentage global d'opinions positives sur une personnalité politique, mais également connaître le pourcentage d'opinions positives chez les ouvriers ou chez les cadres (actifs sur le réseau social), chez les électeurs de gauche ou de droite, etc. La méthode du panel permet enfin suivre l'évolution de ces données croisées sur le temps long, au niveau à la fois individuel et collectif.

Enfin, notre dispositif méthodologique permet d'obtenir un nombre important d'informations comparables sur une population à la fois suffisamment nombreuse pour autoriser le traitement statistique et suffisamment réduite pour permettre au chercheur de la qualifier « manuellement ». Cette volonté de travailler sur un échantillon raisonné s'oppose à la démarche adoptée par la majorité des recherches quantitatives sur Twitter⁵⁰³. Ces travaux reposent en effet sur des recensions exhaustives de tweets publiés sur un sujet donné, et analysent ainsi des corpus gigantesques de plusieurs millions de messages. La taille et le

⁵⁰² De la même manière, dans le cas des consultations électorales, une hausse de participation entre les deux tours d'une élection ne signifie pas que tous les votants du premier tour aient participé au second, on enregistre alors un *turn-over*.

⁵⁰³ Pour avoir un aperçu de ces travaux Cf. chapitre 1.

caractère exhaustif des corpus collectés semblent être un véritable motif de satisfaction scientifique pour ces recherches. Tout se passe comme si sur internet, les opérations d'échantillonnage étaient devenues totalement hors de propos, voire obsolètes, alors même que les sondeurs et les quantitativistes n'ont eu de cesse de les légitimer scientifiquement durant des dizaines d'années. La nature numérique du terrain et les logiciels de veille et de *tracking* rendent en effet réalisable cet autre vieux rêve positiviste : celui d'analyser *le tout*, et non uniquement *la partie*. Sur internet, le chercheur peut en effet s'affranchir de l'opération d'échantillonnage, moyennant rappelons-le l'impossibilité de caractériser socio-démographiquement la population étudiée.

Le véritable paradoxe de cette quête d'exhaustivité est que les logiciels de veille et de *tracking* permettent en réalité de réaliser un échantillonnage aléatoire de façon beaucoup plus rigoureuse statistiquement que les sondages. Ces logiciels permettent en effet de générer une liste exhaustive des comptes qui ont émis un tweet mentionnant une occurrence donnée sur une plage de temps donnée. À partir de cette liste exhaustive, le chercheur peut échantillonner aléatoirement cette population. L'échantillon retenu sera alors tout à fait représentatif de la population totale.

Ces opérations de recension exhaustives de la population et d'échantillonnage ont été précisément les premières étapes de la constitution de notre panel.

3. Les étapes de constitution du panel

Nous allons évoquer ici les quatre étapes qui ont présidé à la constitution de notre panel. Bien que techniques, chacune de ces étapes repose sur des hypothèses et des partis-pris théoriques qui méritent d'être explicités et discutés.

4.3.1 Définition et délimitation de la population-mère

La première question posée est de savoir quelle est précisément la population-mère étudiée. Instinctivement, on pourrait postuler qu'il s'agit de l'ensemble des individus disposant d'un compte Twitter. Or, nous l'avons vu plus haut, il n'existe aucune liste

exhaustive des individus inscrits sur le réseau social. En revanche, il est possible, avec les logiciels de veille et de *tracking*, de générer une liste exhaustive d'utilisateurs ayant publié un tweet mentionnant un mot-clé donné. Tout l'enjeu est alors de déterminer quels mots-clés seront retenus pour définir notre population-mère. Initialement, nous avons pensé choisir des mots-clés « neutres » socialement et politiquement. Deux raisons nous ont conduit à abandonner cette idée. D'une part, il est en réalité assez difficile de trouver des mots-clés « neutres » socialement (le mot « bonjour », d'apparence neutre, n'est pas employé par tous les groupes sociaux, etc.). D'autre part, en choisissant des mots-clés « non politiques », on prenait le risque d'intégrer dans le panel des individus qui ne publieraient peut-être jamais de tweets politiques. Un tel résultat serait loin d'être inintéressant sociologiquement. Cependant, nous avons besoin, pour valider nos hypothèses de recherche, de collecter un nombre suffisant de tweets politiques pour autoriser le traitement statistique. Or, nous avons d'emblée fait le choix de limiter la taille du panel à 1 500 individus tout au plus⁵⁰⁴, constatant que la caractérisation sociologique des comptes serait extrêmement coûteuse en temps. Nous avons donc fait le choix de scinder conceptuellement en deux la population des individus inscrits sur Twitter : une première sous-population composée d'individus qui ont des chances mêmes minimales de publier un tweet politique et une seconde composée d'individus qui, au contraire, ont très peu voire aucune chance de publier un tweet politique. La première sous-population constituera notre population-mère. La question est alors de savoir comment mesurer les chances d'un individu de publier ou non des tweets politiques. Ne pouvant pas *mesurer* les chances de publication à proprement parler, nous avons choisi de répondre à cette question en formulant l'hypothèse suivante : si un individu disposant d'un compte Twitter actif⁵⁰⁵ ne publie pas de tweet politique durant une séquence temporelle marquée par une très forte politisation des discussions politiques ordinaires hors-ligne, il aura alors très peu de chance d'en publier par la suite, en période de plus faible politisation des débats. Plusieurs travaux ethnographiques⁵⁰⁶ ont démontré qu'en France, la séquence électorale présidentielle était une période de très forte politisation des discussions ordinaires, une séquence où « *le temps politique se synchronise avec le temps social* »⁵⁰⁷. Une séquence où la probabilité d'émettre un tweet politique est donc sans doute la plus forte. Afin de repérer les individus qui

⁵⁰⁴ En se référant à la table de Gausse et à la théorie des grands nombres, ce nombre serait suffisant pour que l'échantillon soit représentatif de la population totale, ici la population qui a des individus inscrits sur Twitter qui ont des chances de publier des tweets politiques.

⁵⁰⁵ Cf. Glossaire p. 464.

⁵⁰⁶ Vincent Goulet, *Médias et classes populaires. op. cit.* ; Céline Braconnier et Jean-Yves Dormagen, *La démocratie de l'abstention, op. cit.*

⁵⁰⁷ Vincent Goulet, *Médias et classes populaires, op. cit.*

ont publié un tweet politique durant cette séquence électorale et ainsi délimiter notre population-mère, nous avons enregistré et archivé de manière exhaustive l'ensemble des tweets mentionnant le nom d'au moins un des dix candidats à l'élection présidentielle, et ce sur une période d'un mois, du 1^{er} au 31 mars 2012⁵⁰⁸. Les mots-clés retenus pour collecter de façon exhaustive les tweets dits « politiques » ont donc été les noms des dix candidats à l'élection. La sélection des tweets par l'entrée « candidats » (et non par exemple par une entrée thématique ou par une entrée « partis politiques ») a été motivée par la volonté d'adopter, dans une perspective comparative, une base de données similaire à celles des principales études commerciales⁵⁰⁹ et académiques⁵¹⁰. Or, dans leur majorité, ces études ne prennent pour objet que les tweets évoquant le nom des candidats aux élections. Cette réduction des tweets politiques aux seuls noms des candidats n'est selon nous pas préjudiciable d'un point de vue sociologique. Si, comme c'est le cas dans cette recherche, nous adoptons une définition volontairement restrictive et dominante de la politique, on peut émettre l'hypothèse selon laquelle une sélection des tweets par l'entrée « candidats » – de surcroît dans une élection aussi personnalisée que l'élection présidentielle française – n'exclura pas de notre base de données un nombre significatif de tweets politiques. Dit autrement, on peut penser que les chances de parler de l'élection présidentielle sans en évoquer les candidats sont faibles. Chaque requête comprenait le nom du candidat, une configuration de noms mal orthographiés (afin que la requête soit la moins discriminante socialement – sarkozy OR sarkozi, etc.), les éventuels surnoms du candidat (« sarko ») et les principaux hashtags⁵¹¹ de campagne (« ns2012 », « fh2012 », etc.). Sur la période du 1 au 31 mars 2013, le logiciel a au final enregistré 2 807 612 tweets, provenant de 248 628 comptes uniques. Ces 248 628 comptes ont donc publié au moins un tweet mentionnant l'un des dix

⁵⁰⁸ Cette plage temporelle est discutable. On peut objecter qu'elle aurait pu être prolongée jusqu'au 6 mai 2012, jour du second tour, en avançant avec raison l'argument selon lequel c'est à quelques jours du premier tour de scrutin et durant la campagne d'entre-deux-tours que l'intérêt pour la campagne électorale est à son maximum, et que la probabilité de publier un tweet politique est donc la plus forte. Trois raisons ont motivé cette couverture circonscrite au seul mois de mars. La première repose sur une volonté de comparer les résultats obtenus avec ceux d'autres études commerciales publiées dans les médias. Or, dans la majorité de ces études, c'est une période mensuelle qui est retenue. Et le mois de mars était le dernier où l'ensemble des candidats était recensé. La deuxième raison tient dans la volonté de commencer la construction du panel avant la fin de la campagne présidentielle, afin de le rendre opérationnel dès la fin du second tour. Une mauvaise appréciation du temps nécessaire à la constitution du panel a finalement retardé sa « mise en service ». Enfin et surtout, collaborant avec un ingénieur dans cette phase de collecte, nous étions dépendant de son aide et ne pouvions générer seul ces recensions exhaustives de centaines de milliers de tweets.

⁵⁰⁹ On peut citer entre autres le « Politivox » de libération.fr, le « Twittoscope » du journal gratuit Métro, « Le Scan » du figaro.fr, ou encore le « Twittrack2012 » d'atlantico.fr.

⁵¹⁰ Cf. chapitre 1.

⁵¹¹ Cf. glossaire p. 464.

candidats à l'élection présidentielle. En suivant notre hypothèse, ces comptes avaient donc une chance supérieure à zéro de publier un tweet politique.

4.3.2 Échantillonnage et analyse d'éligibilité

Une rapide lecture des profils de notre base de données a permis de nous rendre compte qu'un certain nombre de ces comptes étaient en réalité des comptes de médias, d'organisations politiques, d'entreprises, etc. Une part non négligeable de ces comptes était de plus rédigée en langue étrangère. Or, nous avons conçu notre dispositif méthodologique et nos hypothèses de recherche dans l'optique d'étudier des individus, des agents sociaux, de choisir en clair la même unité statistique que les sondages. Trois critères d'éligibilité au panel ont été alors instaurés. Pour être éligible : 1) le compte devait être « ouvert » au moment de l'analyse d'éligibilité, 2) les tweets publiés par le compte devait être majoritairement rédigés en langue française, 3) le compte devait être tenu par un individu unique, anonyme ou non. Nous avons ensuite extrait aléatoirement, à partir de notre base de données, une liste de 20 000 comptes. Ces 20 000 comptes ont été évalués « manuellement » afin de vérifier s'ils étaient éligibles au panel. L'analyse d'éligibilité a débuté le 26 avril 2012 et s'est terminée le 28 août 2012. Le premier élément de filtrage de cette base a donc été d'exclure les comptes temporairement inactifs ou définitivement supprimés au moment de l'analyse manuelle des comptes. Un délai s'est instauré entre la période de collecte et de constitution de la base (finalisée le 31 mars 2012) et le début de l'analyse exhaustive des comptes (cette phase a débuté près d'un mois après, le 26 avril 2012). Entre ces deux dates, l'utilisateur a pu désactiver ou supprimer son compte. Ce sont au final près de 8,2% des comptes qui n'ont pas été trouvés sur le site de micro-blogging au moment de l'analyse d'éligibilité. Il est très difficile de connaître les raisons et motivations de ces suppressions de compte, les individus étant désormais injoignables.

Tableau 2.1 Activité des comptes soumis à l'analyse d'éligibilité au panel

Activité des comptes	Base extraite <i>n</i> =20 000		Base totale (extrapolation) <i>n</i> =248 628
	Effectifs	%	Effectifs
Comptes actifs	18 366	91,8	228 315
Comptes suspendus ou supprimés	1 634	8,2	20 313
Total	20 000	100,0	248 628

Le choix de ne pas introduire une restriction de langue *a priori*, au moment de la collecte des tweets, a été une contrainte imposée par notre logiciel. Le programme ne permettait pas, à l'époque de la mise en place du protocole de collecte, de paramétrer un filtrage de langue automatisé des tweets. Ce filtrage par langue a donc été effectué de façon manuelle et *a posteriori*, au moment de l'analyse d'éligibilité. Bien qu'ils aient publié des tweets sur la vie politique française, nous avons fait le choix d'exclure les comptes en langue étrangère, pour des raisons de compétences linguistiques mais également par refus de mener une analyse comparative transnationale. En effet, bien que la nature virtuelle du terrain d'enquête mette à égale distance le chercheur d'un enquêté chinois ou français, il n'en demeure pas moins que les cadres conceptuels et analytiques mobilisés ne sont pas aisément transposables d'un pays à l'autre. La question fut ensuite de savoir si, parmi les comptes francophones, seuls les comptes français devaient être éligibles. Le refus d'adopter une analyse comparatiste incitait logiquement à ne garder que les comptes français et donc à exclure les autres comptes francophones (maghrébins, africains sub-sahariens et canadiens pour l'essentiel). Nous avons néanmoins choisi de rendre éligibles l'ensemble des comptes francophones, pour ne pas exclure *a priori* les « français de l'étranger », difficilement différenciables des étrangers francophones. Il a néanmoins été facile d'isoler par la suite les comptes non-français.

Tableau 2.2 Langue des comptes soumis à l'analyse d'éligibilité au panel

Compte francophones/étrangers	Base extraite n = 18 366		Base totale (extrapolation) n = 228 315
	Effectifs	%	Effectifs
Comptes francophones	10 856	59,1	134 955
Compte étrangers	7 510	40,9	93 360
Total	18 366	100,0	228 315

Enfin, le dernier critère d'éligibilité concernait la nature du compte recensé. De nombreuses organisations, institutions, médias, partis politiques et sociétés possèdent un compte Twitter et ont, pour différentes raisons, mentionné le nom d'un ou plusieurs candidats à l'élection présidentielle. C'est pour cette raison que leurs tweets ont été scannés par le logiciel et que leurs comptes se trouvent recensés au sein de notre base de données. Nous avons fait le choix d'exclure ces comptes et de ne conserver que des comptes individuels. Ce choix a été motivé par la volonté de conserver la même unité statistique que les sondages d'opinion : l'individu. Pour autant, nous le verrons par la suite, ce choix est discutable : exclure les comptes de médias et d'organisation politiques, et ne garder que des comptes de « citoyens » revient d'une certaine manière à nier ou occulter la réalité sociale de Twitter. L'opinion mesurée sur le réseau social est directement et indirectement le produit des activités de publication des médias et des organisations politiques. Autrement dit, Twitter met au jour ce que les sondages avaient dissimulé : le rôle central des institutions productrices d'opinion dans la formation de l' « opinion publique ». Or une analyse panéalisée centrée exclusivement sur des individus « citoyens » ne tient pas compte de cette réalité. Cependant, nous le verrons, l'analyse de notre panel nous conduira à « réintégrer » à notre dispositif empirique ces comptes de médias et d'organisations politiques.

Tableau 2.3 Nature des comptes soumis à l'analyse d'éligibilité au panel

Nature des comptes	Base extraite <i>n</i> = 10 856		Base totale (extrapolation) <i>n</i> = 134 955
	Effectifs	%	Effectifs
Individu	10 299	94,9	128 031
Média	229	2,1	2 847
Organisation	230	2,1	2 859
Entreprise	98	0,9	1 218
Total	10 856	100,0	134 955

Ce sont au final 10 299 comptes (51,4 % des comptes de la base de données initiale) qui ont été déclarés éligibles au panel. Il s'agit de comptes actifs, rédigés en langue française et tenus par des individus uniques. 72 % de ces comptes ont émis un tweet (politique ou non) dans les quinze jours précédant leur analyse d'éligibilité. Les comptes qui n'ont pas émis de tweets depuis plusieurs semaines voire plusieurs mois ont été néanmoins conservés. Ce choix repose sur l'hypothèse selon laquelle des individus peuvent ne publier des tweets que très rarement, à l'occasion de grands événements tels que des élections par exemple. Il nous est arrivé d'observer sur certains comptes des délais de plusieurs mois entre la publication de deux tweets. Aucune discrimination n'a été faite quant à la fréquence d'émission des tweets. Un risque demeure cependant : celui de déclarer éligibles des comptes « fantômes », créés artificiellement par des militants durant la campagne et qui ont été abandonnés (sans être supprimés) par la suite.

Tableau 2.4 Date du dernier tweet publié par les individus éligibles au panel

Dernier tweet émis	Effectifs	%
Dans les 15 derniers jours	6108	72
Entre 15 jours et un mois	562	7
Entre 1 et 2 mois	599	7
Entre 2 et 3 mois	449	5
Entre 3 et 4 mois	429	5
Supérieur à 4 mois	301	4
Total	8448	100

4.3.3 Administration du questionnaire et taux de réponse

Un court questionnaire⁵¹² a été ensuite administré en ligne aux 10 299 individus éligibles. Il est composé de 20 questions relatives à des données sociodémographiques (âge, sexe, ville, niveau d'étude, statut, PCS, etc.) et politiques (préférence partisane, participation aux élections, expériences politiques conventionnelles et non-conventionnelles, etc.). La focalisation quasi-exclusive du questionnaire sur des questions « signalétiques », socio-démographiques, se justifie par la volonté de proposer un questionnaire le plus court possible, afin de ne pas décourager les individus les moins enclins à répondre. D'autre part, nous n'avons pas jugé utile de poser des questions relatives aux habitudes et comportements des enquêtés, ces informations pouvant être précisément tirées de l'observation directe des panels⁵¹³. Administrer le questionnaire aux individus éligibles a constitué une réelle difficulté. Ne disposant pas des adresses électroniques des individus éligibles, nous ne pouvions pas envoyer le questionnaire de manière ciblée et personnalisée. Nous avons alors créé un profil Twitter où était déclinée notre identité et où était présentée notre démarche de recherche : « *Doctorant en science politique, je cherche à dresser un portrait sociologique des réseaux sociaux. Merci de m'aider en répondant à mon questionnaire : lien URL du questionnaire* ». Nous nous sommes alors abonné aux comptes des 10 299 individus éligibles.

Trois filtres s'interposent entre la demande d'abonnement et la réponse au questionnaire :

1) L'individu doit prendre connaissance de notre abonnement à son compte. Cette information est donnée automatiquement par Twitter dans la rubrique « @connecter ». L'individu doit donc consulter cette rubrique pour visualiser l'information. Si d'autres utilisateurs s'abonnent dans les jours suivant ou précédant notre abonnement, notre nom d'utilisateur peut ne pas apparaître (« *jocelyne Herrera et 9 autres vous ont suivi* »). On peut émettre l'hypothèse selon laquelle plus un individu a de nouveaux abonnés, moins l'information de notre abonnement à son compte est visible.

2) L'individu doit ensuite consulter notre profil. En prolongeant l'hypothèse précédente, on peut penser que plus un individu a de nouveaux abonnés, moins il prendra le

⁵¹² Cf. Annexe 2 p. 467. Le questionnaire a été conçu et hébergé sur le site *Google documents*.

⁵¹³ Même si l'on peut objecter qu'il aurait été intéressant de comparer données déclarées et données objectivées.

temps de consulter un à un leur profil. De plus, il n'est pas dit qu'un individu aille consulter le profil d'un individu qu'il ne connaît pas et dont il ne sait rien.

3) Enfin, après avoir consulté notre profil, l'individu doit ouvrir l'adresse URL et répondre au questionnaire. Une majorité d'individu a consulté notre profil sans répondre au questionnaire. Un certain nombre s'est même abonné en retour à notre compte. Le fait de s'abonner à notre compte témoigne bien du fait que l'individu n'a pas émis d'hostilité à notre démarche, et que sa non-réponse ne signifie pas nécessairement un refus de répondre. Pour diverses raisons, l'individu peut ne pas se sentir concerné ou autorisé à répondre au questionnaire. Lorsque les individus se sont abonnés à notre compte sans avoir répondu au questionnaire, nous avons néanmoins eu la possibilité de leur envoyer un message privé et personnalisé *via* la messagerie interne du site en les invitant à répondre au questionnaire⁵¹⁴.

711 individus ont répondu au questionnaire en ligne. Cependant, 53 individus étaient hors-cible (des individus qui ont répondu au questionnaire alors qu'ils ne figuraient pas dans la liste des individus éligibles). 50 autres n'ont pas dévoilé leur pseudonyme et ont donc refusé de faire partie du panel.

Ce sont donc au final 608 individus qui ont intégré le panel, soit 5,9% des individus éligibles⁵¹⁵.

4.3.4 Constitution du panel de contrôle des individus qui n'ont pas répondu au questionnaire

Enfin, la dernière étape de notre dispositif méthodologique a consisté à élaborer un panel de contrôle, composé d'un échantillon représentatif d'individus qui n'ont pas répondu à notre questionnaire. Un des avantages de notre dispositif est en effet de pouvoir caractériser – partiellement – les individus qui ont refusé de participer à l'enquête.

À partir de la liste des comptes éligibles qui n'ont pas répondu au questionnaire ($n = 9\ 641$) nous avons extrait un échantillon aléatoire de 620 comptes. Nous avons ensuite analysé

⁵¹⁴ Il est possible d'envoyer un message privé sur Twitter uniquement lorsque le destinataire du message est abonné à l'expéditeur.

⁵¹⁵ À partir d'un seuil de 500 répondants, les résultats enregistrés ne varient pratiquement plus. On a alors émis l'hypothèse selon laquelle la population des répondants était significativement représentative de la population des twittos (pré)disposés à répondre au questionnaire.

manuellement un à un ces comptes afin de recueillir le plus d'informations possible : âge, sexe, profession, niveau de diplôme, etc.

Nous avons collecté ces informations à partir d'une analyse des profils des comptes Twitter, d'une analyse du contenu des tweets publiés et enfin à partir d'un profilage et d'un croisement d'informations collectées sur d'autres réseaux sociaux et/ou sites internet sur lesquels sont inscrits les enquêtés⁵¹⁶.

Tout individu disposant d'un compte Twitter est invité au moment de son inscription⁵¹⁷ à écrire une courte présentation (de moins de 140 caractères). Il est également demandé d'indiquer son lieu d'habitation et un lien vers un site personnel (ou un site que l'on souhaite promouvoir). Ces informations auto-déclarées par les individus constituent une première modalité de caractérisation sociologique. Il est ainsi souvent indiqué dans les quelques lignes de présentation du compte le lieu de résidence de l'enquêté et sa profession ou son niveau d'étude (les étudiants communiquent majoritairement cette information).

Analyser le contenu des tweets publiés peut être également un moyen d'obtenir des informations sociologiques sur les enquêtés. Les tweets peuvent en effet contenir des informations relatives à la profession ou au statut social de l'enquêté (« je file en cours » => étudiant), à ses préférences politiques ou à d'autres informations sociodémographiques.

Enfin, nous avons également obtenu des données sociologiques en recherchant des informations complémentaires sur d'autres sites internet ou d'autres réseaux sociaux sur lesquels peuvent être également inscrits les enquêtés. Les internautes peuvent être plus prolixes sur certains autres réseaux sociaux. Ils renseignent ainsi en moyenne davantage leur date de naissance sur Facebook. Sur les réseaux socio-professionnels, où les profils s'apparentent à des *curriculum vitae* en ligne, les individus indiquent, parfois de façon très précise, leur parcours scolaire et professionnel⁵¹⁸.

Au final, nous avons pu identifier la PCS d'un non-répondant sur deux, et le niveau de diplôme d'un peu plus d'un non-répondant sur trois.

⁵¹⁶ Cf. chapitre 1.

⁵¹⁷ Le renseignement de ces informations n'est pas obligatoire.

⁵¹⁸ Cependant, nous le verrons par la suite, les individus qui ont un compte sur ces réseaux sociaux professionnels sont très majoritairement diplômés et occupent des emplois de cadre.

Ces informations ont permis de caractériser les individus qui n'ont pas répondu au questionnaire et donc de les comparer au panel « répondants ». Nous le verrons dans le prochain chapitre, ces individus ne présentent pas tout à fait le même profil sociologique que les enquêtés répondants. Nous avons également panéalisé ces 620 comptes et suivi leur production.

Tableau 2.5 Taux de remplissage des informations socio-démographiques des non-répondants tirés au sort

Taux de remplissage	Effectifs	%
Sexe	590	95,2
Âge déclaré	72	11,6
Âge estimé	240	38,7
Lieu de résidence	373	60,2
Niveau de diplôme	225	36,3
PCS	326	52,6
Statut	331	53,4
Appartenance politique	264	42,6

Nous avons donc au final suivi l'activité de deux panels : un panel de 608 individus qui a répondu au questionnaire et dont on connaît précisément et pour chacun d'eux les attributs sociologiques (que l'on nommera désormais le panel « répondants ») ; et un panel de 620 individus qui n'ont pas répondu au questionnaire et dont on connaît pour certains des informations sociodémographiques (le panel « non répondants »).

Conclusion du chapitre 2

Dans ce deuxième chapitre, nous avons tenté de démontrer que le cadre conceptuel et théorique de la sociologie critique permettait de donner sens aux instruments de mesure des opinions numériques. Plus encore, ces nouvelles technologies d'étude permettent de matérialiser empiriquement certaines hypothèses de la théorie dispositionnelle des modes de production des opinions politiques.

Notre dispositif méthodologique entend concilier les exigences théoriques de la sociologie critique et les potentialités techniques offertes par ces nouveaux instruments de mesure. Pour ce faire, nous avons emprunté une « troisième voie » méthodologique, refusant d'une part une approche qualitative et localisée qui ne permet pas le traitement statistique et renonçant d'autre part à une analyse exhaustive de très importants volumes de données qui ne permet pas de qualifier sociologiquement la population étudiée.

En échantillonnant la population des individus inscrits sur le réseau social Twitter (individus qui ont une probabilité de publier des messages politiques) et en leur administrant un court questionnaire comprenant une vingtaine de questions socio-démographiques, notre dispositif méthodologique permet de concilier l'exigence de qualification sociologique des producteurs d'opinion et celle du traitement statistique.

Les chapitres suivants tenteront de vérifier empiriquement les hypothèses évoquées ici en analysant de manière longitudinale l'activité de production du panel.

**SECONDE PARTIE : TWITTER, UN OBSERVATOIRE DU MARCHÉ
DES OPINIONS POLITIQUES**

Dans cette seconde partie, nous chercherons à démontrer, à l'appui de données générées par notre dispositif méthodologique, que Twitter constitue un nouvel observatoire du « marché » des opinions politiques, un « marché » où se rencontrent une « offre » et une « demande ».

Notre verrons d'abord que la majorité des individus qui publient des messages politiques sur Twitter présentent des caractéristiques sociologiques et politiques très singulières. Ils sont ainsi issus d'un milieu social relativement aisé et sont très fortement politisés (chapitre 3).

Cependant, bien que fortement politisés, leur propension à publier des tweets politiques est corrélée au niveau d'intensité de production de « biens » politiques par le champ de production de l'information et de l'opinion – « l'offre » (chapitre 4).

De plus, les informations et les opinions évoquées ou formulées dans les messages de nos enquêtés sont dans leur grande majorité des « biens » produits et diffusés par les agents et les institutions du champ de production de l'information et de l'opinion (chapitre 5).

CHAPITRE 3 – TWITTER, UN NOUVEL OBSERVATOIRE

DU « PAYS LÉGAL »

Un mois avant le premier tour de l'élection présidentielle française du 22 avril 2012, du 1^{er} au 31 mars, 128 031 individus ont publié *spontanément* et *publiquement* un ou plusieurs tweets (ou retweets) en langue française mentionnant le nom d'un candidat. Plus de 100 000 individus se sont donc sentis autorisés à s'exprimer politiquement sur Twitter.

Pierre Bourdieu qualifie de membres du « pays légal » l'ensemble des agents sociaux qui se sentent autorisés et appelés à exprimer leur opinion politique, « *même de la manière la plus rudimentaire* »⁵¹⁹. Dans ce troisième chapitre, nous soutiendrons que Twitter constitue un véritable observatoire du « pays légal ».

Le réseau social offre en effet la possibilité d'analyser à très grande échelle et de manière relativement inédite une population sociologiquement atypique. Les individus qui ont publié des tweets politiques durant la dernière campagne présidentielle présentent en effet des propriétés sociologiques tout à fait singulières.

Ordinairement, deux postures méthodologiques permettent d'étudier les propriétés des individus qui se sentent autoriser à prendre la parole politique. La première consiste à mesurer, à partir d'un *échantillon représentatif*, la part des différentes catégories sociales qui accepte de répondre – ou non – à des questions politiques⁵²⁰ ou de participer à des consultations électorales⁵²¹.

La seconde posture consiste à établir une sociographie d'un *échantillon* dit *spontané* d'individus qui acceptent de dévoiler d'eux-mêmes leurs opinions politiques, par exemple en répondant volontairement à des enquêtes par questionnaire publiées dans les médias⁵²², en

⁵¹⁹ Pierre Bourdieu, *La Distinction. Critique sociale du jugement*, Paris, Éditions de Minuit, 1979, p. 480.

⁵²⁰ On pense là bien sûr aux très nombreuses recherches sur les « non réponses » aux sondages d'opinion. On peut citer entre autres Guy Michelat et Michel Simon, « Les “sans réponse” aux questions politiques », *Pouvoir*, n°33, 1985, p. 41-56, ou Daniel Gaxie, *Le cens caché. Inégalités culturelles et ségrégation politique*, Paris, Seuil, 1978.

⁵²¹ L'abstention, ou le « non vote » est fortement déterminé socialement. Cf. par exemple Céline Braconnier et Jean-Yves Dormagen, *La démocratie de l'abstention : Aux origines de la démobilisation électorale en milieux populaires*, Paris, Gallimard, 2007.

⁵²² On peut citer ici l'exemple de l'enquête sur l'enseignement supérieur que Bourdieu mobilise dans *La Distinction*. Organisée par l'Association d'étude pour l'expansion de la recherche scientifique, cette enquête a

participant à des manifestations politiques ou encore à des dispositifs de démocratie participative⁵²³.

Pour Pierre Bourdieu, « *l'échantillon spontané est la limite vers laquelle tend la population des répondants d'une enquête auprès d'un échantillon représentatif* »⁵²⁴. C'est cette hypothèse que nous allons discuter ici. Les individus qui publient spontanément leurs opinions politiques sur Twitter sont-ils les mêmes que ceux qui acceptent de répondre aux enquêtes d'opinion ou de participer à des réunions politiques ?

Répondre à cette question nécessite de prendre en considération trois caractéristiques propres à notre échantillon. La première est que les individus qui ont publié un tweet politique n'ont pas été directement incités à le faire. Ils n'ont participé à aucune enquête⁵²⁵ ni à aucun dispositif politique consultatif. Aucune question ni aucun cadrage thématique ne leur ont été directement imposés⁵²⁶. Il s'agit donc d'un échantillon *doublement spontané*. Ces individus ont *spontanément* publié un message politique à partir d'une thématique ou d'une problématique qu'ils ont eux-mêmes *spontanément* définie.

La deuxième particularité de notre échantillon a trait à sa taille. Il est composé, rappelons-le, de plus de 100 000 individus. Si elle ne le rend pas davantage représentatif, la taille importante de notre échantillon permet néanmoins d'observer avec plus de précision et de finesse les différences qui existent au sein même de cette population sociologiquement atypique.

Enfin, une troisième caractéristique – qui nuance et précise celle précédemment évoquée – est que notre échantillon est en réalité un *échantillon représentatif* de cet *échantillon spontané*. Pour des raisons méthodologiques, exposées lors du précédent chapitre, nous avons en effet décidé d'échantillonner aléatoirement cet *échantillon spontané* de plus de 100 000 individus. Nous avons administré un questionnaire d'identification sociodémographique à près de 10 300 individus tirés au sort parmi la population totale. 658

été réalisée à partir d'un questionnaire sur l'enseignement publié dans un très grand nombre de titres de presse. Bourdieu a étudié les classes et fractions de classe qui ont répondu au questionnaire. Cf. *La Distinction*, p. 480.

⁵²³ La majorité des travaux ayant établi une sociographie de ces dispositifs constatent généralement la sous-représentation des catégories les plus jeunes, les moins diplômées, et les plus exclues socialement. Cf. entre autres Daniel Gaxie, *Le cens caché, op. cit.*

⁵²⁴ Pierre Bourdieu, *La Distinction, op. cit.*, p. 482.

⁵²⁵ En mars 2012, lorsque les futurs membres de notre panel ont publié un ou plusieurs tweets politiques, l'existence de notre dispositif ne leur était pas connue.

⁵²⁶ En revanche, des problématiques ou des cadrages thématiques peuvent s'imposer *indirectement* aux producteurs de tweets, *via* leur consommation de biens informationnels par exemple. Cette question sera au cœur des chapitres 3 et 4.

individus ont répondu à notre questionnaire (soit 6 % de la population). Rien n'indique *a priori* que ces individus « répondants » présentent les mêmes caractéristiques sociologiques que les individus qui n'ont pas répondu au questionnaire. Pour contrôler ce risque propre à toute enquête par questionnaire, nous avons donc également analysé un échantillon aléatoire de 620 individus qui n'ont pas répondu. La comparaison de ces deux sous-échantillons – « répondants » et « non répondants » – a permis de caractériser plus finement notre population.

Nous démontrerons dans une première section que les producteurs de tweets politiques présentent des propriétés sociologiques et politiques singulières, relativement proches de celles des membres du « pays légal ». Ils forment une population relativement homogène sociologiquement et politiquement (1). Cependant, nous verrons dans un second temps que l'on peut observer certaines différences au sein de cette population. Ces différences se manifestent d'abord par la manière dont les individus se mettent – ou non – en scène et en récit à travers leur « profil »⁵²⁷, laissant paraître certaines dispositions constitutives de leur habitus. De manière analogue, la place et la fonction que les membres de notre panel assignent à la politique trahit là encore certaines différences dispositionnelles (2).

⁵²⁷ Cf. glossaire. p. 464.

1. Les propriétés sociologiques et politiques singulières des producteurs de tweets politiques

Malgré certaines spécificités que nous tenterons de mettre en lumière, les producteurs de tweets politiques présentent des propriétés sociologiques et politiques relativement proches de celles des membres du « pays légal »⁵²⁸.

Ils sont ainsi davantage masculins, diplômés, Parisiens et occupent une position élevée dans la hiérarchie sociale. Cependant, leur appartenance aux catégories les plus « jeunes » de la population constitue une certaine spécificité (1.1).

D'un point de vue politique, les membres de notre panel témoignent d'un intérêt élevé pour la politique, participent assidûment aux élections et sont très souvent engagés politiquement. La surreprésentation des électeurs de gauche constitue une autre spécificité de notre échantillon (1.2).

1.1 Les propriétés sociologiques des producteurs de tweets politiques

Près d'un producteur de tweets politiques sur deux a moins de 25 ans. Ce résultat différencie *a priori* notre échantillon des populations participationnistes et répondant aux études sociologiques classiques. Ce sont en effet les catégories les plus jeunes de la population qui participent le moins politiquement⁵²⁹. Cependant nous démontrerons que c'est en réalité la fraction la plus politisée – et donc sociologiquement la moins représentative – de la jeunesse qui s'exprime politiquement sur Twitter (1.1.1).

Les producteurs de tweets politiques occupent en réalité des positions élevées dans l'espace social. Les (filles et fils de) cadres, professions intellectuelles supérieures et professions libérales sont en effet nettement surreprésentés au sein de notre échantillon. Occupant souvent des professions « *de présentation et de représentation* » en lien avec la communication, les études de marché ou encore le marketing, ces individus semblent

⁵²⁸ Pierre Bourdieu, *La Distinction, op. cit.*, p. 480.

⁵²⁹ Anne Muxel, *L'expérience politique des jeunes*, Paris, Presses de Science Po, 2001, p. 42.

s'apparenter à la « *petite bourgeoisie nouvelle* » décrite par Pierre Bourdieu dans *La Distinction*⁵³⁰ (1.1.2).

Cette position sociale élevée donne alors leur forme spécifique à d'autres propriétés secondaires telles que le sexe, le lieu d'habitation ou le niveau de diplôme des producteurs de tweets politiques. La surreprésentation des hommes, des plus diplômés et des Parisiens au sein de notre échantillon peut ainsi s'expliquer par une prédominance des catégories sociales supérieures : on sait en effet que la catégorie des cadres est composée très majoritairement d'hommes et qu'une grande partie des professions « *de présentation et représentation* » est située géographiquement au sein de la capitale (1.1.3).

1.1.1 Une jeunesse surreprésentée mais non représentative

La propriété *a priori* la plus remarquable des membres de notre panel est en effet leur jeune âge. Les catégories les plus jeunes de la population sont nettement surreprésentées au sein de notre échantillon. Alors qu'en 2012 les 18-25 ans représentaient 9,7 % de la population française, 39,8 % des individus qui ont répondu au questionnaire et 45,8 % de ceux qui n'ont pas répondu (et dont l'âge a pu être identifié) appartenaient à cette classe d'âge.

Tableau 3.1 Répartition par classe d'âge (en %) des populations « répondants » et « non répondants »

Classe d'âge	Population Répondants (n = 658)	Population Non Répondants (n = 240*)	Population totale INSEE 2012
Moins de 18 ans	8,1	19,6	22,2
18 à 25 ans	39,8	45,8	9,7
26 à 34 ans	20,1	14,2	11,1
35 à 45 ans	18,8	14,6	14,7
46 à 60 ans	9,3	4,6	20,0
61 ans et plus	4,0	1,3	22,2
Total	100,0	100,0	100,0

*Individus dont l'âge a pu être identifié (soit 38,8% des non répondants tirés au sort)

Source : INSEE

⁵³⁰ Pierre Bourdieu, *La Distinction*, op. cit., p. 397.

La surreprésentation des moins de 18 ans et des 18-25 ans est sensiblement moins marquée chez les répondants⁵³¹. Le questionnaire semble donc avoir joué comme un filtre pour les individus les plus jeunes, et notamment pour les mineurs qui n'ont pas le droit de vote. Deux éléments de cadrage méthodologique doivent être ici abordés afin de mieux comprendre ces résultats. Il convient d'abord de préciser que bien que le questionnaire n'ait pas été réservé aux individus inscrits sur les listes électorales, ni aux individus de plus de 18 ans, un certain nombre de questions faisait néanmoins référence à des préférences politiques et à des choix électoraux passés. Ces questions, et l'orientation plus générale du questionnaire, ont donc pu laisser penser aux individus de moins de 18 ans qu'ils n'étaient pas concernés ou qu'ils n'étaient pas habilités à répondre. Certains d'entre eux nous ont d'ailleurs directement interpellés pour savoir s'ils devaient répondre au questionnaire, étant donné qu'ils étaient mineurs. On peut donc logiquement supposer que le questionnaire ait joué comme un filtre pour ces individus les plus jeunes. Un deuxième élément à prendre en compte a trait à la construction même du panel des non répondants. Tous les individus inscrits sur Twitter n'indiquent pas systématiquement leur âge sur leur profil. Nous n'avons ainsi pu identifier l'âge que de 38,8% des non répondants. Or rien n'indique que les individus qui ont indiqué leur âge soient représentatifs de l'ensemble des non répondants. Nous le verrons plus loin, indiquer son âge comme indiquer sa profession est une pratique socialement située. Les individus les plus jeunes ont ainsi d'avantage tendance à indiquer leur âge et surtout leur statut d'étudiant ou de lycéen. Or nous avons déduit l'âge des enquêtés en fonction de leur statut (lycéens, étudiants, retraités, etc.), en raisonnant en tranche d'âge estimée. Les catégories les plus âgées de la population sont donc sans doute sous-représentées – ou plutôt sous-estimées – au sein du panel non répondants.

Ce résultat peut sembler surprenant. En effet, les catégories les plus jeunes de la population sont précisément celles qui participent et qui votent le moins⁵³². C'est en effet parmi les 18-24 ans que l'on compte proportionnellement le plus d'individus qui ne sont pas inscrits sur les listes électorales et le plus d'abstentionnistes. Pourtant ce sont bien les 18-24 ans qui, à l'intérieur de notre panel, se sont le plus exprimés politiquement sur Twitter.

Ce résultat est d'autant plus inattendu que les producteurs de tweets politiques semblent encore plus jeunes que l'ensemble des individus inscrits sur Twitter (qui, rappelons-le, ne publient pas tous nécessairement des tweets politiques)⁵³³. La population de ce réseau social est pourtant dans son ensemble déjà relativement jeune : 29% des inscrits auraient entre 18 et 24 ans⁵³⁴. Au sein de notre échantillon de répondants, ce sont 41% des individus qui

⁵³¹ Afin de ne pas entraver notre développement, les paragraphes méthodologiques seront mis en retrait.

⁵³² Anne Muxel, *L'expérience politique des jeunes*, op. cit., p. 42.

⁵³³ Rappelons ici que la population des individus inscrits au réseau social a été au début de cette recherche scindée en deux, en fonction des chances de publier un tweet politique. Cf. chapitre 2.

⁵³⁴ Étude IFOP parue en 2010 : « Twitter vu par ses utilisateurs ».

appartiennent à cette même classe d'âge. Les individus qui produisent des messages politiques sont donc encore plus jeunes que la moyenne des individus inscrits sur Twitter.

Tableau 3.2 Répartition par classe d'âge (en %) des « répondants » et des inscrits sur Twitter

Classe d'âge	Population Répondants (n = 605)	Étude IFOP 2010 Twitter (n = 306)	Population française INSEE 2012
18 à 24 ans	40,7	29	11,1
25 à 34 ans	24,5	35	15,8
35 à 49 ans	24,0	30	26,2
50 à 64 ans	8,3	5	24,9
65 ans et plus	2,6	1	22,0
Total	100	100	100

Source : IFOP, INSEE

Il existe peu de données relatives à la sociologie des individus inscrits sur Twitter. Les résultats présentés ici sont issus d'une étude publiée par l'IFOP, réalisée du 18 octobre au 24 novembre 2010. La notice méthodologique de l'étude indique que « les interviews ont eu lieu par questionnaires auto-administrés en ligne » et que « les répondants ont été recrutés directement via Twitter, afin de maximiser l'échantillon » sans toutefois préciser les modalités d'échantillonnage, de « recrutement », de passation du questionnaire, et les taux d'acceptation et de refus. Ce manque de précision méthodologique et le faible effectif de l'échantillon (n = 306) invitent donc à la prudence. Précisons enfin que l'étude de l'IFOP n'a pas comptabilisé les moins de 18 ans. Pour obtenir des résultats comparables, nous avons alors exclu momentanément les moins de 18 ans de notre échantillon (53 individus).

Sur le réseau social, à l'inverse de ce que l'on observe dans le domaine électoral, les individus les plus jeunes sont donc ceux qui « participent » le plus politiquement. On pourrait alors supposer, à la lecture de ces résultats, que les « jeunes » ne seraient peut-être pas moins politisés, mais simplement moins attirés par les formes « classiques » de participation (voter, militer dans un parti, etc.) et trouveraient dans les réseaux sociaux une nouvelle forme d'« engagement » ou d'expression politique qui leur conviendrait davantage. Ainsi, pour Dominique Cardon, publier un tweet politique pourrait être considéré comme une nouvelle forme de participation politique, moins exigeante socialement et culturellement. Cette

nouvelle forme de participation, « *infime, futile et murmurante* », permettrait alors de corriger les inégalités sociales de participation politique⁵³⁵.

Les données issues de notre questionnaire semblent nuancer cette hypothèse. Les individus les plus jeunes qui ont publié des tweets politiques sont précisément ceux qui ont voté au premier tour des élections présidentielles. Alors que les Français âgés de 18 à 24 ans sont proportionnellement ceux qui ont le moins voté le 22 avril 2012 (seuls 72,4 % de cette classe d'âge a voté), les individus du même âge qui ont publié des tweets politiques ont *déclaré* voter à près de 92 % – soit le taux de participation par classe d'âge le plus élevé de notre panel « répondants ».

Comme pour toute enquête déclarative, il est probable qu'une partie des enquêtés sur-déclare leur participation aux élections. Néanmoins, à biais déclaratif constant, les résultats issus de notre étude sont significativement plus élevés que ceux enregistrés par des enquêtes auprès d'échantillons représentatifs de l'électorat.

Tableau 3.3 Taux de participation au premier tour des élections présidentielles 2012 par tranche d'âge (en %)

Classe d'âge	Taux de participation Données INSEE (analyse à partir des listes d'émargement)	Taux de participation des individus qui ont répondu au questionnaire inscrits sur les listes électorales (n = 550)
18 à 24 ans	72,4	91,8
25 à 34 ans	76,1	87,9
35 à 49 ans	85,3	89,5
50 à 64 ans	88,1	85,7
65 ans et plus	77,5	86,7
Total	81,4	89,3

Source : INSEE

Il aurait été bien sûr intéressant de comparer le vote des répondants avec celui des non répondants. On peut en effet légitimement supposer que la passation du questionnaire, comme lors de toute enquête électorale, entraîne un effet de sur-sélection des individus qui participent le plus. Néanmoins,

⁵³⁵ Dominique Cardon, *La démocratie internet. Promesses et limites*, Paris, Seuil, 2010, p. 80.

le trop faible nombre de « non répondants » déclarant d’eux-mêmes leur vote sur leur profil ou par leurs tweets empêche toute mesure statistique rigoureuse.

Comme l’avait observé Pierre Bourdieu, l’échantillon spontané – ici les individus qui ont publié des tweets politiques – est bien « *la limite vers laquelle tend la population des répondants [ou des votants] d’une enquête [ou d’une consultation électorale] auprès d’un échantillon représentatif [ou de la population totale]* »⁵³⁶. Autrement dit, les jeunes qui ont publié des tweets politiques durant la séquence présidentielle sont également ceux qui ont voté et probablement répondu aux enquêtes électorales sur internet⁵³⁷. Les 18-24 ans qui composent notre panel ne sont donc pas représentatifs de l’ensemble des Français âgés de 18 à 24 ans. C’est en réalité seulement la fraction la plus politisée de cette classe d’âge qui publie des messages politiques sur Twitter.

Comme le laisse suggérer leur haut niveau de participation électorale, les individus les plus jeunes de notre panel sont issus pour la plupart de milieux sociaux relativement aisés. Comme l’avait souligné Pierre Bourdieu, la « jeunesse n’est qu’un mot » : ce concept subsume en réalité des univers sociaux qui n’ont pratiquement rien de commun⁵³⁸. Ainsi, au sein d’une même classe d’âge – les 15-19 ans ou les 20-24 ans, pour reprendre les catégories de l’INSEE – tout oppose les individus qui sont toujours dans le système scolaire et universitaire et ceux qui en sont sortis pour entrer dans la vie active, en occupant – ou non – un emploi. Or cette seconde catégorie de « jeunes » est quasiment absente de notre panel. Les jeunes de 15 à 19 ans sortis du système scolaire ne publient pas de messages politiques sur Twitter.

Parmi les individus âgés de 20 à 24 ans qui ont répondu à notre questionnaire, 78,6 % sont étudiants, alors que les étudiants ne représentent que 21,7 % de cette classe d’âge⁵³⁹. Cet écart de près de 57 points témoigne avec force du caractère tout à fait atypique des « jeunes » qui produisent des tweets politiques. Si durant la dernière séquence présidentielle près d’un producteur de tweet politique sur deux avait moins de 25 ans – soit approximativement 50 000 individus – l’importance de ce nombre *absolu* ne doit pas être perçue comme le signe d’une plus grande représentativité statistique. Malgré leur nombre, les individus de moins de 25 ans qui

⁵³⁶ Pierre Bourdieu, *La Distinction, op. cit.*, p. 482.

⁵³⁷ On sait en effet que les jeunes sont plus difficilement joignables par téléphone.

⁵³⁸ Pierre Bourdieu, « La jeunesse n’est qu’un mot », *Questions de sociologie*, Paris, Éditions de Minuit, 1984, p. 143-154.

⁵³⁹ Enquête de la DARES, « Emploi et chômage des 15-29 ans en 2012 », novembre 2013.

publient des tweets politiques ne sont pas représentatifs de leur classe d'âge. Au sein de notre panel, la catégorie des 18-24 ans recoupe en réalité davantage celle des étudiants.

Tableau 3.4 Situation de l'emploi des jeunes de 15 à 24 ans (en %)

	15-19 ans		20-24 ans	
	Répondants (n = 113)	INSEE	Répondants (n = 182)	INSEE
Taux d'activité	1,8	14,4	21,0	60,1
Taux de chômage	0,0	32,7	15,4	21,8
Part de chômage ⁵⁴⁰	-	4,7	3,3	13,1

Source : INSEE

Mais les étudiants de notre panel ne sont pas pour autant représentatifs de l'ensemble des étudiants français. La répartition des étudiants par type d'établissement fréquenté en est une preuve. Alors que les étudiants engagés dans des formations universitaires courtes (type BTS, STS, etc.) représentaient 16 % des effectifs de l'enseignement supérieur en 2011-2012⁵⁴¹, seuls 6 % des étudiants de notre panel étaient inscrits dans ces filières. De manière encore plus remarquable, au sein de notre panel, le taux d'étudiants inscrits dans les grandes écoles et leurs classes préparatoires était près de 20 points supérieur à la moyenne nationale. Au sein de notre échantillon, près d'un étudiant sur trois est effet inscrit dans un établissement de type école de commerce (5,9 %), école d'ingénieur (4,5 %), Institut d'Études Politiques (3,6 %) ou encore école de journalisme (2,7 %).

Dans la majorité de ces formations, les étudiants sont amenés et formés à s'exprimer et disserter sur des questions d'« intérêt général »⁵⁴². La production et la publication d'opinions

⁵⁴⁰ Pour reprendre la définition proposée par l'INSEE, « la part des chômeurs est la proportion de chômeurs dans l'ensemble de la population. Cet indicateur est plus faible que le taux de chômage qui mesure la proportion de chômeurs dans la seule population active. Il est utilisé pour nuancer le très fort taux de chômage parmi les jeunes de moins de 25 ans. Comme beaucoup de jeunes sont scolarisés et que relativement peu ont un emploi, leur taux de chômage est très élevé alors que la proportion de chômeurs dans la classe d'âge est beaucoup plus faible ». Source : <http://www.insee.fr/fr/methodes/default.asp?page=definitions/part-chomeurs.htm>

⁵⁴¹ Atlas régional, « Effectifs d'étudiants en 2011-2012 » : <http://www.enseignementsup-recherche.gouv.fr/cid67140/atlas-regional-effectifs-d-etudiants-en-2011-2012.html>

⁵⁴² Claire Oger, *Le façonnage des élites de la République. Culture générale et haute fonction publique*, Paris, Presses de Sciences Po, 2008.

politiques ou l'échange et la communication d'articles de presse⁵⁴³ s'inscrivent donc dans une économie de pratiques socialement encouragées et valorisées par ces formations.

Tableau 3.5 Répartition (en %) des étudiants en fonction du type d'établissement en 2011-2012

Type d'établissement fréquenté	Répondants (n = 222)	INSEE
Supérieur court (STS, IUT, etc.)	6,0	15,8
Université	62,2	71,4
Grandes écoles et classes préparatoires	31,8	12,9
Total	100,0	100,0

Source : INSEE

Mais cela serait méconnaître les logiques sociales de sélection des grandes écoles⁵⁴⁴ que de se limiter au constat selon lequel la formation dispensée favorise et encourage la prise de parole politique. Dit autrement, les grandes écoles « consacrent » plus que ne provoquent le haut niveau de compétence politique de leurs étudiants. En réalité, la forte proportion d'étudiants issus des grandes écoles est avant tout révélatrice d'une origine sociale élevée et d'un haut niveau de capital culturel. Les chances d'entrer dans une grande école sont ainsi significativement plus élevées pour les fils et filles de cadres et de professeurs que pour les fils et filles d'ouvriers ou d'employés.

Mesurée par la PCS du père, l'origine sociale des étudiants de notre panel confirme cette hypothèse. Près de 42 % des étudiants qui ont répondu au questionnaire sont des filles et fils de cadres et de professions intellectuelles supérieures. Au niveau national, ce taux était de 30,3 % en 2011-2012⁵⁴⁵.

⁵⁴³ Nous le verrons par la suite, une grande partie des tweets politiques sont des liens vers des articles de presse politique. Cf. chapitres 4 et 5.

⁵⁴⁴ Valérie Albouy et Thomas Wanecq, « Les inégalités sociales d'accès aux grandes écoles », *Économie et statistique*, n° 361, 2003 ; Pierre Bourdieu, *La noblesse d'État. Grandes écoles et esprit de corps*, Paris, Éditions de Minuit, 1979.

⁵⁴⁵ Repères et références statistiques sur les enseignements, la formation et la recherche. Statistiques - publications annuelles - Édition 2012 http://media.education.gouv.fr/file/2012/36/9/DEPP-RERS-2012_223369.pdf

Tableau 3.6 Répartition (en %) des étudiants selon leur origine sociale

PCS du père des étudiants	Répondants (n = 222)	Étude RERS 2011-2012
Agriculteurs, artisans, commerçants et chefs d'entreprise	14,4	10,3
Cadres et professions intellectuelles supérieures	41,9	30,3
Professions intermédiaires	11,7	12,5
Employés	12,2	11,2
Ouvriers	12,6	11,3
Autres personnes sans activité professionnelle	7,2	-
<i>Retraités et inactifs</i>	-	11,3
<i>Non-renseigné</i>	-	12,0
Total	100,0	100,0

Source : Ministère de l'Éducation Nationale

L'étude intitulée « Repères et références statistiques sur les enseignements, la formation et la recherche (RERS) » publiée chaque année par le Ministère de l'Éducation Nationale fournit des informations détaillées sur l'origine sociale des étudiants français. Cependant, une différence de catégorisation empêche de comparer nos résultats terme à terme. Comme pour notre recherche, l'origine sociale des étudiants est mesurée à partir de la profession du chef de famille. Cependant, contrairement à la RERS, nous n'avons pas proposé comme item la PCS « Retraités et inactifs ». Nous estimions que cette catégorie ne permettait pas de déterminer l'origine sociale des étudiants, elle n'a donc pas été soumise aux enquêtés (la question de la profession du père était posée en « fermée », seule la question de la profession de l'enquêté était posée en « ouverte »). En 2011-2012 (soit l'année universitaire correspondant à l'administration de notre questionnaire), 11,6 % des étudiants avaient répondu avoir un père retraité ou inactif. On ne sait donc pas quelle était la PCS de ces retraités lorsqu'ils étaient actifs. De plus, contrairement à nos enquêtés, les enquêtés de la RERS avaient la possibilité de ne pas répondre à cette question. Nous ne savons donc pas comment se distribue la catégorie « non-renseigné » parmi les différentes PCS. Notons enfin qu'en raison de leurs effectifs trop peu nombreux, l'étude de la RERS a regroupé en une catégorie les agriculteurs d'une part, et les artisans, commerçants et chefs d'entreprise d'autre part.

Ces résultats confirment l'origine sociale relativement élevée des étudiants – et plus largement des « jeunes » – qui ont publié des tweets politiques durant la dernière séquence présidentielle. Issus majoritairement des classes moyennes supérieures, les étudiants de notre panel témoignent d'un haut niveau de capital culturel et de compétence politique. En cela, les « jeunes » qui publient des tweets politiques présentent un profil proche de ceux qui participent aux enquêtes électorales et aux consultations politiques. Leur surreprésentation au

sein de notre panel (47,9 % des répondants ont, rappelons-le, moins de 25 ans) ne doit donc pas laisser sous-entendre qu'ils soient représentatifs de la « jeunesse » dans son ensemble. Les « jeunes » de notre panel ne sont représentatifs que de la fraction la plus politisée de la « jeunesse ».

Les individus les plus jeunes de notre échantillon présentent en réalité des particularités sociologiques très proches de celles des membres du panel dans son ensemble. Ainsi, l'origine sociale des lycéens et des étudiants est très similaire à celle des autres membres du panel. Ils sont majoritairement issus des classes supérieures et des classes moyennes supérieures.

Tableau 3.7 Répartition (en %) des individus panélisés selon leur origine sociale

PCS père	Lycéen et étudiants (n = 280)	Non lycéen et étudiants (n = 378)
Agriculteurs exploitants	0,0	1,3
Artisans, commerçants et chefs d'entrepr.	14,3	16,4
Cadres et professions intellectuelles sup.	42,9	36,8
Professions intermédiaires	12,9	13,0
Employés	10,4	10,3
Ouvriers	11,8	16,1
Autres personnes sans activité pro.	7,9	6,1
Total	100,0	100,0

Nous allons maintenant étudier les propriétés sociologiques réellement structurantes de notre échantillon et démontrer que ces propriétés sont en fait très proches de celles des membres du « pays légal ». Parmi ces propriétés, la position sociale, saisie à travers la profession et la catégorie socio-professionnelle, joue un rôle déterminant.

1.1.2 Une position sociale élevée

Les producteurs de tweets politiques occupent des positions relativement élevées dans l'espace social. Les cadres et professions intellectuelles supérieures sont ainsi très nettement surreprésentés au sein de notre échantillon. En effet, alors que les cadres ne représentent que 14,8 % des Français en âge de travailler et n'étant plus en cours d'étude, 49,2 % des individus qui ont répondu à notre questionnaire et 71,1 % de ceux qui n'ont pas répondu (et dont la position sociale a pu être identifiée) appartiennent à cette catégorie socio-professionnelle. À l'inverse, on enregistre une sous-représentation des ouvriers et des employés. Seuls 2,4 % des répondants et 1,0 % des non répondants (dont la profession a pu être identifiée) sont des ouvriers, alors que cette catégorie socio-professionnelle représente 19,1 % de la population française en âge de travailler et n'étant plus en cours d'études.

Tableau 3.8 Répartition (en %) des panélisés selon la PCS

PCS (hors lycéens et étudiants)	Répondants (n = 378)	Non Rép. (n = 201)**	Twitter IFOP (n = 306)	INSEE 2012	INSEE 2012 Etudiants et retraités exclus
Agriculteurs exploitants	0,3*	0,0*	0,0*	1,0	1,5
Artisans, commerçants et chefs d'entreprise	2,4*	4,0*	9,5*	3,4	5,2
Cadres et professions intellectuelles sup.	49,2	71,1	67,6	9,6	14,8
<i>Professions libérales et assimilés</i>	4,0*	2,0*	-	1,0	1,5
<i>Cadres de la fonction publique, professions intell.</i>	21,7	33,8	-	3,2	4,9
<i>Cadres d'entreprise</i>	23,5	35,3	-	5,4	8,3
Professions intermédiaires	25,7	16,9	14,9	13,3	20,5
<i>Professions intermédiaires du public</i>	12,4	3,0	-	5,7	8,8
<i>Professions intermédiaires du privé</i>	8,2	9,5	-	4,1	6,4
<i>Techniciens, contremaîtres et agents de maîtrise</i>	5,1*	4,5*	-	3,5	5,3
Employés	10,6	3,5*	5,4*	16,0	24,7
Ouvriers	2,4*	1,0*	1,4*	12,4	19,1
Retraités	5,8	1,0*	1,4*	26,5	-
Autres personnes sans activité pro.	3,7*	2,5*	-	17,7	14,0***
Total	100	100,0	100,0	100,0	100,0

Source : INSEE, 2012 ; IFOP, 2010

*Effectifs faibles **Individus dont la profession a pu être renseignée***Uniquement des chômeurs et des femmes/hommes au foyer.

Après avoir placé notre focale sur la population étudiante, nous avons fait le choix de recentrer ici l'analyse sur la population qui n'est plus en cours d'étude. Nous avons alors retiré les lycéens et les étudiants de notre population afin de recentrer l'analyse sur la population susceptible de travailler (comprenant la population active et les inactifs autres qu'étudiants). Nous avons comparé la distribution par PCS des populations répondants et non répondants avec les données de l'INSEE. Afin de corriger – assez sommairement – les effets spécifiques liés à la structure par âge de notre échantillon (les catégories les plus âgées de la population sont nettement sous-représentées sur Twitter, diminuant ainsi mécaniquement le poids de la catégorie « Retraités », à l'inverse les étudiants sont surreprésentés et donnent un poids prépondérant à la catégorie « autres personnes sans activité professionnelle »), nous avons exclu les retraités et les étudiants de la population INSEE (ne restent alors que les Français en âge de travailler et n'étant plus en cours d'études).

Concernant le panel non répondants, la PCS a pu être identifiée dans 32,4 % des cas. Il convient néanmoins de préciser que certaines PCS sont plus facilement identifiables que d'autres. La PCS des non répondants dont la profession a pu être identifiée ne se distribue pas de la même manière que la PCS des non répondants dont la profession n'a pas pu être identifiée. (Auto-) indiquer sa profession sur internet est en réalité une activité socialement située, qui ne se distribue pas de la même manière selon les différentes catégories socio-professionnelles. Les cadres ont ainsi plus tendance que la moyenne à indiquer leur profession sur leur profil Twitter ou sur d'autres réseaux sociaux ou blogs et sites personnels. Leur présence sur des réseaux sociaux dits « professionnels » tels que LinkedIn ou Viadeo (davantage destinés aux emplois qualifiés), sur lesquels les comptes d'utilisateurs s'apparentent à des curriculum vitae en ligne, facilite ainsi l'identification de leur profession. En revanche, les professions intermédiaires, les ouvriers et les employés indiquent très rarement leur profession sur internet. À titre d'exemple, 74 % des cadres du panel « répondants » avaient rempli leur présentation de profil Twitter contre seulement 33 % des ouvriers. Plusieurs hypothèses peuvent être avancées pour expliquer ces écarts. La plus probable étant que la recherche « active » d'un emploi sur internet, qui passe par la création et l'entretien de réseaux professionnels, n'est socialement encouragée (et sans doute socialement utile) que pour la recherche d'emplois qualifiés (d'encadrement et de direction), et non pour des emplois peu qualifiés ou intermédiaires.

Enfin précisons que les résultats concernant les inscrits sur le réseau social (et non uniquement les inscrits qui publient des tweets politiques) sont tirés d'une étude précédemment citée, réalisée par l'IFOP en 2010. Rappelons que les résultats de cette étude sont à interpréter avec beaucoup de prudence. D'une part car, nous l'avons déjà dit, les modalités d'échantillonnage, de « recrutement » et de passation du questionnaire restent relativement floues ; de plus la taille relativement réduite de l'échantillon interrogé (n = 306) implique d'importantes marges d'erreur. D'autre part, la date de réalisation de l'étude (2010, soit deux ans avant notre enquête) ne doit pas être négligée. La démographie du réseau social évolue sensiblement et de façon continue, presque jour après jour, et

tout laisse penser que cette croissance démographique impacte et modifie la sociologie du réseau social. Dit autrement, la distribution des inscrits par PCS n'est sans doute plus la même aujourd'hui qu'elle ne l'était en 2010 (ni en 2012, au moment de notre enquête, nous y reviendrons). On peut donc raisonnablement supposer que les catégories les plus sous-représentées statistiquement en 2010 l'étaient légèrement moins en 2012.

Cependant, à la lecture de ces résultats, il convient de se demander, pour être tout à fait précis, si la sous-représentation statistique des catégories sociales inférieures (ou, corrélativement, la surreprésentation des catégories supérieures) est imputable uniquement à des inégalités de compétence politique ou s'explique davantage par des effets de sélection propres au réseau social et à la sociologie de sa population⁵⁴⁶. Autrement dit, on peut se demander si la sous-représentation des ouvriers et des employés à l'intérieur de notre panel n'est pas avant tout la conséquence d'une plus faible présence de ces catégories sociales sur Twitter.

D'après l'IFOP⁵⁴⁷, en 2010, le réseau social ne comptait que 4 % d'employés et seulement 1 % d'ouvriers. La répartition par PCS des inscrits serait donc relativement proche de celle des producteurs de tweets politiques qui n'ont pas répondu au questionnaire. Une fois inscrit sur Twitter, il n'existerait donc pas de filtre spécifique, ou de « cens caché » à la publication de messages politiques. Le véritable filtre serait donc bien d'être inscrit sur le réseau social, avant même d'y parler de politique.

Étudier les représentations des individus qui ne sont pas inscrits sur Twitter permettrait sans doute de mieux saisir les logiques sociales d'inscription au réseau social et de comprendre ainsi la plus faible inscription des ouvriers et employés. Or nous ne disposons pas de données empiriques relatives aux représentations des non-inscrits ; et si nous avons réalisé des entretiens avec certains membres de notre panel, rien n'indique que leurs représentations soient les mêmes que celles des non-inscrits. Pour tenter d'expliquer cette sous-représentation, on peut alors émettre l'hypothèse (non étayée empiriquement) selon laquelle Twitter serait moins associé à un espace de sociabilité interpersonnelle (comme l'est Facebook, qui bénéficie d'un taux de pénétration élevé parmi les catégories sociales

⁵⁴⁶ Rappelons que tous les inscrits sur Twitter ne sont pas des producteurs de tweets politiques.

⁵⁴⁷ IFOP, « Twitter vu par ses utilisateurs », 2010.

*inférieures*⁵⁴⁸) qu'à un espace de discussion centré sur l'actualité – et notamment sur l'actualité politique –, et que, partant, cette représentation « politisée » de Twitter pourrait jouer comme un « frein » pour les individus les moins politisés, ou du moins les individus les moins habilités à s'exprimer politiquement. C'est alors cette représentation davantage « politisée » de Twitter qui pourrait expliquer l'écart d'inscription entre les catégories inférieures, intermédiaires et supérieures au réseau social.

Les ouvriers, employés et professions intermédiaires inscrits sur Twitter sont issus de la fraction la plus politisée – et donc certainement la moins représentative – des classes sociales inférieures. Ces agents se sont d'ailleurs particulièrement prêtés au jeu de l'enquête : les employés, ouvriers et professions intermédiaires ont proportionnellement plus répondu à notre questionnaire – qui peut être appréhendé comme une enquête politique – que les cadres et professions intellectuelles supérieures.

La majorité des cadres producteurs de tweets politiques évoluent quant à eux dans des « professions de présentation et de représentation »⁵⁴⁹, des professions de production et de vente de « biens et services symboliques » : « attachés de presse », « journalistes », « chargés de communication », « chargés d'études », « chefs de projet », « cadres commerciaux », « responsables marketing », etc. Pierre Bourdieu avait qualifié de « *petite bourgeoisie nouvelle* » l'ensemble des agents sociaux, aux trajectoires sociales souvent très diverses, qui occupaient ces professions « nouvelles ou rénovées », associées à des positions sociales encore mal déterminées. Pour le sociologue, ces professions nouvelles étaient « *le lieu d'élection de tous ceux qui n'avaient pas obtenu du système scolaire les titres leur permettant de revendiquer avec succès les positions établies auxquelles les promettait leur position d'origine* »⁵⁵⁰. On y trouvait donc à la fois des agents déclassés issus des classes supérieures et des agents en ascension sociale issus des classes inférieures. Pierre Bourdieu intuitionnait néanmoins que ces agents en ascension avaient « *toutes les chances d'être expulsés des positions où ils se sont trouvés conduits [les professions nouvelles des années 1960] par un effet d'allodoxia suscité par le système scolaire, à mesure qu'elles viendront à se revaloriser (entre autres, du fait de leur élimination)* »⁵⁵¹. Trente ans après *La Distinction*, beaucoup de

⁵⁴⁸ Précisons que nous n'accordons bien sûr aucune portée normative au terme « inférieure », qui s'oppose ici à « supérieure » en référence à la place objective occupée par une PCS sur l'échelle sociale.

⁵⁴⁹ Pierre Bourdieu, *La Distinction*, op. cit., p. 420

⁵⁵⁰ *Ibid.*, p. 409.

⁵⁵¹ *Ibid.*, p. 421.

ces professions se sont en effet « revalorisées » et ne sont aujourd’hui plus « nouvelles » : les métiers de « chargé d’études », d’« attaché de presse » ou de « responsable marketing » se sont institutionnalisés, comme en témoignent notamment la création de formations spécifiques sanctionnées par des titres scolaires.

Lise Bernard⁵⁵², à partir des enquêtes Emploi de l’INSEE réalisées entre 2003 et 2007 a cherché à réaliser une cartographique de l’espace social des années 2000⁵⁵³ (en mettant en évidence les groupes socio-professionnels dont les membres sont issus de milieux sociaux proches, en étudiant l’origine sociale des individus et leurs alliances matrimoniales). L’analyse factorielle des correspondances fait ressortir trois principaux pôles : un premier composé de classes supérieures (professions libérales, médecins salariés, etc.), un deuxième formé d’ouvriers et de certaines catégories d’employés, et un troisième pôle regroupant différentes catégories de professions intermédiaires et de travailleurs indépendants. Lise Bernard note qu’un des trois sous-ensembles qui composent ce dernier pôle intermédiaire « présente une morphologie professionnelle qui rappelle, dans une certaine mesure, celle de la “petite bourgeoisie nouvelle” » et observe que ces agents ont plus souvent que la moyenne un père cadre. Ce résultat semble donc confirmer que certaines des professions « nouvelles » se sont revalorisées : les agents exerçant ces professions se recrutent davantage au sein des catégories sociales supérieures. Les cadres de la publicité apparaissent même désormais rattachés au pôle des classes supérieures, à côté des professions libérales.

Jean-Baptiste Comby, dans une perspective de recherche qui nous intéresse ici plus directement, a quant à lui étudié les rapports différenciés des catégories sociales supérieures à l’information politique en ligne⁵⁵⁴. L’auteur dissocie au sein des catégories supérieures deux principaux groupes, définis par deux rapports spécifiques à l’information politique en ligne. Un premier groupe, dont les membres sont qualifiés d’« usagers conventionnels », ne ferait pas d’internet une source principale d’information. Selon l’auteur, lorsqu’ils s’informent *via* internet, ces usagers reproduiraient en ligne leurs pratiques hors-ligne (consultant ainsi principalement les sites des trois grands quotidiens nationaux : lemonde.fr, liberation.fr et

⁵⁵² Lise Bernard, « Réflexions sur la “petite bourgeoisie nouvelle” dans les années 2000 », in Philippe Coulangeon et Julien Duval (dir.), *Trente ans après "La Distinction" de Pierre Bourdieu*, La Découverte, « Recherches », Paris, 2013, p. 266-277.

⁵⁵³ Beaucoup de ceux qui trouvent « datées » les analyses de *La Distinction*, ou qui pointent des changements dans l’analyse que fait Bourdieu de certaines positions sociales au fil de sa carrière peinent à considérer que l’espace social est un espace par essence mouvant (le chapitre 2 s’intitule d’ailleurs « l’espace social et ses transformations » et que l’analyse diachronique des trajectoires sociales collectives sont tout aussi voire plus importantes que l’analyse synchronique des positions sociales à un temps *t*).

⁵⁵⁴ Jean-Baptiste Comby, « Les appropriations différenciées de l’information en ligne au sein des catégories sociales supérieures », *Réseaux*, 2011/6, n°170, p. 75-102.

lefigaro.fr). Un deuxième groupe, celui des « usagers en tâche de fond », ferait au contraire d'internet une source principale voire exclusive d'information. Ces individus consulteraient davantage des sites dits « pure player »⁵⁵⁵ et « navigueraient en continu sur les réseaux sociaux et notamment Twitter ». On peut donc logiquement supposer que ces « usagers en tâche de fond » décrits par Comby présentent un profil très similaire à celui des producteurs de tweets politiques. Ces individus partagent en effet les mêmes propriétés sociales que les individus de notre panel. Jean-Baptiste Comby note en effet que les usagers en « tâche de fond » exercent « *des métiers en lien avec les médias et la communication* », des professions « *de la présentation et de la représentation* », ou des « *professions du symbolique* » : en clair les professions occupées dans les années 1960 par la « petite bourgeoisie nouvelle ». Cependant, contrairement à ce qu'observe Lise Bernard et à ce qu'avait anticipé Pierre Bourdieu, pour Jean-Baptiste Comby la majorité de ces agents se recrute encore parmi des fractions très hétérogènes des classes moyennes et inférieures. Ces cadres connaîtraient donc une trajectoire sociale ascendante qui expliquerait leur rapport distinctif à l'information en ligne⁵⁵⁶.

Au sein de notre échantillon, la majorité des producteurs de tweets politiques qui exercent ces « professions nouvelles » ont un père cadre et sont issus des classes moyennes supérieures (la petite bourgeoisie nouvelle des années 1960) ou des classes supérieures. Contrairement à Jean-Baptiste Comby, nous n'affirmons donc pas que les agents sociaux qui occupent aujourd'hui ces positions connaissent une trajectoire sociale ascendante et que leur trajectoire sociale explique leurs pratiques sociales numériques. En revanche, on observe parmi ces professions du symbolique de « nouvelles » professions qui doivent être selon nous distinguées des « professions nouvelles » des années 1960. Ces nouvelles professions des années 2000, relativement proches de celles apparues dans les années 1960, sont nées avec l'émergence d'internet et sont orientées vers la production symbolique de nouveaux biens et services « numériques » : « webmaster », « social community manager », « chargé de web marketing » ou de « marketing digital », etc.⁵⁵⁷ Tout comme les professions nouvelles des années 1960 décrites par Bourdieu, ces professions semblent associées à des positions qui « *n'offrent aucune garantie mais ne demandent en retour aucune garantie, n'exigent aucun*

⁵⁵⁵ Cf. glossaire, p. 464.

⁵⁵⁶ Nous développerons et discuterons dans le chapitre 5 cette idée.

⁵⁵⁷ Il s'agit en fait des professions associées à l'émergence du nouveau secteur d'activité professionnel autour des autopublications que nous avons présenté dans le chapitre 1.

droit d'entrée »⁵⁵⁸ et qui laissent aux agents la possibilité de créer eux-mêmes l'avenir de leur profession en réalisant des « *coups de force symboliques nécessaires pour produire et imposer de nouveaux produits* »⁵⁵⁹. Si des effectifs trop faibles nous empêchent de mener une analyse statistique de la trajectoire sociale (individuelle mais surtout collective) des agents occupant ces nouvelles positions, on peut néanmoins émettre l'hypothèse selon laquelle ces agents sont issus, comme l'était la petite bourgeoisie nouvelle dans les années 1960, d'horizons sociaux très divers.

Notons enfin qu'à côté de ces métiers de la présentation et de la représentation (numérique ou non), on retrouve parmi les classes sociales supérieures une part importante de professeurs du secondaire et du supérieur et, dans une moindre mesure, de cadres administratifs d'entreprise. Plus d'un producteur de tweets politiques sur deux (voire trois sur quatre, parmi les non répondants) occupe donc une position sociale relativement élevée.

Nous allons maintenant voir que la position sociale des producteurs de tweets politiques donne leur forme spécifique à d'autres variables secondaires, tels que le sexe, le lieu d'habitation ou encore le niveau de diplôme.

1.1.3 Des producteurs masculins, Parisiens, et fortement diplômés

À l'instar des populations qui répondent aux enquêtes électorales ou qui participent politiquement, les producteurs de tweets politiques sont davantage masculins, Parisiens et diplômés que la moyenne des Français. La position sociale relativement élevée – saisie à travers la profession – des membres de notre panel explique pour partie l'état de ces propriétés secondaires. La part relative des femmes au sein d'une catégorie professionnelle, la distribution sociale des agents issus de cette catégorie dans l'espace géographique ou encore le niveau de diplôme nécessaire à l'entrée dans un champ professionnel sont en effet des variables en grande partie définies par la position des agents dans l'espace social.

⁵⁵⁸ Pierre Bourdieu, *La Distinction*, op. cit.

⁵⁵⁹ *Ibidem*.

1.1.3.1 Sexe et production de tweets politiques

Les producteurs de tweets politiques sont ainsi très majoritairement des hommes. Alors que la population internautes française se compose à part égale d'hommes et de femmes⁵⁶⁰, les femmes sont nettement sous-représentées au sein de notre échantillon. 59,4% des producteurs de tweets politiques qui ont répondu à notre questionnaire et 66,4% de ceux qui n'ont pas répondu sont des hommes.

Tableau 3.9 Distribution de la population selon le sexe (en %)

Sexe	Répondants (n = 658)	Non répondants (n = 590)	Pop. Twitter IFOP 2012 (n = 306)	Facebook IFOP 2012 (n = 1083)	Pop. internet IFOP 2012 (n = 2005)	Pop. française INSEE 2012
Hommes	59,4	66,4	59,7	46,9	49,7	48,5
Femmes	40,6	33,6	40,3	53,1	50,3	51,6
Total	100,0	100,0	100,0	100,0	100,0	100,0

Sources : IFOP, INSEE

Cependant, affirmer, comme nous l'avons fait plus haut, que la surreprésentation masculine constitue une propriété sociale distinctive des populations qui participent ordinairement aux consultations politiques peut apparaître aujourd'hui comme un fait sociologique discutable, voire daté.

En effet, les écarts de participation politique entre hommes et femmes se sont considérablement réduits au cours des trois dernières décennies⁵⁶¹. Au premier tour des élections présidentielles du 22 avril 2012, le taux de participation était quasiment identique chez les femmes (80,9 %) et chez les hommes (81,9 %)⁵⁶². Parmi les 25-34 ans, le taux de participation des femmes était même de six points supérieur à celui des hommes (79,1 % contre 73,1 %), créant ainsi un nouveau *gender gap*, cette fois-ci à la faveur des femmes. L'accroissement de la part des femmes accédant à l'enseignement secondaire et supérieur a favorisé une élévation de leur niveau de compétence politique, à la fois technique, et, dans

⁵⁶⁰ Sur Facebook, réseau social bénéficiant du meilleur taux de pénétration au sein de la population internautes, les femmes sont en revanche majoritaires : 53,1 % des inscrits seraient des femmes (source : IFOP, « Observatoire des réseaux sociaux », vague 2012).

⁵⁶¹ Pippa Norris « The Gender Generation Gap in British Elections », *British Elections and Parties Yearbook*, Vol. 3, Issue 1, 1993, p. 129-142.

⁵⁶² Données INSEE, à partir des listes d'émargement.

une moindre mesure, statutaire. L'augmentation du capital scolaire des femmes, corrélée à une élévation de leur niveau de compétence politique, permet alors d'expliquer ce rééquilibrage en termes de participation électorale⁵⁶³. Suivant cette logique, on constate que, parmi les générations les plus âgées, les femmes continuent de participer toujours moins que les hommes⁵⁶⁴. Parmi les 65 ans et plus, les hommes ont en effet davantage voté en 2012 (84,0 %) que les femmes (72,9 %).

La distribution par sexe et par âge des répondants laisse entrevoir le même phénomène au sein de notre échantillon. Parmi les répondants, la part des femmes décroît également en fonction de l'âge. Chez les moins de 18 ans, les femmes sont deux fois plus nombreuses que les hommes à avoir publié des tweets politiques. À partir de 35 ans, au contraire, les hommes sont en moyenne deux fois plus nombreux à parler de politique sur Twitter.

Tableau 3.10 Distribution de la population « répondants » selon le sexe et l'âge (en %)

Tranche d'âge/ Sexe	Homme	Femme	Total
Moins de 18 ans (<i>n</i> = 48)	38	62	100
18 à 25 ans (<i>n</i> = 249)	55	45	100
26 à 34 ans (<i>n</i> = 116)	58	42	100
35 à 45 ans (<i>n</i> = 111)	68	32	100
46 à 60 ans (<i>n</i> = 59)	66	34	100
61 ans et plus (<i>n</i> = 25)	64	36	100
Total (<i>n</i> = 608)	58	42	100

Cependant, et bien que la distribution hommes/femmes se rééquilibre parmi les plus jeunes, les hommes sont toujours majoritaires, quelles que soient les classes d'âge étudiées (à l'exception notable des moins de 18 ans).

La position sociale des enquêtés peut permettre alors de mieux comprendre cette surreprésentation masculine. On sait en effet que chaque catégorie socio-professionnelle est définie par un sex-ratio qui lui est propre. En 2012, au sein de la population française âgée de 15 ans et plus, la catégorie des « employés » était ainsi composée de 76,6 % de femmes alors que ce taux, chez les cadres, n'était que de 40,2 %. La surreprésentation des cadres au sein de

⁵⁶³ Inversement, parmi les 50 ans et plus, les hommes ont davantage voté (72,7%) que les femmes (67,45%).

⁵⁶⁴ Il convient toutefois de préciser que les femmes connaissent une plus forte espérance de vie que les hommes et sont ainsi en moyenne plus âgées dans cette catégorie, or le « grand âge » (plus de 75 ans) produit de l'abstention.

notre échantillon (cette catégorie représente, nous l'avons vu, 49,2 % des répondants, si l'on exclut les étudiants) donne donc un élément d'explication à la sous-représentation statistique des femmes au sein de notre panel.

*Tableau 3.11 Distribution de la population « répondants » selon le sexe et la PCS
(en %)*

PCS / Sexe (en %)	Part des Femmes (Répondants) <i>n</i> = 658	Part des Femmes Population de 15 ans et + INSEE 2012
Agriculteurs exploitants	0,0*	28,3
Artisans, commerçants et chefs d'entreprise	0,0*	27,6
Cadres et professions intellectuelles sup.	32,9	40,2
Professions intermédiaires	38,2	51,2
Employés	41,7	76,6
Ouvriers	0,0*	19,6
Retraités	36,4	54,2
Etudiants	51,4	52,0
Total général	41,9	52,2

*effectifs faibles ou nuls

Source : INSEE 2012

On remarque alors que plus une PCS est représentée statistiquement sur Twitter, plus son sex-ratio se rapproche de celui de la PCS dans son ensemble. Ainsi, la part des femmes « étudiantes »⁵⁶⁵ est quasiment identique sur Twitter et dans la population étudiante totale. De la même manière, le sex-ratio des cadres sur Twitter est également relativement proche de celui des cadres dans leur ensemble. Et inversement, moins une PCS est représentée statistiquement sur Twitter, plus on enregistre un écart important en terme de sex-ratio⁵⁶⁶.

On s'aperçoit donc bien que la structure sociale de notre échantillon détermine la surreprésentation statistique des hommes au sein de notre panel. De la même manière, nous allons voir que la position sociale relativement élevée des producteurs de tweets politiques permet de comprendre la surreprésentation des Parisiens au sein de notre échantillon.

⁵⁶⁵ Qui composent presque exclusivement la catégorie des autres personnes sans activité professionnelle.

⁵⁶⁶ Cette relation statistique peut alors possiblement s'expliquer par le fait les catégories les moins bien représentées statistiquement sur Twitter sont aussi celles qui comportent les marges d'erreurs les plus importantes.

1.1.3.1 Lieu d'habitation et production de tweets politiques

Le lieu d'habitation des producteurs de tweets politiques est une variable (secondaire) particulièrement structurante de notre échantillon. La surreprésentation des habitants de Paris et de son agglomération⁵⁶⁷ est en effet sensible. Alors que les Parisiens ne représentent que 3,4 % de la population française, 12,5 % des producteurs de tweets politiques qui ont répondu à notre questionnaire et 39,6 % de ceux qui n'ont pas répondu déclarent – ou « auto-déclarent » habiter à Paris *intra-muros*.

Tableau 3.12 Distribution des populations répondants et non répondants selon le lieu de résidence (en %)

Lieu de résidence (en %)	Répondants (n = 658)	Non répondants* (n = 373)	Pop. française INSEE 2012
Paris <i>intra-muros</i>	12,5	39,6	3,4
Agglo. parisienne hors Paris <i>intra-muros</i>	14,9	12,1	12,4
Province	72,6	48,2	84,2
Total	100,0	100,0	100,0

Source : INSEE

* Le lieu d'habitation a pu être identifié dans 60% des cas

On enregistre un écart très important entre les panels répondants et non répondants concernant le lieu d'habitation. Cet écart peut s'expliquer par des raisons méthodologiques. Pour le panel répondants, la variable « lieu d'habitation » a été codée à partir du code postal à cinq chiffres que l'on a demandé aux enquêtés d'indiquer (question ouverte obligatoire). Concernant le panel non répondants, la variable « lieu d'habitation » a été codée à partir des indications géographiques (auto)déclarées par les membres du panel sur leur profil Twitter ou sur d'autres sites et réseaux sociaux. Ces informations auto-déclarées sont cependant plus ou moins précises. Un même individu pourrait ainsi indiquer comme lieu d'habitation « la Bretagne », « Brest », ou une ville de l'agglomération brestoise. Ainsi, si 39,6 % des non répondants (dont le lieu d'habitation a pu être identifié) ont indiqué habiter à Paris, il est fort probable qu'une majorité de ces individus aient une vision très extensive de « Paris » et habitent en réalité dans une commune de la petite ou de la grande

⁵⁶⁷ D'après l'INSEE, l'agglomération parisienne, ou l'aire urbaine de Paris, comprend quatre départements : Paris, les Hauts-de-Seine, la Seine-Saint-Denis et le Val-de-Marne, soit, en 2010, 412 communes et 10 460 118 habitants.

couronne. Au sein du panel répondants, un certain nombre d'individus ont ainsi indiqué sur leur profil habiter à Paris alors qu'ils résidaient en réalité dans une ville de la petite ou de la grande couronne. Cependant, si on additionne les catégories « Paris intra-muros » et « Agglomération parisienne hors Paris », l'écart entre les deux panels reste manifeste. 27,4 % des répondants habitent dans la capitale ou son agglomération contre 51,7 % des non répondants. On peut alors supposer que les habitants de province indiquent moins souvent sur leur profil Twitter leur lieu d'habitation que les Parisiens. Cette hypothèse a pu être vérifiée empiriquement via l'étude du panel répondants (on connaît en effet pour les membres de ce panel à la fois les données auto-déclarées et les données renseignées par questionnaire). 81 % des individus habitant Paris intra-muros indiquent sur leur profil leur lieu d'habitation contre 70 % des individus habitant en province. On peut alors supposer, sans toutefois pouvoir cette fois le vérifier, que cet écart est amplifié parmi les non répondants.

La surreprésentation des Parisiens sur Twitter peut paraître de prime abord surprenante, voire paradoxale. En effet, Internet permet de neutraliser, au moins *virtuellement*, les distances géographiques entre les agents sociaux. À la différence de l'offre culturelle qui est géographiquement concentrée dans quelques grandes métropoles (dont principalement Paris), Internet et Twitter sont accessibles depuis n'importe quel site du territoire géographique⁵⁶⁸. Pourtant, on observe bien sur Twitter une surreprésentation des Parisiens, à savoir des agents résidant dans un périmètre géographique concentrant les principales institutions de pouvoir et de culture.

Et c'est encore une fois la position sociale des producteurs de tweets politiques qui permet de comprendre ce paradoxe. La distribution des agents sociaux dans l'espace géographique n'est en effet jamais neutre socialement⁵⁶⁹. Les agents issus de la petite bourgeoisie nouvelle et des classes intellectuelles supérieures – catégories d'agents, nous l'avons vu, surreprésentées au sein de notre échantillon – résident en effet plus que la moyenne des Français au sein de la capitale. Les entreprises de communication, d'études, de conseil, et plus généralement l'ensemble des « professions nouvelles » décrites par Bourdieu sont en effet essentiellement localisées à Paris. C'est donc la prédominance de ces professions sur Twitter qui expliquerait en grande partie cette surreprésentation parisienne.

⁵⁶⁸ Même si les facilités d'accès aux infrastructures qui facilitent la navigation sur internet – le haut ou très haut débit, la fibre optique, la 4G, etc. – ne sont pas les mêmes partout sur le territoire géographique.

⁵⁶⁹ Pierre Bourdieu, *La Distinction*, *op. cit.*, p. 113.

Tableau 3.13 Distribution (en %) des répondants selon la PCS et le lieu de résidence
(n = 658)

PCS / lieu d'habitation (en %)	Paris <i>intra-muros</i>	Agglo. parisienne hors Paris	Province
Agriculteurs exploitants (n = 1)	-	-	-
Artisans, commerçants et chefs d'entreprise (n = 9)	0,0*	20,0*	80,0*
Cadres et professions intellectuelles sup. (n = 189)	21,5	14,1	64,4
Professions intermédiaires (n = 97)	8,6	11,1	80,2
Employés (n = 40)	5,9	2,9	91,2
Ouvriers (n = 9)	0,0*	0,0*	100,0*
Retraités (n = 22)	5,0	5,0	90,0
Autres personnes sans activité pro. (n = 291)	10,3	19,4	70,3
Total général (n = 658)	12,3	15,2	72,5

*effectifs faibles

Toujours afin de comprendre ce paradoxe, on peut reprendre – en l'élargissant – une autre hypothèse formulée par Pierre Bourdieu selon laquelle « *c'est seulement dans les dîners ou dans les conversations de l'existence bourgeoise que les noms propres d'intérêt général [...] renvoient à des personnes familières, concrètement connues et fréquentées, qui, comme les voisins et les cousins de l'espace villageois, appartiennent à l'univers de l'interconnaissance* »⁵⁷⁰. La surreprésentation des Parisiens au sein de notre échantillon s'expliquerait alors selon une hypothèse – que nous n'avons pas les moyens de démontrer empiriquement – qui s'articulerait comme suit : 1) on enregistre sur Twitter une forte proportion d'agents issus des champs médiatique et politique (dont la plupart, à l'instar des étudiants en écoles de journalisme ou de science politique, n'ont pas – encore – des « *noms propres d'intérêt général* »)⁵⁷¹ 2) les champs politique et médiatique sont géographiquement circonscrits au périmètre de la capitale, 3) les logiques d'inscription au réseau sont en partie tributaires d'un phénomène de « bouche-à-oreille » et se diffusent donc au sein des réseaux d'interconnaissances de ces agents « Parisiens ». Autrement dit, si l'on enregistre parmi notre population une surreprésentation de Parisiens, c'est aussi sans doute parce que Twitter concentre une population d'agents qui appartiennent ou qui « gravitent » autour des champs médiatique et politique, et que ces champs sont concentrés géographiquement au sein de la capitale.

Au final, si internet permet bien de neutraliser les distances *géographiques*, l'analyse fait néanmoins ressortir que les distances *sociales* qui se superposent – ou se dissimulent – derrière ces distances *géographiques* sont en revanche toujours tenaces. Au final, on se rend

⁵⁷⁰ Pierre Bourdieu, *La Distinction*, op. cit., p. 518.

⁵⁷¹ Nous développerons ce point dans les chapitres suivants.

compte que la position dans l'espace géographique des agents sociaux est déterminée par la position de ces agents dans l'espace social.

Davantage masculins et Parisiens, les producteurs de tweets politiques sont également plus diplômés que la moyenne des Français. Cette propriété sociologique singulière rapproche encore une fois les producteurs de tweets politiques des membres du « pays légal ».

1.1.3.1 Niveau d'étude et production de tweets politiques

Les producteurs de tweets politiques sont en effet également plus diplômés que la moyenne des Français.

La sociologie politique a de longue date établi que le niveau de diplôme est fortement corrélé au niveau de compétence politique des individus, à leur niveau d'intérêt pour la politique ainsi qu'à leur propension à voter et à répondre aux questions politiques⁵⁷². Les résultats issus de notre questionnaire permettent d'affirmer que le niveau de diplôme est également fortement corrélé à la probabilité de produire des tweets politiques.

Les individus qui ont accepté de répondre à notre questionnaire sont en effet nettement plus diplômés que la moyenne des Français. Cette surreprésentation ne concerne pas uniquement les individus en cours d'études, comme nous l'avons vu précédemment. Si l'on exclut cette catégorie de notre échantillon, trois quart des producteurs répondants sont diplômés de l'enseignement supérieur (74,9 %) et plus d'un tiers ont même un niveau de diplôme égal ou supérieur à Bac +5 (33,9 %).

⁵⁷² Voir entre autres Daniel Gaxie, *Le cens caché*, *op. cit.* ou Pierre Bourdieu, *La Distinction*, *op. cit.* Pour des données plus récentes, on peut citer ici par exemple un sondage réalisé par l'IFOP à l'occasion de l'élection parlementaire européenne du 25 mai 2014, d'après lequel 33% des individus possédant un niveau de diplôme inférieur au bac déclarent s'intéresser « beaucoup ou assez » à la campagne contre 42% de ceux disposant d'un niveau supérieur à bac+2. Cf. IFOP, « Sondage Jour du Vote : profil des électeurs et clefs du scrutin européen », 25 mai 2014.

*Tableau 3.14 Distribution de la population répondants selon le niveau de diplôme
(en %)*

Dernier diplôme obtenu (hors lycéens et étudiants)	Panel Répondants (n = 378)
Diplôme inférieur au Baccalauréat	11,4
Niveau Baccalauréat	13,8
Deug, BTS ou équivalent Bac+2	14,8
Licence ou équivalent Bac+3	14,6
Maîtrise ou équivalent Bac+4	10,6
Master (ex. DEA ou DESS) ou équivalent Bac+5	31,2
Doctorat	3,7
Total général	100,0

Comparée à la population française dans son ensemble, la population des producteurs de tweets politiques est donc largement surdiplômée. Cependant, comparer le niveau de diplôme des répondants avec celui des Français dans leur ensemble n'a guère de sens sociologiquement. L'état du marché des titres scolaires et la valeur relative de ces titres ne sont en effet pas les mêmes d'une génération à l'autre. Dit autrement, les chances d'être titulaire aujourd'hui d'un diplôme de Master ne sont pas les mêmes pour les plus de 65 ans et les 25-49 ans. Or toutes les générations ne sont pas équitablement représentées au sein de notre échantillon. Nous l'avons vu, les catégories les plus âgées sont ainsi nettement sous-représentées. Leur sous-représentation élève donc presque mécaniquement le niveau général d'étude de notre population.

Afin de contrôler ce biais, nous avons comparé le niveau d'étude d'une classe d'âge (les 25-49 ans) relativement bien représentée statistiquement au sein de notre panel (38,1 % des répondants appartiennent à cette classe d'âge contre 32,5 % des Français). Or à l'intérieur de cette classe d'âge, le niveau de diplôme des répondants est tout de même sensiblement plus élevé que la moyenne. Alors que la majorité des Français âgés de 25 à 49 ans ont un niveau de diplôme inférieur ou égal au Baccalauréat (64,1 % des Français contre 21,4 % des répondants), la majorité des répondants ont au contraire un niveau de diplôme égal ou supérieur à Bac+4 (49,0 % des répondants contre 20,5 % des Français).

Tableau 3.15 Distribution des populations âgées de 25 à 49 ans selon le niveau de diplôme (en %)

Niveau de diplôme des 25-49 ans (hors lycéens et étudiants)	Panel Répondants (n = 271)	INSEE 2011
Niveau Bac et inférieur	21,4	64,1
Bac+2/+3	29,5	15,4
Bac+4 /+5	49,0	20,5
Total général	100,0	100,0
<i>Part des 25-49 ans dans la population</i>	<i>38,1</i>	<i>32,5</i>

Source : INSEE

Nous n'avons pas pu comparer le niveau d'étude des membres du panel avec celui des inscrits sur Twitter. Contrairement au sexe, à l'âge et à la PCS, le niveau de diplôme n'est pas une variable mobilisée par les études de l'IFOP sur la sociologie des réseaux sociaux.

La surreprésentation des plus diplômés au sein de notre panel confirme donc que le niveau d'étude, ou le capital scolaire, est fortement corrélé à la propension à produire des tweets politiques.

Mais encore une fois, la structure par âge de notre population ne permet pas d'expliquer à elle seule cette surreprésentation des plus diplômés. La position sociale des producteurs de tweets politiques est également un élément primordial à prendre en compte. Le niveau de capital scolaire constitue en effet pour un certain nombre de professions, et notamment pour les plus valorisées symboliquement et socialement, un droit d'entrée. Ainsi, alors que 65,3 % des producteurs de tweets politiques appartenant à la catégorie des cadres et professions intellectuelles supérieures ont un niveau de diplôme supérieur ou égal à Bac+4, ce taux n'est « que » de 31,5 % pour les professions intermédiaires et de 0% pour les ouvriers (dont les effectifs sont rappelons-le, très faibles). La surreprésentation statistique des cadres et professions intellectuelles supérieures élève donc le niveau général d'étude des producteurs de tweets politiques.

D'un point de vue plus général, la surreprésentation des plus diplômés au sein de notre échantillon n'est sans doute pas en soi étonnante et confirme l'hypothèse selon laquelle les producteurs de tweets politiques présentent des propriétés sociologiques et un niveau de politisation bien plus élevés que la moyenne de la population française. La propension à prendre publiquement et spontanément la parole politique est bien fonction du niveau de

compétence politique statutaire des individus, et le niveau de compétence politique statutaire est lui-même fonction du capital scolaire.

Malgré leur relatif jeune âge, les producteurs de tweets politiques présentent des propriétés sociologiques en réalité très proches de celles des membres du « pays légal ». Ils sont davantage masculins, Parisiens, diplômés et occupent une position relativement élevée dans l'espace social.

Nous allons maintenant voir qu'au-delà de ces propriétés sociologiques, les producteurs de tweets politiques présentent des propriétés politiques qui les singularisent non seulement par rapport à la population française mais également par rapport aux membres du « pays légal ».

1.2 Une population fortement politisée et politiquement ancrée à gauche

Les producteurs de tweets politiques sont en effet très fortement politisés⁵⁷³. Ils sont ainsi significativement plus nombreux que la moyenne des Français à déclarer s'intéresser « beaucoup » à la politique, à participer aux élections et à être engagés au sein de partis, syndicats ou associations politiques (1.2.1). Cependant, contrairement à d'autres populations politisées, les producteurs de tweets politiques sont définis par une identité politique relativement homogène. Les électeurs de gauche sont en effet largement surreprésentés au sein de notre échantillon (1.2.2).

⁵⁷³ Affirmer que les producteurs de tweets politiques sont fortement politisés peut de prime abord sembler tautologique. Il n'est en effet pas surprenant qu'une population composée uniquement d'individus publiant des messages politiques témoigne d'un niveau élevé de politisation. Cependant, nous le verrons plus loin dans ce chapitre, cet échantillon réunit en réalité des individus tout de même hétérogènes, certains ne publiant que très occasionnellement des messages politiques, et ce uniquement en période de forte politisation. La corrélation entre la production de tweets politiques et le niveau de politisation pourra alors sans doute paraître moins évidente au lecteur.

1.2.1 Une population fortement politisée

Plusieurs indices permettent de mesurer le niveau de politisation d'un individu ou d'une population. Le niveau d'intérêt *déclaré* pour la politique est l'un de ceux les plus communément mobilisés par les travaux de sociologie politique. Nous avons également mobilisé cet indice en posant la question suivante aux producteurs de tweets politiques : « *D'une façon générale, vous diriez que vous vous intéressez à la politique : beaucoup, assez, un peu ou pas du tout* ».

56 % des individus qui ont répondu à notre questionnaire ont déclaré « beaucoup » s'intéresser à la politique, un résultat très nettement supérieur aux résultats enregistrés par des enquêtes administrées auprès d'échantillons représentatifs de l'électorat français. Ainsi, d'après le sondage post-électoral du CEVIPOF⁵⁷⁴, en juin 2012, seuls 21 % des électeurs français en âge de voter déclaraient s'intéresser « beaucoup » à la politique, un taux inférieur de 35 points à celui que nous avons enregistré auprès des producteurs de tweets politiques. La quasi-totalité des répondants à notre questionnaire (87 %, en additionnant les catégories « beaucoup » et « assez ») témoignent alors d'un intérêt élevé pour la politique.

Tableau 3.16 *Distribution de la population répondants selon l'intérêt déclaré pour la politique (en %)*

Intérêt pour la politique (en %)	Panel Répondants (n = 658)	Base enquête post-électorale Cevipof 2012 (n = 2504)	Base enquête CEE FES 2012 (n = 2014)	CEVIPOF janvier 2011 (n = 1908)⁵⁷⁵
Beaucoup	56	21	29	17
Assez	31	37	45	41
Un peu	11	31	16	32
Pas du tout	3	11	11	9
Total général	100	100	100	100

Sources: CEE, CEVIPOF

De manière analogue à ce que l'on peut observer avec des d'enquêtes administrées auprès d'échantillons représentatifs⁵⁷⁶, l'intérêt déclaré pour la politique des membres de notre panel varie significativement en fonction de variables sociologiques telles que le sexe

⁵⁷⁴ CEVOPOF, « Sondage post-électoral 2012 », 2012.

⁵⁷⁵ CEVIPOF, « Le baromètre de la confiance politique », janvier 2011.

⁵⁷⁶ Daniel Boy et Nona Mayer, *L'électeur à ses raisons*, Paris, Les Presses de Sciences Po, 1997.

($\chi^2 = 0,010$; $V = 0,131$), l'âge ($\chi^2 = 0,016$; $V = 0,202$) ou le niveau de diplôme ($\chi^2 = 0,020$; $V = 0,168$). Les hommes producteurs de tweets politiques sont ainsi significativement plus nombreux que les femmes à déclarer s'intéresser « beaucoup » à la politique (61,1 % contre 47,6 %). Mais si les femmes qui ont répondu à notre questionnaire s'intéressent moins à la politique que les hommes, elles restent proportionnellement plus nombreuses à s'y intéresser que l'ensemble des Françaises. Ainsi, d'après l'enquête post-électorale du CEE réalisée en 2012⁵⁷⁷, ce sont 24 % des électrices françaises qui déclaraient s'intéresser « énormément » ou « beaucoup »⁵⁷⁸ à la politique (contre 24 % des hommes), soit un taux inférieur de 24 points à celui des femmes de notre panel (47,6 %).

De la même manière, on observe à l'intérieur de notre panel comme à l'échelle de l'électorat français, une corrélation positive entre l'intérêt déclaré pour la politique et le niveau de diplôme. Plus un individu a un niveau d'étude élevé, plus il a tendance à s'intéresser à la politique. Cependant, quel que soit le niveau d'étude, la part de ceux qui déclarent s'intéresser « beaucoup » à la politique est nettement plus importante au sein de notre panel que dans l'ensemble de la population. Ainsi, les producteurs de tweets politiques ayant un niveau d'étude inférieur au Baccalauréat sont proportionnellement deux fois plus nombreux que les électeurs français du même niveau scolaire à s'intéresser « beaucoup » à la politique (39 % contre 20 %). À un niveau d'étude plus élevé (Bac+4 et Bac+5), les différences d'intérêt pour la politique entre producteurs de tweets politiques et électeurs restent sensibles, bien que moins marquées. Les membres de notre panel disposant d'un niveau d'étude équivalent à Bac+3/Bac+4 sont 63 % à s'intéresser beaucoup à la politique, contre 48% des électeurs français ayant atteint le même niveau d'étude.

La structuration de notre population, et notamment la surreprésentation statistique des catégories sociales supérieures, peut sans doute expliquer, du moins en partie, la plus grande politisation au sein de notre panel des catégories les moins intéressées par la politique (les moins diplômés, les femmes, etc.). Nous l'avons vu, la position des agents sociaux dans l'espace social détermine les variables secondaires telles que le sexe, l'âge, le lieu d'habitation ou encore le niveau de diplôme. Ainsi, si les femmes de notre panel sont plus intéressées par la politique que les femmes françaises dans leur ensemble, c'est peut-être avant tout parce qu'elles appartiennent davantage aux catégories sociales supérieures. On sait

⁵⁷⁷ Nicolas Sauger, « The French Electoral Study 2012 ». Données disponibles sur : <http://cee.sciences-po.fr/fr/elections-2012/enquete-electorale-francaise-2012.html>.

⁵⁷⁸ Nous avons regroupé les catégories « énormément » et « beaucoup » afin de pouvoir les comparer à notre modalité « beaucoup », ce qui n'est pas sans poser problème.

en effet que les femmes cadres sont plus politisées que les femmes employées⁵⁷⁹ et qu'elles sont aussi plus jeunes (et nous avons vu que les femmes les plus jeunes étaient aussi les plus politisées).

Cependant, la surreprésentation des catégories supérieures au sein de notre échantillon ne suffit pas à expliquer entièrement le plus grand intérêt déclaré pour la politique des producteurs de tweets politiques. Il suffit pour s'en convaincre de comparer l'intérêt déclaré par PCS au sein de notre panel et au sein de l'électorat français. Si on s'intéresse précisément aux cadres et professions intellectuelles supérieures – PCS surreprésentées sur Twitter – on s'aperçoit que les cadres de notre panel sont sensiblement plus nombreux (62 %) à déclarer s'intéresser « beaucoup » à la politique que les cadres pris dans leur ensemble (52 %, soit 10 points de moins)⁵⁸⁰. Cette plus forte politisation s'observe également au niveau des catégories sociales inférieures. Les employés de notre panel sont ainsi 40 % à déclarer s'intéresser « beaucoup » à la politique, soit un taux deux fois supérieur à celui des employés pris dans leur ensemble (19 %). Ces résultats confirment donc que les fractions populaires qui participent à Twitter sont particulièrement peu représentatives de leurs groupes sociaux.

Cette plus grande politisation des membres de notre panel s'observe également au niveau de leur participation électorale. Seuls 7 % des individus qui ont répondu à notre questionnaire déclarent ne pas avoir voté lors du premier tour des élections présidentielles du 22 avril 2012, alors que le taux d'abstention national s'élevait à 20,52 %⁵⁸¹. Cette plus forte participation concerne là encore toutes les catégories socio-professionnelles. Parmi les répondants, le taux d'abstention déclaré des cadres était de seulement 7 %, alors qu'au niveau national, il s'élevait à près de 14 %.

La forte participation électorale des producteurs de tweets politiques s'observe également, et peut-être avec encore plus de force, lors de scrutins générant traditionnellement plus d'abstention, comme les élections européennes. Si, comme au niveau national, c'est lors des élections européennes de 2009 que les membres de notre panel se sont le plus abstenus, leur taux d'abstention était cependant près de 25 points inférieur à la moyenne nationale⁵⁸².

⁵⁷⁹ Daniel Gaxie, *Le cens caché*, *op. cit.*

⁵⁸⁰ Cependant, si l'on additionne les modalités « beaucoup » et « assez », les résultats deviennent alors identiques : 92 % des cadres « électeurs » et 90 % des cadres de notre panel s'intéressent à la politique. On peut néanmoins supposer que la différence qui se joue entre la modalité « beaucoup » et l'ensemble des autres modalités est la plus discriminante et sans doute la plus significative.

⁵⁸¹ Résultats officiels du Ministère de l'intérieur.

⁵⁸² Précisons toutefois que contrairement au taux d'abstention qui est calculé au niveau national à partir d'indicateurs objectifs, le taux d'abstention des membres de nos panels a été obtenu à partir de données

Alors que les électeurs français n'ont été en effet qu'une minorité à participer à ce scrutin (41%), deux tiers des producteurs de tweets politiques se sont au contraire exprimés (59,7 %).

Au-delà de cette forte participation électorale, le niveau élevé de politisation des producteurs de tweets politiques se manifeste également par une pratique importante d'activités politiques dites « non-conventionnelles ». Les individus qui ont répondu à notre questionnaire sont ainsi une grande majorité à avoir déjà signé une pétition ou à avoir déjà manifesté.

Tableau 3.17 Pratiques politiques des répondants

Avez-vous déjà...	Répondants (n = 658)
Signé une pétition	80,7%
Manifesté	61,1%
Fait grève	32,4%
Adhéré à un syndicat	15,8%
Adhéré à un parti politique	30,5%
Adhéré à une ou plusieurs association(s)	48,5%

Mais le résultat politiquement le plus remarquable de notre recherche réside dans la très forte proportion de militants politiques au sein de notre panel. 31 % des individus qui ont répondu à notre questionnaire déclarent en effet avoir déjà adhéré à un parti politique ou à un syndicat. Près d'un producteur de tweets politiques sur trois est, ou a été, un militant politique.

Ce taux s'élève même à 59 % si l'on inclut les individus qui ont adhéré à une ou plusieurs association(s). Néanmoins, nous ne mobiliserons pas ce résultat agrégé dans notre démonstration. N'ayant pas demandé aux enquêtés de préciser le nom des associations auxquelles ils déclarent « avoir adhéré ou déjà adhéré », nous n'avons pas pu différencier les associations qui revêtent une dimension politique (du type défense d'une cause, comme Greenpeace, ATTAC, DAL, etc.) de celles qui ne sont au contraire pas définies par une identité politique (associations sportives, de voisinage, etc.). Nous avons en effet fait le choix de ne pas demander aux répondants de préciser les syndicats, partis, ou associations auxquelles ils déclareraient adhérer (cependant, concernant les militants des partis politiques, on a pu hypothétiquement identifier ces partis en croisant la variable « militant » avec la question « de quel parti vous sentez-vous le plus proche... »). Nous partions du principe que ces demandes supplémentaires d'information rallongeraient sensiblement le temps de remplissage du

déclaratives formulées *a posteriori* (nous avons interrogé nos enquêtés en 2012 sur les votes de 2007 à 2012), ce qui n'est pas sans poser problème.

questionnaire, et que surtout, ces informations pourraient être obtenues directement par le chercheur via l'observation des profils. Nous avons en effet constaté, lors d'observations préliminaires, que sur Twitter un grand nombre d'individus indiquaient explicitement sur leur profil leur activité militante et les noms des partis, syndicats ou associations politiques pour lesquelles ils militaient. Ce que nous n'avions pas anticipé, c'est que des individus pouvaient adhérer à un parti, un syndicat ou une association sans le spécifier sur leur profil, sans publiciser leur engagement. Ainsi, au sein du panel « répondants », seuls 24% des individus qui déclarent avoir déjà adhéré à un parti politique ou un syndicat ont indiqué sur leur profil être militant et pour quel parti ou syndicat ils militaient. Enfin, précisons que la formulation de la question pose également problème (« avez-vous déjà... ») : elle ne permet pas de dissocier une action révolue (l'individu a adhéré mais n'adhère plus aujourd'hui) d'une action toujours en cours (l'individu est toujours adhérent au moment de l'enquête). En résumé, nous désignerons désormais sous l'appellation « militants politiques » les individus qui ont répondu « adhérer » ou avoir « déjà adhéré » à un parti politique.

Par rapport aux populations étudiées dans la majorité des travaux de sociologie politique, notre population se situe dans une sorte « d'entre deux ». Lorsque ces travaux prennent pour objet l'électorat français, les militants sont « invisibles », précisément parce que leur poids relatif dans la population totale est tout à fait marginal. De l'autre côté, les nombreux travaux sur le militantisme centrent exclusivement leur focale sur les militants et ont tendance à les étudier en « vase clos »⁵⁸³, s'intéressant assez peu aux autres champs dans lesquels ces agents sont également positionnés.

Dans notre échantillon, les militants sont très largement « visibles » sans être pour autant majoritaires : 70 % des individus qui ont répondu à notre questionnaire n'ont en effet jamais adhéré à un parti politique. De plus, sur Twitter, ces agents sociaux ne sont pas analytiquement réduits à leur statut de militant. En effet, nous le verrons, la majorité des militants ne font pas de leur profil Twitter un instrument exclusivement dédié à leur activité militante et laissent entrevoir par ce biais certaines autres dimensions de leur habitus.

D'un point de vue sociologique, le profil des militants de notre panel tend à se rapprocher davantage de celui des militants « hors-ligne »⁵⁸⁴. Les militants sur Twitter sont ainsi sensiblement plus âgés que les non-militants (39 % des militants de notre panel ont

⁵⁸³ Frédéric Sawicki et Johanna Siméant, « Décloisonner la sociologie de l'engagement militant. Note critique sur quelques tendances récentes des travaux français », *Sociologie du travail*, 51, 2009, p. 97-125.

⁵⁸⁴ Anaïs Théviot, « Qui milite sur Internet ? Esquisse du profil sociologique du « cyber-militant » au PS et à l'UMP », *Revue Française de Science Politique*, 2013/3-4 (Vol.63), p. 663-678.

moins de 30 ans, contre 65 % des non-militants), même s'ils restent tout de même significativement plus jeunes que les militants « hors-ligne » (10 % des militants du PS auraient moins de 30 ans⁵⁸⁵). La répartition hommes/femmes des militants sur Twitter se rapproche de façon encore plus nette de celle des militants « hors-ligne ». Le taux d'hommes parmi les militants est en effet quasiment identique au sein de notre panel (71 %) et parmi les militants du PS (70 %⁵⁸⁶) ou de l'UMP (68 %⁵⁸⁷), alors que ce taux n'est que de 54 % pour les non-militants du panel.

Concernant la catégorie socio-professionnelle, on enregistre en revanche très peu de différences entre ces trois populations (les militants Twitter, les non-militants Twitter, et les militants « hors-ligne ») : les catégories sociales supérieures y sont chaque fois surreprésentées, et ce dans des proportions relativement identiques (ils représentent 53 % des militants « Twitter », 50 % des non-militants « Twitter » et 49 % des militants « hors-ligne » du PS). Cependant, si l'on s'intéresse au secteur d'activité (public/privé), on retrouve là encore une plus grande similitude entre les militants sur Twitter et les militants « hors-ligne » qu'entre les militants et les non-militants Twitter. 48 % des militants sur Twitter appartiennent ainsi au secteur public (fonctionnaires ou contractuels de la fonction publique), contre 34 % des non-militants. Ce taux se rapproche donc davantage de celui des militants « hors-ligne » du PS, qui appartiennent à 59 % au secteur public. Ces résultats confirment donc que le profil des militants sur Twitter est très proche de celui des militants en général.

Fortement intéressés par la politique, votants assidus, militants : les producteurs de tweets politiques témoignent d'un niveau très élevé de politisation et s'apparentent aux fractions les plus politisées des populations qui participent « spontanément » à la politique.

Cependant, les producteurs de tweets politiques se singularisent par une identité politique relativement homogène.

⁵⁸⁵ Henri Rey, « Les adhérents socialistes : permanences et changements », *Revue socialiste*, 42, 2011, p. 125-135.

⁵⁸⁶ *Ibidem*.

⁵⁸⁷ Florence Haegel, « La mobilisation partisane de droite. Les logiques organisationnelles et sociales d'adhésion à l'UMP », *Revue française de science politique*, 59 (1), 2009, p. 7-27.

1.2.2 Une population fortement ancrée à gauche

Au-delà de ce niveau très élevé de politisation, les producteurs de tweets politiques se caractérisent en effet par une identité politique relativement homogène : l'électorat de gauche est prédominant sur Twitter. 63 % des membres de notre panel qui ont voté lors du premier tour des élections présidentielles du 22 avril 2012 ont choisi un candidat de gauche ou d'extrême-gauche.

Tableau 3.18 Vote au premier tour des élections présidentielles du 22 avril 2012

Vote premier tour présidentielle 2012	Population Répondants et votants (n = 491)	Suffrages exprimés (Ministère de l'intérieur)
François Hollande (PS)	34,2 %	28,6 %
Jean-Luc Mélenchon (Front de gauche)	20,8 %	11,1 %
Nicolas Sarkozy (UMP)	19,1 %	27,2 %
François Bayrou (MODEM)	13,2 %	9,1 %
Éva Joly (EELV)	6,7 %	2,3 %
Marine Le Pen (FN)	3,1 %	17,9 %
Nicolas Dupont-Aignan (DLR)	1,8 %	1,8 %
Philippe Poutou (NPA)	0,8 %	1,1 %
Jacques Cheminade (Solidarité et Progrès)	0,2 %	0,3 %
Nathalie Arthaud (Lutte ouvrière)	0,0 %	0,6 %
Total	100,0 %	100,0 %

Source : Ministère de l'intérieur

La surreprésentation de l'électorat de Jean-Luc Mélenchon est un exemple particulièrement manifeste de cette prédominance de la gauche : le candidat du Front de Gauche enregistre en effet un score deux fois supérieur (20,8 %) à sa moyenne nationale (11,1 %). Et inversement, l'électorat de droite, et plus encore celui d'extrême-droite, sont nettement sous-représentés sur Twitter. Seuls 3 % des producteurs de tweets politiques qui ont répondu à notre questionnaire déclarent avoir voté pour Marine Le Pen. Moins de 5 % des

répondants se situent d'ailleurs « très à droite » et seuls 6 % se définissent comme étant « ni de gauche, ni de droite »⁵⁸⁸.

Cette sous-représentation des électorats de droite et d'extrême-droite peut sans doute s'expliquer, une fois encore, par la structuration de notre population. Les ouvriers, employés, artisans, commerçants, indépendants et retraités sont très nettement sous-représentés sur Twitter, nous l'avons vu. Or ces catégories constituent précisément les électorats sociologiques de la droite et de l'extrême-droite⁵⁸⁹. Leur faible présence sur Twitter constitue donc un élément d'explication de cette sous-représentation de la droite et de l'extrême-droite.

Cependant, si l'on recentre l'analyse sur le vote des cadres – catégorie relativement bien représentée au sein de notre échantillon – des différences importantes continuent à se faire jour entre les producteurs de tweets politiques et les électeurs français pris dans leur ensemble. Alors que les cadres auraient voté à environ 30 % pour Nicolas Sarkozy lors du premier tour de scrutin⁵⁹⁰, seuls 19 % de ceux qui ont répondu à notre questionnaire déclaraient avoir voté pour le candidat de l'UMP le 22 avril 2012. Sur Twitter, les catégories sociales supérieures seraient donc tout de même plus à gauche que la moyenne⁵⁹¹.

Pour tenter d'expliquer cette prédominance de l'électorat de gauche sur Twitter, une autre hypothèse, davantage contextuelle, peut être ici mobilisée. Selon deux politistes américains, Christopher Anderson et Aida Paskevicate, les citoyens d'un pays parleraient plus fréquemment de politique lorsqu'ils ne partagent pas les mêmes convictions politiques que le gouvernement en place⁵⁹². Notre panel a été constitué en mars 2012, alors qu'un parti de droite, l'UMP, était à la tête de l'ensemble des institutions nationales – la présidence de la République, le Gouvernement et l'Assemblée Nationale – et ce depuis presque 10 ans (et même 17 ans pour la présidence de la République). En suivant la thèse de Anderson et Paskevicate, on pourrait alors émettre l'hypothèse suivante : c'est parce que les instances

⁵⁸⁸ D'après le sondage post-électoral du CEVIPOF, ce sont 22 % des Français qui ne se déclarent « ni de gauche, ni de droite ».

⁵⁸⁹ Pascal Perrineau, *Le vote normal. Les élections présidentielle et législatives d'avril-mai-juin 2012*, Paris, Presses de Sciences Po, 2013.

⁵⁹⁰ http://www.bva.fr/data/sondage/sondage_fiche/1130/fichier_intention_de_vote_-_le_parisien21a91.pdf

⁵⁹¹ Il serait évidemment plus rigoureux de mobiliser ici des catégories plus fines, tenant compte du secteur d'activité (privé/public), du volume et de la structure du capital (économique et culturel), du patrimoine, etc. Or la taille relativement réduite de notre échantillon de répondants ($n = 608$) ne nous permet pas d'opérer ces distinctions.

⁵⁹² Christopher J. Anderson et Aida Paskevicate, « Macro-politics and micro-behavior: mainstream politics and the frequency of political discussion in contemporary democracies », in Alan Zuckerman (dir.), *The social logic of politics. Personal networks as contexts for political behaviour*, Philadelphie, Temple University Press, p. 228-250, 2005.

dirigeantes de leur pays étaient détenues par un parti dont ils partageaient les convictions politiques que les électeurs de droite se seraient moins exprimés politiquement sur Twitter. Et inversement, c'est parce qu'ils se situaient dans « l'opposition » au Gouvernement et au Président en place que les électeurs de gauche se seraient davantage exprimés politiquement sur le réseau social. En suivant cette hypothèse, Twitter pourrait donc être appréhendé comme un « tribunal de la raison » des gouvernants, comme un espace essentiellement critique.

Afin de discuter cette hypothèse, plusieurs résultats empiriques peuvent être ici exposés. En termes de volume de production de messages, Nicolas Sarkozy, lorsqu'il était Président de la République, était, et de très loin, la personnalité politique la plus citée sur Twitter, et ce jusqu'à sa défaite électorale du 6 mai 2012. Ainsi, en mars 2012 (le mois où a été constitué notre panel), Nicolas Sarkozy était deux fois plus cité (199 810 comptes uniques ont publié au moins un tweet citant le candidat) que François Hollande, qui était alors candidat (85 826 comptes uniques). Mais Nicolas Sarkozy était également la personnalité politique qui générait le plus de messages hostiles sur Twitter⁵⁹³. D'après une étude commerciale de l'institut Netscope⁵⁹⁴, en mars 2012, 39 % des tweets évoquant Nicolas Sarkozy étaient négatifs (54 % étaient neutres et 7 % positifs), contre 20 % des tweets citant François Hollande (71 % neutres, 9 % positifs)⁵⁹⁵. Et toujours au mois de mars 2012, d'après notre enquête, 63 % des individus qui ont émis ces tweets politiques étaient de gauche. Au sein de notre panel « répondants », les électeurs de François Hollande avaient également davantage cité Nicolas Sarkozy (76 %) que François Hollande (57 %).

Cependant, exposer ces résultats ne suffit pas à valider – ou invalider – notre hypothèse. Pour cela, il faudrait également étudier ces mêmes données depuis que la Présidence de la République, le Gouvernement et la majorité parlementaire ont changé de bord politique au printemps 2012. Les individus qui publient des tweets politiques depuis cette date sont-ils davantage de droite ? Il aurait alors été intéressant de reproduire à l'identique notre protocole de recherche un an après, au mois de mars 2013 et ainsi recenser l'exhaustivité des tweets politiques publiés sur cette période, envoyer un questionnaire à un échantillon d'au minimum 10 000 individus, etc. Ces résultats nous auraient alors permis de déterminer dans quelle mesure la composition politique des inscrits sur le réseau social se

⁵⁹³ On pourrait également émettre l'hypothèse selon laquelle Nicolas Sarkozy serait une personnalité politiquement plus clivante que François Hollande et que ce serait la raison pour laquelle il susciterait plus de réactions sur le réseau social.

⁵⁹⁴ Nous avons nous-mêmes participé à la création de l'institut Netscope. Cf. introduction.

⁵⁹⁵ Étude réalisée du 1^{er} mars au 19 avril 2012, à partir d'une extraction aléatoire de plusieurs milliers de tweets codés manuellement.

serait rééquilibrée – ou non – en faveur de l'électorat de droite depuis l'arrivée de la gauche au pouvoir en mai 2012. On peut simplement ici se borner à constater qu'en mars 2013, François Hollande, cette fois Président de la République, était à son tour la personnalité politique la plus citée sur Twitter (463 940 tweets), loin devant les figures d'opposition de la droite telles que Nicolas Sarkozy (169 540 tweets), Jean-François Copé (64 710 tweets) ou François Fillon (26 070 tweets).

S'il nous est donc impossible de vérifier précisément l'hypothèse de Christopher Anderson et Aida Paskeviciute, elle reste cependant une piste de réflexion intéressante afin de comprendre la surreprésentation de la gauche au sein de la population des producteurs de tweets politiques.

L'analyse des propriétés sociologiques et politiques des producteurs de tweets politiques permet de valider l'hypothèse selon laquelle l'échantillon spontané est la limite vers laquelle tend la population des répondants d'une enquête auprès d'un échantillon représentatif⁵⁹⁶. Ce sont en effet les fractions les plus politisées des catégories sociales elles-mêmes les plus politisées de la population qui ont publié des tweets politiques durant la dernière séquence présidentielle. Twitter constitue en cela donc bien un véritable observatoire du « pays légal ».

Cependant, nous allons maintenant démontrer que de nombreuses différences se font jour au sein de notre échantillon. Si le plus petit dénominateur commun des membres de notre panel est d'avoir publié au moins un tweet évoquant un acteur politique, l'analyse de leurs profils et de leur production de tweets laisse entrevoir l'existence d'un continuum de dispositions relativement étendu.

⁵⁹⁶ Pierre Bourdieu, *La Distinction, op. cit.*, p. 482.

2. Les usages socialement distinctifs de Twitter

Si la population des producteurs de tweets politiques est donc relativement homogène sociologiquement et politiquement, différant ainsi nettement de l'ensemble de la population française, nous allons maintenant démontrer que l'on peut y observer également certaines différences significatives. Ces différences peuvent s'objectiver notamment en analysant les pratiques sociales en ligne de nos enquêtés.

En effet, si nous avons jusqu'ici objectivé la position sociale des producteurs de tweets politiques à partir de données *déclaratives* (sexe, âge, profession, niveau d'intérêt pour la politique, etc.), obtenues principalement par le biais de l'interrogation⁵⁹⁷, bien d'autres indices sociologiques s'offrent au chercheur lorsqu'il analyse l'activité et les pratiques des agents sociaux sur Twitter.

Comme l'a souligné Dominique Cardon, « *l'identité [des internautes] se trouve de plus en plus déposée dans les produits [des] activités que les [internautes] exhibent aux autres* »⁵⁹⁸. Publier une représentation photographique de soi, manifester une humeur ou une opinion, s'abonner au compte d'un artiste ou d'un journaliste, etc. constituent autant de pratiques objectivement classables et potentiellement révélatrices de la position des internautes dans l'espace social. Ainsi, « *sous forme objective ou subjective, ce sont donc les caractéristiques de l'individu les plus incorporées à sa personne qui peuvent être alors identifiées* »⁵⁹⁹. Nous proposons de nous inscrire dans la continuité de cette réflexion, en affirmant que ces pratiques « numériques » peuvent être appréhendées comme des manifestations empiriques de certaines dispositions constitutives de l'habitus.

En reprenant les termes de Pierre Bourdieu, nous définirons l'habitus comme un système de dispositions durables incorporées dès la prime enfance, « *générateur de pratiques objectivement classables [et] comme un système de classement de ces pratiques* »⁶⁰⁰. Les activités des agents sociaux sur Twitter peuvent alors être considérées comme des pratiques objectivement classables (par le chercheur), générées par l'habitus. Mais ces pratiques « numériques » peuvent être également évaluées, classées voire hiérarchisées par les agents

⁵⁹⁷ Mais aussi par celui de l'observation, dans le cas des non répondants.

⁵⁹⁸ Dominique Cardon, « Le design de la visibilité. Un essai de cartographie du web 2.0 », *Réseaux*, 2008/6, n° 152, p. 93-137.

⁵⁹⁹ *Ibidem*.

⁶⁰⁰ Pierre Bourdieu, *La Distinction, op. cit.*, p. 190

sociaux eux-mêmes. Une majorité de ces agents développent ainsi, plus ou moins explicitement et plus ou moins consciemment, des *usages socialement distinctifs* de Twitter.

Ces usages sociaux différenciés et distinctifs sont avant tout révélateurs d'une pluralité de dispositions dont l'analyse statistique⁶⁰¹ ne peut à elle seule rendre compte. Cette diversité de dispositions se manifeste notamment par la façon dont les individus se mettent en scène et en récit à travers leur description de profil (2.1). La place et la fonction assignées à la politique et à l'actualité sur les comptes trahit également certaines différences dispositionnelles. Nous tenterons d'en rendre compte en dressant une typologie des producteurs de tweets politiques (2.2).

Il est alors important de préciser en préambule de cette section que la dénomination de « producteurs de tweets politiques » employée jusqu'ici regroupe à la fois des agents pour lesquels la publication de messages politiques constitue la principale activité sur le réseau social mais aussi des agents qui au contraire ne conçoivent pas la discussion et la veille d'information politique comme la fonction première de Twitter, et qui ne publient en réalité que très peu de tweets à caractère politique. L'appellation de « producteurs de tweets politiques » ne doit alors pas laisser sous-entendre au lecteur que l'ensemble des agents désignés sous ce terme ont pour principale activité de publier des messages politiques.

2.1 L'« identité numérique » comme manifestation empirique de l'habitus

Les recherches sur l'« identité numérique » ont tendance à se focaliser sur les spécificités « numériques » de cette « identité », sans nécessairement chercher à situer ces « identités numériques » dans le monde social « réel ». Pourtant, tout laisse à penser que les pratiques « numériques » des agents sociaux sont générées par le même système de dispositions que leurs pratiques sociales « réelles » (2.1.1). L'analyse des descriptions de profil des producteurs de tweets politiques en apporte une preuve empirique (2.1.2).

⁶⁰¹ Qui a tendance à mettre en avant, comme nous l'avons vu dans la précédente section, l'homogénéité de notre échantillon en termes d'âge, de sexe ou de PCS.

1.1.1 Du concept d'« identité numérique » à celui d'habitus

La notion d'« identité numérique » a été développée dans le courant des années 2000 – principalement en science de l'information et de la communication – dans des travaux prenant pour objet les usages sociaux des espaces de communication et d'autopublication sur internet, et notamment les blogs et les réseaux sociaux⁶⁰². Fanny Georges définit l'identité numérique comme « *une transposition graphique, sonore et visuelle d'une représentation en pensée façonnée par le sujet [l'internaute] dans le matériau de l'interface [c'est-à-dire l'interface du réseau social ou du site internet]* »⁶⁰³. L'auteure identifie trois ensembles de signes qui constitueraient cette identité numérique : l'« *identité déclarative* », l'« *identité agissante* » et l'« *identité calculée* ». L'« *identité déclarative* » se compose « *des données saisies par l'utilisateur (exemple : nom, date de naissance, photographie)* ». Il s'agit donc des données socio-démographiques – ou « *signalétiques* », pour reprendre la terminologie des sondeurs – dont nous nous sommes servis pour qualifier sociologiquement les individus qui n'avaient pas répondu à notre questionnaire. L'« *identité agissante* » est constituée du « *relevé explicite des activités de l'utilisateur par le système « exemple : x et y sont maintenant amis » ; on pourrait également ajouter « W aime Madona » qui nous renseigne sur les goûts culturels des individus)* ». Enfin, l'« *identité calculée* » se manifeste par « *des variables quantifiées produites d'un calcul du Système (exemple : nombre d'amis, nombre de groupes)* »⁶⁰⁴. Dans le cas de Twitter, on pourrait alors citer le nombre d'abonnés, d'abonnements ou de retweets.

Pour Fanny Georges, l'identité numérique se différencie de l'identité réelle en ce qu'elle matérialise ce qui n'a précisément pas de forme matérielle dans le réel, comme par exemple les pensées de l'individu : « *dans la vie quotidienne « non interfacée », l'identité personnelle n'a pas de forme matérielle, exceptée celle du corps. [...] À l'écran, l'identité personnelle s'informe différemment : les logiciels de Communication médiée par ordinateur (CMO) donnent forme à ce qui n'a pas de forme dans le réel, comme les centres d'intérêt ou les pensées* »⁶⁰⁵. L'auteure mobilise alors le concept d'« hexis numérique » pour qualifier

⁶⁰² Fanny Georges, « Représentation de soi et identité numérique. Une approche sémiotique et quantitative de l'emprise culturelle du web 2.0 », *Réseaux*, 2009/2, n°154, p. 165-193 ; Fabien Granjon et Julie Denouël, « Exposition de soi et reconnaissance des singularités subjectives sur les sites de réseaux sociaux », *P.U.F.* 2010/1, n°1, p. 25-43 ; Dominique Cardon, « Le design de la visibilité. Un essai de cartographie du web 2.0 », art. cit.

⁶⁰³ Fanny Georges, « Représentation de soi et identité numérique. Une approche sémiotique et quantitative de l'emprise culturelle du web 2.0 », art. cit.

⁶⁰⁴ *Ibidem*.

⁶⁰⁵ Fanny Georges, art. cit.

l'ensemble de ces signes identitaires, qui serait alors l'équivalent sur internet du concept d'hexis corporelle.

On retrouve alors avec le concept d'« hexis numérique » une certaine analogie avec le concept d'habitus, ce système de représentations faites corps. De la même manière, lorsque Dominique Cardon souligne qu'à travers le « *récit intime, l'introspection, la représentation photographique, la manifestation d'humeur* », ce sont « *les caractéristiques les plus incorporées* »⁶⁰⁶ des internautes qui se trouvent matérialisées sur internet, on peut également repérer une certaine analogie avec le concept d'habitus.

Cependant, une recherche qui mobiliserait le concept d'habitus pour étudier les pratiques numériques des agents sociaux se différencierait selon nous des travaux précédemment cités sur deux principaux points. D'une part, l'habitus étant un principe générateur s'appliquant, « *par simple transfert, aux domaines les plus différents de la pratique* »⁶⁰⁷, on ne chercherait pas alors à analyser les pratiques numériques dans ce qu'elles ont de plus singulier, mais au contraire dans ce qu'elles ont de plus générique. Les travaux précédemment cités ont en effet tendance à opposer les notions d'« identité numérique » et d'« identité réelle », plutôt que de voir dans l'« identité numérique » une déclinaison transfigurée de l'« identité réelle ». Le postulat de ces travaux, qui est plus généralement celui des recherches en sciences de l'information et de la communication, est de considérer que « l'interface » – c'est-à-dire l'architecture d'un site internet – « module » l'identité des internautes, voire la modifie. Ainsi, pour Fanny Georges, « *étudier l'identité dans le contexte du web 2.0 permet de mieux comprendre comment les interfaces numériques changent le regard porté sur soi, l'Autre et le monde* »⁶⁰⁸. Certes, porter un regard critique sur ces dispositifs sociotechniques peut s'avérer salutaire, tant ces dispositifs semblent aujourd'hui, comme le souligne à juste titre l'auteure, devenir invisibles aux yeux des « utilisateurs », qui n'ont plus conscience des effets qu'ils peuvent produire sur leurs activités en ligne. Cependant, affirmer que ces interfaces transforment l'identité des individus, et plus généralement que l'« identité numérique » est différente par essence de l'« identité réelle », revient à nier le fait que les pratiques des internautes sont définies, comme l'ensemble des pratiques sociales, par les mêmes conditions matérielles d'existence.

⁶⁰⁶ Dominique Cardon, « Le design de la visibilité. Un essai de cartographie du web 2.0 », *Réseaux*, 2008/6, n°152, p. 93-137.

⁶⁰⁷ Pierre Bourdieu, *La Distinction, op. cit.*, p. 190.

⁶⁰⁸ Fanny Georges, « Représentation de soi et identité numérique. Une approche sémiotique et quantitative de l'emprise culturelle du web 2.0 », art. cit.

Sur internet, les individus ont certes la possibilité, comme l'a analysé Dominique Cardon notamment, d'« *endosser des rôles qui échappent à leur univers quotidien* » et de pouvoir ainsi « *styliser et modifier leurs apparences jusqu'à menacer l'authenticité des définitions de soi* », instaurant ainsi une réelle « *distance entre identité numérique et identité réelle* »⁶⁰⁹. Cependant, cette production d'une identité virtuelle totalement fantasmée ne devrait pas être étudiée comme un objet à part entière, que l'on opposerait à l'« identité réelle », mais devrait être au contraire analysée au prisme des conditions matérielles d'existence des individus qui en sont les auteurs. La mise en relation de l'habitus d'un individu avec son identité virtuelle – même la plus fantasmée – permettrait sans doute de mieux saisir les logiques sociales qui président à cette volonté – voire ce besoin – de « *rêve de vol social* »⁶¹⁰, cette volonté de se projeter dans un monde « virtuel », où toutes les virtualités – physiques, sexuelles, sociales – sont encore permises.

Le second point sur lequel peut s'observer une différence entre une recherche qui aurait recours au concept d'habitus pour analyser les pratiques numériques et les travaux qui mobilisent la notion d'« identité numérique » tient justement au caractère proprement « collectif » du concept d'habitus. En effet, l'habitus n'est pas, comme l'identité, une notion individuelle. Un habitus est commun à l'ensemble des agents sociaux qui partagent des conditions matérielles d'existence relativement homogènes. Le concept d'habitus est donc étroitement lié à celui de classe. Si les travaux consacrés à l'« identité numérique » entreprennent pour certains de classer ou de catégoriser les pratiques numériques des internautes, ils ne le font pas à partir de critères socio-économiques « réels », mais uniquement à partir de données générées par les réseaux sociaux (le nombre d'amis ou d'abonnés par exemple, c'est-à-dire en fonction de l'« identité agissante » ou « calculée » pour reprendre les termes de Fanny Georges⁶¹¹). La non-prise en compte de ces données socio-économiques s'explique sans doute par des raisons essentiellement pratiques : seule l'administration d'un questionnaire peut permettre d'obtenir des données de ce type et ce de façon systématique, comme nous l'avons vu dans les précédents chapitres.

⁶⁰⁹ Cardon, « Le design de la visibilité. Un essai de cartographie du web 2.0 », art. cit., p. 101-102.

⁶¹⁰ Pierre Bourdieu, *La Distinction*, op. cit., p. 429.

⁶¹¹ Fanny Georges, « Représentation de soi et identité numérique. Une approche sémiotique et quantitative de l'emprise culturelle du web 2.0 », art. cit.

Nous allons donc précisément nous attacher à analyser les pratiques numériques des producteurs de tweets politiques à l'aune de leurs caractéristiques socio-économiques « réelles », obtenues *via* l'administration de notre questionnaire notamment. Nous placerons alors plus particulièrement la focale sur un type de pratique socialement distinctive : la rédaction d'une description de profil.

2.1.2 Se « décrire » en moins de 160 caractères, une pratique socialement distinctive

L'inscription au réseau social Twitter se déroule en trois étapes. Après avoir indiqué une adresse *email* et choisi un « nom d'utilisateur » et un mot de passe, le nouvel inscrit est invité à s'abonner à des comptes d'autres utilisateurs⁶¹². Enfin, la troisième et dernière étape consiste à choisir une photo de profil et à rédiger « une biographie » d'au maximum 160 caractères, afin de « personnaliser » son profil.

Illustration 3.1 Dernière étape de l'inscription à Twitter



Personnalisez.
Ajoutez une photo et décrivez-vous.

Charger une image
Taille maximale de 700Ko.
JPG, GIF, PNG.

Biographie 0 / 160
Astronaute, amateur de bons vins, adore le cinéma, papa comblé.

Passer >

Source : <https://twitter.com/welcome/profile>

⁶¹² Nous analyserons les logiques sociales d'abonnement aux comptes de médias et d'hommes politiques dans le chapitre 5.

Cette injonction de « personnalisation » est alors suivie d'une « biographie-témoin », un exemple illustratif censé guider les nouveaux inscrits dans leur rédaction : « *Astronaute, amateur de bons vins, adore le cinéma, papa comblé* »⁶¹³. Au travers de sa « biographie », le nouvel utilisateur est donc invité à dévoiler sa profession (appréhendée avec humour mais non sans arrière-plan normatif dans l'exemple donné comme une vocation voire une passion), ses goûts en matière culturelle ainsi que son statut familial (on se gardera alors de préciser que ces quatre informations suffisent à faire du nouvel inscrit une potentielle cible marketing). La description « biographique » telle qu'elle est conceptualisée par les « architectes » du réseau social est donc avant tout « fonctionnelle » : elle consiste à décrire brièvement, sans véritable intention stylistique, une position (sociale et/ou familiale) et des goûts culturels.

Cependant, tous les utilisateurs ne respectent pas nécessairement cette règle du jeu. Comme l'avait observé Pierre Bourdieu, les usages sociaux ne se réduisent pas aux modes d'emploi⁶¹⁴. Il convient en effet de dissocier *les propriétés utiles* des *usages réels*. L'analyse des « biographies » des membres de notre panel en apporte une preuve empirique.

Avant de catégoriser les différents types de « biographies » et de tenter d'y déceler certaines intentions « distinctives », il convient de préciser qu'un premier usage *possible* consiste à ne pas rédiger sa « biographie » (le nouvel inscrit a en effet la possibilité d'ignorer cette étape) et de refuser ainsi cette injonction à « personnaliser » son profil. À l'instar de l'abstention, on peut alors interpréter ce « refus de rédaction biographique » comme *une intention consciente et délibérée* (l'utilisateur peut craindre que ces informations soient communiquées voire vendues à des sites commerciaux par exemple) ou comme *une incapacité sociale* à se « raconter ». L'analyse statistique permet alors d'apporter certains éléments de réponse à cette interrogation.

24 % des individus qui ont répondu à notre questionnaire n'ont pas rempli leur description de profil. Ce taux s'élève à près de 40 % pour les plus de 46 ans et n'est au contraire que de 19 % pour les moins de 25 ans. De la même manière, ce taux est de 43 % parmi les enquêtés ayant un niveau d'étude inférieur ou égal au baccalauréat, alors qu'il n'est que de 25 % pour les individus disposant au minimum d'une maîtrise.

⁶¹³ Source : Twitter.com.

⁶¹⁴ Pierre Bourdieu, *La Distinction, op. cit.*, p. 111.

Tableau 3.19 Distribution des répondants selon l'âge, le niveau scolaire et le fait d'avoir rédigé ou non sa biographie (en %, n = 608)

Âge et niveau scolaire	Rempli	Non-rempli
Moins de 18 ans	81,3	18,8
18 à 25 ans	81,9	18,1
26 à 34 ans	74,1	25,9
35 à 45 ans	73,9	26,1
46 à 60 ans	61,0	39,0
61 ans et plus	60,0	40,0
Total	76,0	24,0
Niveau bac et inférieur	57,5	42,5
Bac +2/+3	69,6	30,4
Supérieur ou égal àBac+4	75,2	24,8
En cours d'études	85,0	15,0
Total	76,0	24,0

La propension à dévoiler des informations « biographiques » est donc corrélée à l'âge ($\chi^2 = 0,005$; $V = 0,166$) et au niveau d'étude des enquêtés ($\chi^2 = 0,000$; $V = 0,224$). Ce résultat peut donner *a priori* un certain crédit à l'hypothèse d'une incapacité sociale à « se décrire » et à exposer ses « titres » et ses « trophées » pour les individus les plus démunis en capital culturel. On peut en effet supposer que les individus disposant d'un capital culturel moins élevé se sentent moins à l'aise à l'écrit pour rédiger une « biographie », aussi brève soit-elle. Cette étape de l'inscription peut d'ailleurs, pour certains, être assimilée à un exercice scolaire.

Un enquêté de 18 ans, titulaire d'un baccalauréat technologique, exprime explicitement, non sans une certaine ironie, cette analogie avec l'exercice scolaire de la rédaction : « *Pas du tout fort dans les redactions quand j'etais au collège alors me demande pas de faire ma bio* ». Cependant, en acceptant tout de même de rédiger sa « biographie », en ayant à son tour recours à l'impératif (en réponse à l'adjonction : « *ajoutez une photo et décrivez-vous* ») et en s'adressant directement à « Twitter » en « l' » anthropomorphisant et en « le » tutoyant (« *ne me demande pas* »), on peut supposer que cet enquêté, contrairement à d'autres, n'a pas réellement subi cette étape comme une violence symbolique.

Mais les individus disposant d'un capital culturel moins élevé peuvent d'autre part se sentir socialement moins légitimes à rédiger une « biographie ». Le terme même

de « biographie » peut sembler être réservé aux « grands hommes », ou aux hommes qui ont une histoire (« *des astronautes* », et non à des demandeurs d'emploi ou à des tourneurs-fraiseurs).

Mais pour tenter de comprendre ce refus de « se décrire », une seconde hypothèse peut être également formulée. Si certains individus n'ont pas rédigé de « biographie », c'est peut être avant tout parce que leur inscription au réseau social était hasardeuse, hésitante, « tâtonnante », ou n'était en tout cas pas sous-tendue par une réelle volonté de s'inscrire durablement au sein d'une nouvelle « communauté ». Un individu peut en effet *utiliser* Twitter de façon totalement passive : il peut lire et consulter les tweets d'autres utilisateurs sans jamais en produire lui-même⁶¹⁵, ni sans jamais interagir avec les autres utilisateurs.

Cette hypothèse d'un rapport moins investi à Twitter peut alors être vérifiée empiriquement en comparant le nombre de tweets publiés par ces utilisateurs « sans biographie » avec ceux ayant rempli leur description de profil. Alors que 87 % des individus qui ont rédigé une « biographie » publient plus d'un tweet (politique ou non) par semaine, les individus qui n'ont pas rédigé leur « biographie » sont seulement 59 % à le faire (41 % publient donc moins d'un tweet par semaine). Mais c'est en comparant le nombre d'abonnés de ces deux sous-populations que les différences se font les plus manifestes. 55 % des individus qui ont accepté de « se décrire » ont plus de 100 abonnés à leur compte Twitter, contre seulement 10 % de ceux qui n'ont pas rempli leur description. Les individus qui n'ont pas rédigé de « biographie » publient donc peu de tweets et ont un capital relationnel numérique très peu développé. Ces résultats semblent donc bien confirmer l'hypothèse du rapport moins investi à Twitter

Si on s'intéresse maintenant aux individus qui ont accepté de rédiger leur description de profil, trois catégories de « biographies » peuvent être alors distinguées : les descriptions « fonctionnelles » (elles représentent 36 % des biographies des répondants qui ont rédigé une description de profil), les descriptions « fonctionnelles et stylisées » (36 % également), et les descriptions « uniquement stylisées » (28 %).

La première catégorie, la « biographie fonctionnelle », correspond à la « biographie-témoin » évoquée plus haut. Elle consiste à exposer sa position (sociale, familiale, ou politique) et/ou ses goûts, sans véritable intention stylistique. On peut dissocier, au sein de ce

⁶¹⁵ Cependant les membres de notre panel ont tous nécessairement publié au moins un tweet politique, cf. chapitre 2.

premier groupe, trois sous-ensembles. Le premier est composé d'individus se définissant uniquement par leur position, le plus souvent professionnelle (« *Prof des écoles en congé parental.* », « *Ingénieur* », « *étudiante en sociologie* »). Ce premier sous-ensemble représente 72 % des « biographies fonctionnelles ». Un deuxième sous-ensemble est composé d'individus ne se définissant au contraire que par leurs goûts et leurs pratiques culturelles (« *Sport, politique et économie* », « *Passionné par le jeu-vidéo, le web, le livre, la musique, et tant d'autres choses plus ou moins importantes...* »). Ces descriptions représentent seulement 8 % des « biographies fonctionnelles ». Enfin, le dernier sous-groupe est composé d'individus dévoilant à la fois leurs positions et leurs goûts, reproduisant presque à l'identique le schéma de la « biographie-témoin » proposée par Twitter (« *Étudiant en droit (L3) Supporter du PSG, Fan de sport en tout genre (tennis, rugby, ...) et passionné de politique* »). Ce dernier sous-ensemble représente 20 % des « biographies fonctionnelles ».

Tableau 3.20 Répartition (en %, n = 460) des « biographies » selon la forme et le contenu

Type de « biographie » / contenu	Position(s) uniquement	Goût(s) uniquement	Positions et goûts	Mots d'esprit, citations	Total
Fonctionnelle	72	8	20	0	100
Fonctionnelle et stylisée	39	25	31	4	100
Stylisée	1	5	0	94	100
Total	37	13	17	33	100

La « biographie fonctionnelle et stylisée » comporte également des informations relatives à la position et/ou aux goûts de leurs auteurs, mais avec une intention stylistique plus prononcée, voire avec une réelle volonté de distinction. Ce deuxième type de description représente 36 % des « biographies » rédigées par les producteurs de tweets politiques qui ont répondu à notre questionnaire. Plusieurs éléments permettent de « styliser » sa « biographie » : l'humour, l'ironie, la dérision, la mobilisation de références culturelles, la manifestation d'une humeur affirmée ou encore l'usage d'une langue étrangère et notamment de l'anglais.

Encadré 2.1. Verbatims illustratifs de « biographies fonctionnelles et stylisées »

L’humour, l’ironie : « *Presque journaliste, bientôt athlète et prochainement super-héros* », « *Étudiant en école d’ingénieur, spécialité glande* », « *Faux marketeux. Faux doctorant. Faux écrivain. Faux comique. Vrai loser.* »

La dérision : « *Renard Nazi, étudiant dégénéré et littéraire infiltré en école d’ingénieur. Ici sans raison, j’observe et j’écoute.* », « *Webdesigner freelance dans la Loire. Philanthrope parfois (l’espoir n’est pas une denrée tarissable)* »

Une référence culturelle : « *La culture ne s’hérite pas, elle se conquiert ☼Malraux☼ La rose à la main le 22 avril & 6 mai Mktg+ Com, stagiaire à l’international* »

L’usage de l’anglais : « *Digital lover, photographic aficionados & cigars taster. Currently studying at Supdepub Bordeaux.* »

Ces « biographies » n’ont pas uniquement pour *fonction* de communiquer et de publiciser des informations que leurs auteurs évaluent comme définitoires de leur « personne » (ou de leur « personnalité »). En *stylisant* leur « biographie », en s’attachant à l’aspect *formel* de l’exercice, les individus témoignent également d’une volonté de se distinguer, d’affirmer leur « style personnel », cet ensemble de propriétés jugées exclusives, uniques, originales (en faisant preuve d’humour, de dérision, de sarcasme, d’esprit, etc.)⁶¹⁶.

Mais c’est avec le troisième type de « biographie », purement « formelle », que cette volonté de distinction se fait la plus manifeste (elles représentent 28 % des « biographies » étudiées). Les individus ne communiquent plus alors d’informations relatives à leur position ni à leurs goûts – se détournant ainsi du « mode d’emploi » délivré par Twitter – mais mobilisent cette « tribune » de 160 caractères pour exhiber leur « style personnel ». L’enjeu n’est plus alors de préciser qui on est ni ce que l’on fait, mais uniquement de démontrer (sans en expliciter la volonté) que l’on sait faire preuve d’esprit, d’originalité, de culture, d’humour, de sarcasme, de raillerie, etc. : « *A la limite du bon goût, sans jamais y tomber* », « *So we beat on, boats against the current, borne back ceaselessly into the past. F.S.Fitzgerald.* », « *Rien de tel que de résumer un long discours par une phrase assassine* ».

⁶¹⁶ Pierre Bourdieu, *La Distinction*, op. cit., p. 485.

Cependant la qualité formelle de ces « biographies », le type de langage, le type d'humour, ainsi que les citations et les références culturelles mobilisées n'appartiennent pas toutes nécessairement au registre de la culture légitime. Pour paraphraser Jean-Claude Passeron, les classes dominantes n'ont pas le monopole de la stylisation de la vie : « *la stylisation populaire de la vie se voit à l'évidence dans la culture adolescente [...] et, d'une manière générale, dans tout ce qui a trait à la " mise en scène de la vie quotidienne " et à la " présentation de soi "* »⁶¹⁷. Les références à la culture populaire, qui se donne également « *pour fin explicite la production de formes et de signes* »⁶¹⁸, est en effet perceptible dans les « biographies purement formelles » des producteurs de tweets politiques : « ~ Pourquoi prendre la vie au sérieux puisque de toute façon, on en sortira pas vivant ? ~ B. Marley #TeamCitations #BlackPower #Follow à La Méli_Melo », « Follow me !! &` La Vie te Sourira .FollowBack Pas Automatiquement!#BBM : 27E1C6F9 DANCEHALL IS MY LIFE ♡.. #TeamDancehall#TeamSousouLand #TeamDagging », « Coin Coin Coin ».

Qu'elle s'inscrive dans la culture légitime ou populaire, c'est parmi les catégories les plus jeunes de notre échantillon, et notamment les adolescents, que cette volonté de stylisation se fait la plus manifeste. Cependant, nous ne parlons pas seulement ici des adolescents issus des classes populaires (nous avons eu l'occasion de démontrer qu'ils étaient minoritaires au sein de notre échantillon). C'est bien les adolescents dans leur ensemble – qu'ils soient issus des classes inférieures comme des classes supérieures – qui ont le plus tendance à styliser leur « biographie ». Parmi les individus de moins de 18 ans qui ont répondu à notre questionnaire, plus d'une description de profil sur deux présente un aspect purement formel. Au contraire, seulement 4 % des plus individus de plus de 61 ans n'ont pas mentionné dans leur biographie des informations relatives à leur position ou à leurs goûts. La propension à styliser sa « biographie » semble donc être corrélée à l'âge des producteurs de tweets politiques ($\chi^2 = 0,000$; $V = 0,313$).

⁶¹⁷ Jean-Claude Passeron, « Symbolisme dominant et symbolisme dominé », Enquête [En ligne], 1 | 1985, mis en ligne le 03 juin 2013, consulté le 23 décembre 2013. URL : <http://enquete.revues.org/38> ; DOI : 10.4000/enquete.38

⁶¹⁸ Jean-Claude Passeron, « Symbolisme dominant et symbolisme dominé », art. cit.

Tableau 3.21 Type de « biographie » en fonction de l'âge (en %, n = 460)

Classe d'âge / Type de « biographie »	Fonctionnelle	Fonctionnel le stylisée	Stylisée	Non Remplie	Total
Moins de 18 ans	10	19	52	19	100
18 à 25 ans	27	34	21	18	100
26 à 34 ans	27	28	19	26	100
35 à 45 ans	32	20	22	26	100
46 à 60 ans	31	20	10	39	100
61 ans et plus	40	16	4	40	100
Total	27	27	22	24	100

On peut néanmoins se demander si cette plus forte propension à styliser sa « biographie » n'est pas davantage liée à un effet de génération, ou s'il n'existerait pas une variable cachée entre l'intention stylistique et l'âge des producteurs de tweets politiques. Sans pouvoir apporter des éléments de réponses empiriques à ces questions, on peut toutefois supposer l'existence d'une autre variable à prendre en compte dans l'analyse, que Fabien Granjon nomme le « capital technique »⁶¹⁹. L'auteur définit le capital technique comme « *la capacité de maîtrise des technèmes indispensables à l'appropriation de l'informatique connectée. Il se présente, comme pour le capital culturel, essentiellement sous trois aspects : a) à l'état incorporé, c'est-à-dire sous la forme de savoir-faire pratiques, de compétences et de dispositions durables ; b) à l'état objectivé, sous la forme de biens matériels, logiciels et d'accès aux réseaux. [...] c) à l'état institutionnalisé, par l'objectivation dans des titres scolaires, des diplômes, des brevets d'aptitudes* »⁶²⁰. Si l'auteur observe avec raison que cette espèce de capital est très inégalement distribuée dans la population⁶²¹ (les classes populaires en sont ainsi particulièrement dépourvues), on peut néanmoins supposer que l'incorporation de dispositions durables relatives à l'usage d'internet est aujourd'hui plus développée parmi la génération née dans les années 1990 – période marquée par la démocratisation de l'informatique domestique et de l'accès à internet – et ce précisément parce que les dispositions durables s'incorporent dès la (prime) enfance.

Un niveau plus élevé de capital technique pourrait alors expliquer une plus grande propension à connaître les « règles du jeu » des réseaux sociaux et à savoir les déjouer

⁶¹⁹ Fabien Granjon, « Inégalités numériques et reconnaissance sociale. Des usages populaires de l'informatique connectée », *Les Cahiers du numériques*, 2009/1, Vol. 5, p. 19-44.

⁶²⁰ Fabien Granjon, art. cit.

⁶²¹ Il est regrettable que nous n'ayons pas mesuré le niveau de capital technique par le questionnaire, il aurait été intéressant de confronter, comme pour le niveau d'intérêt pour la politique, les données déclaratives aux observations objectives.

(comme par exemple ne pas s'« inspirer » de la « biographie-témoin » délivrée par Twitter pour rédiger sa description de profil). On a ainsi pu observer que les adolescents et les étudiants, contrairement à d'autres catégories sociales, avaient davantage tendance à changer régulièrement de description et de photographie de profil. Si elle peut s'expliquer par différentes raisons, cette plus forte propension à modifier ses paramètres de profil révèle avant tout que les adolescents *savent* que l'on peut et *savent comment* on peut changer ces différents paramètres, une fois terminée la phase « guidée » de l'inscription. Contrairement aux générations plus anciennes, ces savoirs et savoir-faire apparaissent comme « naturels » (c'est-à-dire socialement incorporés) aux adolescents de notre panel.

Si la propension à styliser « sa biographie » semble donc bien corrélée à l'âge – et donc sans doute également au niveau de capital technique – des producteurs de tweets politiques, nous n'avons en revanche enregistré aucune relation significative entre le type de « biographie » et le milieu social. Cependant, et comme nous l'avons déjà évoqué plus haut, on peut dissocier, parmi les « biographies stylisées », les descriptions mobilisant des références et un langage appartenant à la culture légitime de celles mobilisant des références et un langage appartenant à la culture populaire. Or si nous n'avons pas objectivé statistiquement ces différents types de stylisation⁶²², nous avons cependant pu constater, *via* l'observation qualitative des « biographies stylisées », un lien évident entre le milieu social d'origine et le type de références, de signes et de langage mobilisé.

⁶²² D'un part en raison d'effectifs trop faibles, qui empêchent toute représentativité statistique, et d'autre part car dissocier de manière objective culture dominée et culture dominante nécessiterait de réactualiser les références culturelles des classes sociales, ce que notre recherche doctorale n'ambitionne pas de faire.

Tableau 3.22 Exemples de biographies stylisées

Description	Type de description	Caractéristiques
Suis moi je te fuis fuis moi je te suiiiiiiis #TeamChicha# TeamCorleone tu connais la famille	Stylisée « populaire »	Homme, 23 ans, titulaire d'un Bac professionnel, sans emploi.
• #LesMarseillaisARio & #LesAnges6 de la BOMBE ♥ ! Kikou → amaandine.z .1h1h. • ASK	Stylisée « populaire »	Femme, 20 ans, vendeuse prête à porter.
So we beat on, boats against the current, borne back ceaselessly into the past. F.S.Fitzgerald.	Stylisée « légitime »	Femme, 18 ans, étudiante en droit.
Orwello-orphelin. Europe. Politique. Anything that is contemporary history- or geography-related.	Stylisée « légitime »	Homme, 37 ans, enseignant du secondaire.

Ces résultats exploratoires, qui n'ont pas avec eux la force de la preuve statistique, semblent confirmer l'hypothèse selon laquelle les pratiques « numériques » des producteurs de tweets politiques sont générées par des systèmes de dispositions différents, révélateurs de leur position dans l'espace social. Nous allons voir maintenant que ces différences dispositionnelles se trahissent également dans le rapport que les membres de notre panel entretiennent avec la politique sur Twitter.

2.2 Une tentative de typologie des producteurs de tweets politiques

Contrairement à ce que peut laisser sous-entendre le terme générique de « producteurs de tweets politiques », la principale activité des membres de notre panel n'est pas nécessairement de discuter, commenter, (s') informer ou publiciser des opinions sur la politique. On peut en réalité distinguer, au sein de notre population, trois profils⁶²³

⁶²³ Au sens de « profil Twitter ». Cf. glossaire p. 464.

idéaltypiques : le profil « post-it »⁶²⁴, le « profil militant », et le consommateur « ostentatoire » d'informations politiques.

Avant de présenter plus en détail ces trois profils idéaltypiques, trois remarques préliminaires s'imposent. La première concerne la nature précisément idéaltypique de ces catégories. Il convient en effet de préciser qu'on ne rencontre dans la réalité aucun exemple de profil correspondant parfaitement à ces idéauxtypes. Les profils analysés au sein de notre échantillon *tendent* à s'inscrire dans une ou plusieurs de ces catégories, sans jamais en réunir toutes les propriétés. C'est pour cette raison que nous n'avons pas entrepris une mesure statistique de ces différentes catégories de profil.

La deuxième remarque concerne *la nature des tweets* produits, qui doit être dissociée de la *nature des profils* sur lesquels ils sont publiés, même si *certain types de tweets* se retrouvent plus fréquemment *sur certains types de profils*. On peut distinguer six catégories de tweets :

- Les *tweets-réponses* : des réponses à des tweets publiés par d'autres utilisateurs. L'enchaînement de tweets-réponses peut alors donner lieu à *une discussion* entre deux ou plusieurs utilisateurs.
- Les *retweets* : des tweets rédigés et publiés par d'autres utilisateurs qu'un utilisateur va republier à l'identique sur son profil.
- Les *live-tweets* : des tweets publiés durant un événement que l'on commente en direct (des émissions télévisées, ou des réunions publiques par exemple).
- Les *tweets avec lien URL* : des tweets contenant un lien URL renvoyant à une page internet.
- Les *tweets-photos* : des tweets contenant des images, commentées ou non.
- Les tweets sans lien ni photo.

Enfin, la dernière remarque concerne la nécessité de distinguer les usages en vue desquels les producteurs de Twitter ont conçu le réseau social et les usages sociaux réels des utilisateurs. Comme nous avons déjà eu l'occasion de l'évoquer plus haut en paraphrasant Pierre Bourdieu, les usages réels ne se réduisent pas aux modes d'emploi : « *ce que la science doit établir, c'est cette objectivité de l'objet [ici Twitter] qui s'établit dans le rapport entre un objet défini dans les possibilités et les impossibilités techniques qu'il offre et qui ne se livrent*

⁶²⁴ Nous reprenons ce terme employé par Dominique Cardon dans « Le design de la visibilité », art. cit. p.118.

que dans l'univers des usages sociaux (parmi lesquels, s'il s'agit d'un objet technique, l'usage en vue ou en fonction duquel le producteur l'a conçu) et les dispositions d'un agent ou d'une classe d'agents, c'est-à-dire les schèmes de perception, d'appréciation et d'action qui en constitueront l'utilité objective dans un usage pratique »⁶²⁵. Initialement, les usages imaginés par les concepteurs du réseau social étaient en effet sensiblement différents des usages sociaux réels que nous avons pu observer au sein de notre panel. Réaliser une veille d'information, discuter et commenter l'actualité, ou plus encore exprimer ses opinions politiques constituent des pratiques sociales relativement éloignées de celles imaginées par les « architectes » de Twitter.

Dans un article publié en 2008 – soit deux ans après la création de Twitter aux États-Unis – Dominique Cardon proposait alors une analyse qui assimilait les usages « théoriques » aux usages sociaux réels. Son analyse peut alors nous renseigner sur les usages initialement imaginés par les concepteurs du réseau social :

*« Ici [sur Twitter], ce que les personnes rendent visible à leurs amis est un micro-récit enfermant des indexations spatiales et temporelles fréquentes, souvent adressé par le truchement du téléphone mobile. Le micro-blogging marque en effet le croisement de deux familles d'outils de communication très différents, les outils de blogging et de communication quasi-synchrone du type MSN. Ces services étendent et généralisent le principe de la « petite phrase » de statut comme un signal identitaire envoyé fréquemment et en toute circonstance à sa communauté pour marquer un lieu, une information, un état d'esprit ou un événement – « Que fais-tu en ce moment ? » demande Twitter à ses utilisateurs. Cette visibilité en post-it ouvre un espace d'expression original dans lequel les personnes ne livrent ni leur identité civile (souvent déjà connue de ceux qui les « suivent » (followers) en s'abonnant à leurs micropublications) ni leurs productions personnelles, mais leur disponibilité, leur état d'esprit, leurs activités hic et nunc. Le cadrage identitaire qu'opère ce genre de dispositif déplace la figuration figée de l'identité des personnes vers le mouvement et l'agir, en leur enjoignant de produire une narration de soi continue, contextuelle et renouvelée ».*⁶²⁶

2.2.1 Le profil « post-it »

Le premier type de profil que nous avons identifié correspond assez bien à l'analyse proposée par Dominique Cardon : les auteurs de ces profils « post-it » développent en réalité

⁶²⁵ Pierre Bourdieu, *La Distinction*, op. cit., p. 111.

⁶²⁶ Dominique Cardon, « Le design de la visibilité », art. cit., p. 118-119.

un usage de Twitter relativement identique à celui imaginé par les concepteurs du site. Ces individus se livrent ainsi à « *une narration de soi continue, contextuelle et renouvelée* »⁶²⁷, souvent au travers de discussions quasi-synchrones avec leurs abonnés. Ces individus produisent alors un très grand nombre de tweets quotidien mais un très faible nombre de tweets à caractère politique. Contrairement aux autres producteurs de tweets politiques, la fonction assignée à Twitter et à la publication de tweets n'est pas de publiciser des opinions politiques, ni de partager ou de communiquer des articles de presse en lien avec la politique ou l'actualité, mais d'envoyer « *un signal identitaire [...] fréquemment et en toute circonstance à sa communauté pour marquer un lieu, une information, un état d'esprit ou un événement* »⁶²⁸. C'est alors seulement quand la politique devient un « événement », quand on assiste à une synchronisation du temps social avec le temps politique⁶²⁹, que ces individus publient des tweets politiques, notamment à travers des sessions de *live-tweets*⁶³⁰, lorsqu'ils commentent en temps réel des émissions politiques télévisées générant d'importantes audiences, comme le débat télévisé d'entre-deux-tours de l'élection présidentielle par exemple.

Encadré 2. Un exemple de profil « post-it »

C. est un lycéen de 18 ans en classe de terminale ES. Ses parents sont employés de commerce et originaires d'Afrique sub-saharienne. Il publie en moyenne 28 tweets par jour (il a publié 7670 tweets du 1^{er} mars au 30 novembre 2013). Bien qu'il déclare s'intéresser « beaucoup » à la politique, il n'a durant cette période publié que deux tweets à caractère politique, l'un pour se féliciter de l'adoption par l'Assemblée nationale du projet de loi sur le mariage homosexuel (il est lui-même homosexuel), l'autre pour se plaindre du désordre occasionné par une visite de François Hollande à Marseille, où il réside : " *Ils font tout un bordel paske ya hollande en ville" F U C K !* ». L'essentiel de ses tweets consiste à faire part à ses amis de ses activités, mêmes les plus routinières (« *Today Grass Mat pour moi !* »), à partager ses humeurs ou ses envies les plus diverses (« *J'veuw faire la fête ce Week end ! Mais c'est ma santé avant tout.* ») ou bien encore à invectiver ses amis pour plaisanter : « *Bande d'incultes aujourd'hui c'est la Chandeleur ! Vous est votre fête des crêpes ! J'disais ça j'avais 10ans.* ».

⁶²⁷ *Ibidem.*

⁶²⁸ Dominique Cardon, « Le design de la visibilité », art. cit.

⁶²⁹ Vincent Goulet, *Médias et classes populaires. Les usages ordinaires des informations*. Paris, INA Editions, 2010.

⁶³⁰ Cf. glossaire p. 464.

En dehors de ces situations littéralement *extra-ordinaires*, les profils « post-it » ne publient que très rarement des tweets politiques. C'est essentiellement parmi les adolescents et les étudiants que l'on retrouve ce type de profil⁶³¹.

2.2.2 Le « profil militant » et les « auteurs militants »

Le « profil militant » constitue une deuxième catégorie de profils de producteurs de tweets politiques. Précisons d'emblée que nous distinguons ici les « profils militants » des « auteurs militants » : tous les profils militants ne sont pas nécessairement tenus par des auteurs militants – c'est-à-dire *a minima* des adhérents à une organisation politique – et inversement, tous les militants « déclarés » ne sont pas forcément les auteurs d'un profil que l'on qualifierait de « militant ». Cependant, affirmer que les auteurs de profils militants ne sont pas tous militants n'est pas sans soulever un certain paradoxe : comment expliquer que des individus développent un usage militant de leur profil Twitter s'ils ne font partie d'aucune organisation militante ? Comme appréhender alors cette forme « individuelle » d'engagement ?

Jacques Ion, avec le concept d'engagement « distancié »⁶³² ou « affranchi », avait déjà permis de mieux appréhender certaines formes « nouvelles » d'engagement. Le modèle de l'engagement « affranchi », qui serait ainsi venu se substituer progressivement⁶³³ au modèle de l'engagement « militant », se caractérise par une forme d'engagement plus intermittente, plus informelle, plus instable : « à l'engagement symbolisé par le timbre renouvelable et collé sur la carte [d'adhérent], succéderait l'engagement symbolisé par le post-it, détachable et mobile : mise à disposition de soi, résiliable à tout moment »⁶³⁴. Les « nouveaux militants » sont alors multi-positionnés dans différentes organisations. Pour Fabien Granjon, internet permettrait à ces « nouveaux » militants de coordonner plus facilement leurs différents engagements. D'autre part, la notion de réseau et d'horizontalité véhiculée par internet correspondrait « pleinement aux formes d'engagement matériel ou personnel des militants de ces nouveaux mouvements organisés en réseaux, faisant preuve d'une grande défiance vis-à-

⁶³¹ Dominique Cardon, « Le design de la visibilité », art. cit.

⁶³² Jacques Ion, *La fin des militants ?*, Paris, Éditions de l'atelier, 1997.

⁶³³ La linéarité du modèle a été discutée par Annie Colovald et Rémi Lefèbvre, « Le militantisme socialiste n'est plus ce qu'il n'a jamais été. Modèle de l'"engagement distancié" et transformations du militantisme au Parti Socialiste », *Politix*, Vol. 26, n°102, 2013, p. 9-33.

⁶³⁴ Jacques Ion, *op. cit.*, p. 81.

vis des procédures de délégation et défendant une prise de parole aussi horizontale que possible »⁶³⁵.

Si l'on retrouve bien au sein de notre panel ces deux catégories de militants – qui, précisons-le une nouvelle fois, ne développent pas tous nécessairement un usage militant de leur compte Twitter –, nous avons également pu observer une troisième catégorie, que l'on pourrait qualifier de militants « auto-entrepreneurs ». Bien qu'ils ne soient engagés dans aucune organisation militante, ces individus développent un usage de leur profil Twitter très similaire à celui des militants « encartés » au sein d'organisations politiques. Avant de caractériser plus en détail les propriétés des « profils militants », communs à ces différentes catégories de militants, quelques éléments permettant de mieux comprendre cette troisième forme d'engagement « solitaire », « auto-entrepreneuriale », peuvent être ici abordés.

Nous avons eu l'occasion de réaliser un entretien semi-directif avec F., que nous considérons comme un « auto-entrepreneur militant ». F. n'était engagé dans aucune organisation politique (parti politique, syndicat, association de défense d'une cause). F. est un retraité de 65 ans, ancien cadre dans l'informatique. Il entretient un rapport passionné et passionnel à la *politics*, et nourrit un véritable dégoût envers la gauche. Il se dit de droite et soutient l'UMP, et plus particulièrement Nicolas Sarkozy. F. n'hésite pas à publiciser ses orientations/goûts politiques, comme par exemple dans sa description de profil : « *#Gauchophobe #Socialistophobe #Hollandophobe #Ecolophobe Suivi par @NicolasSarkozy – UMP* ». Il mobilise alors son compte Twitter pour commenter l'actualité politique et discuter avec d'autres utilisateurs, principalement des militants ou sympathisants UMP, parfois FN. Il ne connaît pas personnellement les interlocuteurs avec qui il s'entretient pourtant régulièrement sur Twitter. Il partage avec eux son dégoût envers le Président de la République, le Gouvernement et la gauche en général : « *@nom⁶³⁶ C'est vrai, y a pas plus #Vieux que les #Socialos et la #Gauche en général, le pays est #Gangrené par ces #Voyous !!* ». F. se positionne de manière systématique sur toutes les prises de position de la gauche (il réalise une véritable veille d'information plusieurs fois par jour), pour les dénoncer, les discréditer, les houspiller voire les moquer. En dehors de ces multiples sociabilités numériques, F. a en réalité un capital social très peu développé, du fait notamment d'un handicap qui l'isole socialement mais aussi géographiquement (il habite dans un petit village).

⁶³⁵ Fabien Granjon, « L'Internet militant. Entretien avec Fabien Granjon », *Matériaux pour l'histoire de notre temps*, Année 2005, Vol. 79, Numéro 79, p. 24-29.

⁶³⁶ Nous avons modifié les noms d'utilisateurs interpellés.

Ses activités numériques lui ont alors permis de nouer des relations virtuelles, qu'il qualifie souvent d' « amicales », et qui se sont, pour certaines d'entre elles, prolongées dans « la vraie vie ». Des militants UMP lui ont ainsi un jour proposé de venir le chercher en voiture pour l'amener avec eux à un meeting de Nicolas Sarkozy qui se tenait à plusieurs kilomètres de son domicile. Cette nouvelle socialisation militante et les micro-pressions contextuelles qu'elle a exercées, l'ont finalement amené à s'interroger sur une éventuelle adhésion à l'UMP. Outre la constitution d'un capital relationnel qu'il n'avait pas auparavant, F. tire de son auto-entreprise militante d'autres formes de rétribution symbolique. Il est ainsi extrêmement fier de constater que certaines « grandes personnalités » politiques ou médiatiques se sont abonnées à son profil Twitter (comme notamment Nicolas Sarkozy). F. a de plus le sentiment d'avoir créé avec son blog une véritable demande sociale, certains « twittos »⁶³⁷ le sollicitant occasionnellement pour qu'il rédige des articles sur certains sujets.

On peut donc supposer que les « auto-entrepreneurs militants » présentent certaines propriétés qui les prédisposent au militantisme politique (rapport passionné et passionnel à la politique, grille de lecture axiomatique du monde social, attente de rétribution symbolique et de reconnaissance sociale, etc.). Cependant, ils n'ont jamais pu – ou su – concrétiser leur engagement en s'engageant au sein d'une organisation, faute de capital social notamment. Ces individus qui sont socialement – mais parfois aussi géographiquement – éloignés du champ du militantisme ont alors pu trouver avec internet, et notamment Twitter, une possibilité de créer eux-mêmes une petite entreprise militante, et d'en tirer certaines rétributions symboliques, et notamment de développer un capital social militant « numérique » qui leur faisait défaut, et que certains ont pu convertir en capital social « réel ».

Cependant, ces éléments ne constituent qu'une hypothèse de recherche parmi d'autres. Pour tenter de comprendre l'engagement isolé de certains agents, une seconde hypothèse peut être également formulée.

Le cas de C., une professeure de Français de 46 ans, très intéressée par la politique, et « *profondément de gauche* » peut alors nous intéresser. Elle publie très fréquemment des tweets relatifs à l'actualité sociale et politique, et retweete régulièrement des messages publiés par le compte du Front de Gauche ou de certains de ses cadres (dont notamment Jean-Luc Mélenchon) et de ses militants. Elle vote depuis une vingtaine d'années pour les partis et candidats situés à la gauche du PS. Elle n'a pourtant jamais pris sa carte au Parti de gauche ou

⁶³⁷ Terme désignant les utilisateurs de Twitter. Cf. glossaire, p. 464.

au PCF, car « [elle] ne se retrouve jamais complètement dans la ligne d'un parti, [elle] préfère garder sa liberté ».

On retrouve ainsi la « prétention à l'opinion personnelle », cette dimension de l'habitus petit-bourgeois que Pierre Bourdieu analysait dans *La Distinction* : « la revendication du droit à l'“opinion personnelle” et la défiance à l'égard de toutes les formes de délégation, surtout en politique, s'inscrivent logiquement dans le système des dispositions propres à des individus dont tout le passé et tout le projet reposent sur le pari du salut individuel, fondé sur les « dons » et les « mérites » personnels [et] sur la rupture avec des solidarités pesantes »⁶³⁸. La confrontation des propos de cette enquêtée (recueillis en situation d'entretien) avec l'observation directe de ses pratiques en ligne met au jour un paradoxe saisissant : si elle *déclare* bien se méfier de toute forme d'engagement « formel » au sein d'un parti politique, sa propension à retweeter fréquemment des messages publiés par le Front de gauche et Jean-Luc Mélenchon relève cependant d'une certaine forme de délégation politique. Elle-même avoue d'ailleurs être « très souvent d'accord avec Mélenchon et avec ce qu'il dit sur Twitter », et c'est à cette fin qu'elle relaie ses messages. Internet permettrait alors à ces agents de pratiquer une forme de militantisme sans le nommer : soutenir un parti ou un homme politique sur internet ne serait alors pas vécu comme une forme de *fides implicita* pour ces agents.

Les profils de type « militant » ont donc pour auteurs des individus qui connaissent des engagements « militants » différenciés (durables, distanciés, ou individuels). On peut dissocier alors deux sous-catégories de profils militants : le profil militant *professionnel* et le profil militant *personnel*. Les profils militants *professionnels* sont conçus avant tout comme un instrument de communication : leurs auteurs – le plus souvent des cadres militants et des élus – y communiquent des informations relatives à leur agenda politique ou à celui de leur organisation, le plus souvent à l'échelle locale : réunions, meetings, regroupements militants, interventions dans les médias, etc. Ils relaient également des tweets ou retweets vers des articles de presse relatifs à l'actualité du parti ou à divers événements censés intéresser les « abonnés » du profil. Les abonnés sont principalement des militants de section, que les auteurs connaissent réellement, mais aussi des sympathisants ou des militants d'autres sections, qu'ils ne connaissent pas alors nécessairement. On peut enfin observer sur ces profils que les prises de position du parti se substituent souvent à l'expression d'opinions « personnelles ».

⁶³⁸ Pierre Bourdieu, *La Distinction*, op. cit., p. 487.

Encadré 3. Un exemple de profil « militant professionnel »

G. est un militant frontiste de 24 ans de l'agglomération nantaise. Il crée en octobre 2009, en plus de son compte « personnel », le compte Twitter du Front National. Il constate à l'époque que les partis politiques – dont le FN – étaient très peu présents sur ce réseau social qui « *était encore marginal* » ; d'ailleurs « *personne au parti ne s'intéressait vraiment aux réseaux sociaux* ». Il a donc vu dans cette situation une opportunité à saisir : en créant de manière totalement autonome le compte du parti, il créait en même temps une fonction (ou une position) militante amenée à prendre de la valeur. Jusqu'en 2011, il « gère » donc le compte officiel du FN en parallèle de son compte « personnel », et ce de manière relativement indifférenciée. Les informations communiquées sont en effet très similaires d'un compte à l'autre. Il ne perçoit pas son compte Twitter comme un compte « personnel », mais comme un outil de communication au service de ses activités militantes et de celles de son parti. En 2011, « *la centrale du parti* » décide de « *reprendre la main* » sur l'ensemble des comptes (locaux ou nationaux) estampillés « Front National » et le dépossède alors de l'outil qu'il a créé. Il garde une expérience amère de cette dépossession et regrette la gestion de « *la centrale* » qui « *alimente tous les comptes et pages FN avec le même contenu* ». En reprenant la main sur la gestion du compte du parti, « *la centrale* » l'a donc écarté de la position qu'il s'était lui-même créée : « *écoutez c'est très étrange, moi je crée continûment pour le FN des espaces sur les réseaux sociaux, des sites internet, avant que le FN n'y pense, ce qui fait que ces contenus-là deviennent officiels quelques mois plus tard, voire quelques années plus tard pour Twitter, mais après c'est le FN qui capte ces structures-là, pour les gérer de manière centralisée, donc y a un peu un manque d'initiative du FN et encore un fois un manque de conscientisation vis-à-vis du fait c'est aussi important d'être présents sur les réseaux sociaux. Je sais pas comment ça fonctionne au FDG ou au PS mais je pense que c'est beaucoup plus professionnel* ». Cependant, s'il regrette que le FN n'ait pas suffisamment conscience des potentialités offertes par les réseaux sociaux, il gère paradoxalement son compte Twitter de manière relativement « classique ». Il y communique son agenda politique, publie des communiqués de presse, et « *live-tweets de temps en temps des meetings* ». Il ne donne jamais « son avis personnel » et veille à ne pas trop publier de tweets chaque jour car « *c'est pas très bon en terme d'image* ».

Le second type de profil « militant » est le profil *personnel*. Contrairement au profil *professionnel*, on y observe assez peu d'informations « officielles », relatives à l'agenda politique du parti par exemple. Les auteurs de ces profils ont alors davantage tendance à communiquer leurs goûts (plutôt que leurs opinions) à propos des personnalités et partis politiques qu'ils « soutiennent » mais aussi leurs dégoûts envers ceux qu'ils « combattent » (la connotation sportive est à peine exagérée). C'est donc un rapport à la politique

essentiellement émotif, voire passionnel, que l'on observe sur ces profils. Les auteurs de ces profils militants développent de nombreux échanges par tweets interposés avec d'autres « militants »/« supporters » (là aussi la connotation sportive semble particulièrement appropriée), mais aussi, bien que plus rarement, avec des « militants »/ « adversaires », avec qui les échanges sont souvent (verbalement) agressifs, voire violents. Du fait de ces échanges fréquents et répétés, les profils militants *personnels* produisent un nombre quotidien de tweets politiques relativement élevé.

Les auteurs de ces profils conçoivent donc la production de tweets politiques comme une activité militante à part entière. Rien ne le montre mieux que le cas des « *riposte parties* ». Les « *riposte parties* » sont des réunions de militants et de sympathisants (certains se réunissent physiquement au sein d'un local, d'autres « militent » depuis leur domicile, tout en restant en contact permanent avec les autres militants *via* internet) dont la fonction est de publier massivement des tweets afin de « soutenir » un candidat, notamment lors de débats télévisés. La production de tweets est alors appréhendée comme une compétition sportive où deux équipes s'affrontent : les « supporters » du candidat de gauche et les « supporters » du candidat de droite. Le but est donc de produire plus de tweets que le camp adverse, afin de démontrer que le candidat le plus tweeté est le plus soutenu sur les réseaux sociaux. Les « arbitres » de cette compétition sont les cabinets d' « étude web », qui mesurent en temps réel la volumétrie comparée des deux candidats sur Twitter et donnent à la fin du « match » le nom du « vainqueur ».

Nous avons eu l'occasion d'observer une « *riposte party* » au siège de campagne du candidat François Hollande lorsque celui-ci était invité à une émission politique télévisée diffusée à 20h50 sur TF1 quelques semaines avant le premier tour de l'élection présidentielle. Une vingtaine de militants et de sympathisants, chacun muni d'un ordinateur portable, étaient réunis dans une salle où était retransmise en direct l'émission télévisée. L'enjeu était alors de produire le plus grand nombre de tweets de soutien en faveur de François Hollande. Nous jouions le rôle d' « arbitres », en tant qu'analystes de la société Netscope. Présents avec eux dans la salle, nous leur communiquions chaque demi-heure la volumétrie et la tonalité des tweets publiés à propos de chaque candidat. À la fin de l'émission, lorsque nous leur avons annoncé que François Hollande avait généré le plus grand nombre de commentaires sur Twitter, avec le meilleur ratio de tweets « positifs », une véritable effusion de joie a alors retenti dans la salle, exactement comme à la fin d'un match consacrant la victoire de l'équipe soutenue par les supporters. Pour les militants, le véritable enjeu de cette « *riposte party* »

consistait à ce que les résultats de l'institut Netscope soient repris le plus possible dans les médias, afin que leur candidat apparaisse d'une part, « *aux yeux de l'opinion publique* » (pour reprendre les termes d'un militant), comme le plus populaire sur internet et que les sympathisants et militants socialistes apparaissent d'autre part comme les plus mobilisés.

Cette pratique des « *riposte parties* » apparaît donc comme une manifestation tout à fait significative des pratiques sociales générées par – et observables sur – les profils militants *personnels*. Contrairement aux profils militants *professionnels* qui mobilisent Twitter comme un instrument de communication, les auteurs des profils militants *personnels* conçoivent le réseau social comme un instrument d'action militante à part entière.

2.2.3 Le consommateur « ostentatoire » d'informations politiques

Enfin, le dernier type de profil identifié est celui du consommateur « ostentatoire » d'informations politiques. En publiant très régulièrement sur leur profil des tweets relayant des articles de presse, ces auteurs rendent en effet *ostensible* leur consommation d'informations politiques, selon une logique distinctive. Les auteurs de ces profils présentent donc de nombreuses similitudes avec les « *usagers en tâche de fond* » analysés par Jean-Baptiste Comby⁶³⁹. Ils développent ainsi un « *rapport permanent à l'information [...]* « favorisé par l'omniprésence dans leur environnement immédiat d'un ordinateur ou d'un smartphone » et prend la forme d'une véritable veille d'information *via* Twitter. Ces auteurs sont en effet abonnés à un très grand nombre de comptes de médias (nous proposerons une analyse statistique détaillée des logiques sociales d'abonnement à ces comptes d'information dans le chapitre 5) et leur *time-line*, qu'ils consultent plusieurs fois par jour, présentent alors une forme et une fonction équivalentes à celles du « bandeau » où s'actualisent les dépêches de l'AFP.

Mais la consommation d'informations par ces agents n'est pas uniquement « passive » et c'est là leur spécificité : en publiant très régulièrement des tweets contenant des liens (commentés ou non) vers des articles de presse ou des contenus médiatiques, et/ou en relayant des tweets publiés directement par les comptes Twitter des médias auxquels ils sont abonnés, ces auteurs font alors de leur consommation d'informations une pratique « ostentatoire ». Ils

⁶³⁹ Jean-Baptiste Comby *et al.*, « Les appropriations différenciées de l'information en ligne au sein des catégories sociales supérieures », *Réseaux*, 2011/6, n°170, p. 75-102.

dévoilent en effet volontairement et publiquement à l'ensemble de leurs abonnés les informations qu'ils ont lues, qui les ont intéressés, interpellés, ou qu'ils souhaitent « faire partager » à l'ensemble ou à certains de leurs abonnés.

Mais derrière cette volonté revendiquée de « partager » l'information (propos que l'on enregistre régulièrement dans les entretiens avec ce type d'auteurs) on peut en réalité percevoir une volonté de distinction. Consommer des informations « rares » constitue un exemple de pratique informationnelle distinctive. La « rareté » d'un bien informationnel en détermine précisément sa valeur symbolique.

C. est une retraitée de l'Éducation nationale de 59 ans qui publie très régulièrement des tweets « informationnels » relayant des articles sur des sites d'information « alternatifs », associatifs, proches de la gauche radicale et des mouvements des « sans ». Elle déclare vouloir partager avant tout des informations peu connues de ses abonnés :

[Comment vous sélectionnez les articles que vous partagez ensuite sur Twitter ?]
« Alors je sélectionne généralement beaucoup ce qui est social, je suis très très sensible à tous les problèmes de société, que ce soit le problème des immigrés, le problème des sans-papiers, le problème des Roms, le problème des gens qui sont en grande difficulté économique, les problèmes des sans-abris, et donc tout ce que peut publier le DAL par exemple, chômage-actif, des sites comme ça, qui décrivent la réalité. Ça ça m'intéresse beaucoup, je lis beaucoup, je le partage beaucoup, tout ce qui est du domaine social, je partage tout ce qui va dans cette direction-là... alors ce qui m'intéresse moins c'est les politiques genre les bagarres à l'UMP tout ces trucs-là, bon tout le monde en parle, j'ai pas besoin d'en rajouter, j'essaye de diffuser au contraire des informations que tout le monde n'a pas, particulièrement sur l'enfance, sur des infos qui ne passent pas dans les grands médias, ça ne m'intéresse pas par exemple de répercuter les disputes entre Fillon et Copé, mon opinion n'a rien d'original sur le sujet, je n'ai rien à en dire, je ne sais rien de plus, par contre je vais aller chercher sur Basta, qui est un site qui s'occupe beaucoup d'écologie, aller chercher des informations qui ne sont pas diffusées par les grands médias et de les partager, à mon avis c'est ça le sens d'une veille sur Twitter, de partager des choses qui ne sont pas répétées et répétées sur tous les tons ailleurs ».

Derrière sa volonté « citoyenne » de vouloir « diffuser » auprès de son réseau d'abonnés des informations que « tout le monde n'a pas », « qui ne passent pas dans les grands médias » à propos de « causes importantes », on peut en réalité percevoir une volonté

de se distinguer, de se démarquer de ce dont « *tout le monde parle* ». Les biens informationnels « rares » ont donc pour cette enquêtée une plus haute valeur symbolique et sociale. En publiciser la consommation, c'est faire savoir que l'on s'intéresse à des causes « nobles », dignes d'intérêt.

Mais cette volonté de distinction se manifeste également au travers d'autres pratiques sociales qui ont toutes pour point commun d'être empruntées au registre journalistique. Nous ne l'avons pas encore souligné, mais réaliser une veille d'information sur Twitter ou publier des articles sur son profil dans une optique de valorisation et de démarcation sont également des pratiques sociales exercées par les agents du champ journalistique, nous aurons là encore l'occasion de le démontrer dans un prochain chapitre. Mais on peut également observer sur ces profils de consommateurs de biens informationnels deux autres pratiques également empruntées aux journalistes : le « *fact checking* », et une pratique qui peut sembler plus anodine, singer les descriptions de profil des journalistes.

Le « *fact checking* » – ou la vérification par les faits en français – « *désigne une démarche visant à valider l'exactitude des chiffres rapportés, d'affirmations, d'informations, ou de faits allégués dans un texte ou discours* »⁶⁴⁰. Sur Twitter, cette pratique s'observe régulièrement lors de débats ou lors d'émissions politiques télévisées. Lorsque les internautes perçoivent une erreur dans le discours d'un homme politique, ils s'empressent alors de la vérifier sur internet et de la corriger en informant les autres utilisateurs. Pour A., un étudiant de 21 ans à l'IEP de Paris, le « *fact checking* » est une pratique relevant du « *devoir civique* » :

« J'aime bien casser les politiques, je le fais rarement, puisque de toute façon ils me répondront jamais, mais c'est vrai que parfois y a des énormités, à droite, à gauche, un peu partout, qui sont tellement énormes que je suis obligé de répondre, c'est un devoir civique pratiquement de corriger, peut-être le plus maintenant de Twitter, des réseaux sociaux c'est le fact checking, c'est tout de suite la vérification instantanée, les politiques se font généralement recadrer de manière très sèche, et ça leur fait du bien aussi, parce qu'à force de dire des conneries... »

Encore une fois, derrière cette justification normative de l'exercice d'un « *devoir civique* », il est facile de percevoir une pratique socialement distinctive : en corrigeant publiquement des erreurs factuelles prononcées par un homme politique, on fait la

⁶⁴⁰ http://fr.wikipedia.org/wiki/V%C3%A9rification_par_les_faits

démonstration (notamment envers ses pairs) que l'on n'est pas désarmé devant le langage politique, «*globalement situé, comme tout ce qui est symbolique, du côté des dominants* »⁶⁴¹ ; au contraire, non seulement on comprend ce langage, mais mieux, on le corrige. En pratiquant le « fact checking » ces individus entendent démontrer qu'ils se situent « au niveau », voire au-dessus, des hommes politiques, ou en tout cas qu'ils ne se situent pas, comme la majorité des citoyens, à la marge du champ politique.

Une seconde pratique distinctive, là encore empruntée aux journalistes, consiste à conclure sa description de profil par la phrase suivante : « *mes tweets n'engagent que moi* ». On observe en effet assez fréquemment cette mise en garde dans les descriptions de profil des journalistes inscrits sur Twitter. Avec cette phrase (qui n'a aucune valeur juridique) les journalistes entendent signifier que les tweets qu'ils publient sur leur profil personnel (et nominatif) n'engagent pas la responsabilité de leur journal (même si paradoxalement, la plupart citent dans cette même description le(s) média(s) au(x)quel(s) ils appartiennent). Les opinions qui sont exprimées sur leur profil le sont alors à titre personnel, et ne correspondent pas aux prises de positions de leur rédaction. Il n'est alors pas inintéressant de constater que des individus qui n'appartiennent pas au champ journalistique ont repris à leur compte cette mise en garde. Si certains singent ironiquement cette pratique, d'autres la reprennent « sérieusement », en entendant signifier à leur tour que leurs opinions n'engagent pas leur institution, ou leur employeur. Avec cette mise en garde, ces individus prétendent avoir des opinions qui leur sont propres et qui « n'engagent qu'eux ». En imitant cette pratique journalistique, ces auteurs affirment donc en creux leur « droit à l'opinion personnelle », cette dimension de l'habitus petit-bourgeois que nous avons évoquée plus haut. Cette pratique assez anodine est donc là encore une pratique sociale distinctive, propre aux profils de consommateurs ostentatoires d'informations politiques.

À travers cette typologie des producteurs de tweets politiques, nous avons donc souhaité démontrer que la place et la fonction assignées à la politique et à l'actualité par les membres de notre panel témoignaient de l'existence de rapports différenciés et parfois distinctifs à la politique.

⁶⁴¹ Pierre Bourdieu, *La Distinction, op. cit.*, p. 541.

Conclusion du chapitre 3

L'objectif de ce chapitre était double. Il s'agissait d'une part de livrer au lecteur une sociologie des producteurs de tweets politiques et de démontrer que cette population présente des propriétés très singulières, proches de celles des populations dites « répondantes » ou « participationnistes ». Les hommes, les Parisiens, les plus diplômés et les cadres et professions intellectuelles supérieures sont en effet surreprésentés au sein de notre panel. Pour reprendre les mots de Pierre Bourdieu, on peut donc clairement affirmer que Twitter constitue un nouvel observatoire du « pays légal ».

Le second objectif de ce chapitre était de démontrer que l'analyse qualitative des profils des producteurs de tweets politiques laisse cependant entrevoir certaines différences dispositionnelles.

Après avoir étudié la sociologie des producteurs de tweets politiques, nous allons, dans le prochain chapitre, analyser leur production de messages. On pourra alors se demander dans quelle mesure le volume de tweets politiques produits est déterminé par les propriétés sociologiques singulières de leurs producteurs.

CHAPITRE 4 – LES INÉGALITÉS DE PRODUCTION DE TWEETS POLITIQUES

Après avoir dressé dans le précédent chapitre une sociologie des individus qui ont commenté la campagne présidentielle du printemps 2012 sur Twitter, nous allons maintenant analyser leur production de messages politiques sur le temps long. Durant 11 mois, du 1^{er} mars 2013 au 31 janvier 2014⁶⁴², nous avons recensé et archivé, à l'aide de notre logiciel, l'exhaustivité⁶⁴³ des tweets publiés par les membres de nos panels.

Ainsi, sur cette période, les membres du panel « répondants » ont publié 840 251 tweets, dont 81 606 tweets contenant des occurrences politiques⁶⁴⁴, soit 9,7 % des tweets publiés ; les « non répondants » ont quant à eux publié 1 120 896 tweets, dont 42 940 tweets politiques, soit 3,8 % des tweets publiés.

Au-delà de cette différence volumétrique remarquable entre les deux panels, qui, nous le verrons, rend compte du caractère discriminant de notre questionnaire, l'analyse statistique met au jour un second résultat, commun quant à lui aux deux panels : l'existence d'une forte dispersion statistique du nombre de tweets politiques publiés.

Cette dispersion est observable à la fois à un niveau *individuel et synchronique* et à un niveau *collectif et diachronique*. À un niveau individuel et synchronique, c'est-à-dire si l'on analyse le volume *total* de tweets publiés sur la période, on observe une forte dispersion du nombre de tweets politiques publiés par individu. 3 % des membres du panel « répondants » (20 individus) ont ainsi produit à eux seuls plus de la moitié des tweets politiques publiés sur la période. Nous chercherons alors à établir dans quelle mesure les variables sociologiques (âge, sexe, PCS, etc.) et politiques (niveau d'intérêt pour la politique, militantisme, etc.) *individuelles* déterminent le nombre de tweets politiques publiés.

Mais cette forte dispersion s'observe également à un niveau collectif et diachronique. En effet, si l'on analyse la production des panels dans le temps, avec comme unité de mesure

⁶⁴² En raison de problèmes techniques rencontrés par notre logiciel de collecte, la phase d'observation des panels a été retardée de 7 mois. Initialement prévu pour le mois de septembre 2012, l'enregistrement exhaustif et systématique des tweets n'a pu finalement débiter que le 1^{er} mars 2013.

⁶⁴³ Des contrôles manuels répétés de l'activité de collecte nous laissent supposer qu'elle a en effet été réellement exhaustive. Cependant, nous ne pouvons totalement écarter le risque de défauts d'indexation épisodiques.

⁶⁴⁴ Nous définirons précisément ces occurrences dans la suite de ce chapitre.

le nombre *quotidien* de tweets politiques publiés par l'ensemble des producteurs, on observe alors des différences volumétriques très marquées d'un jour à l'autre.

Pour tenter d'expliquer ces différences de production dans le temps, nous mobiliserons des hypothèses de nature davantage *conjoncturelle*. L'analyse statistique de la production de nos panels permettra ainsi de vérifier empiriquement l'hypothèse formulée par Daniel Gaxie selon laquelle « *plus un problème est débattu au sein des espaces de débat politique, plus ces débats sont répercutés au-delà de ces espaces, plus s'accroît le nombre des agents qui se prononcent « politiquement » [...] »*⁶⁴⁵.

L'objectif de ce chapitre sera donc de mettre au jour les facteurs *dispositionnels* et *conjoncturels* qui président à la publication de tweets politiques. Si la sociologie politique a depuis longtemps identifié les facteurs *dispositionnels* de production des opinions politiques – en analysant notamment les inégalités de réponse aux questions politiques – elle a plus rarement cherché à objectiver les facteurs *conjoncturels* qui participent à l'activation de cette production. Les facteurs contextuels sont en réalité un point aveugle de la théorie des sondages. Les sondages tiennent en fait pour résolue la question du contexte de production des opinions de la même manière qu'ils tiennent pour résolue la question des inégalités de compétence politique : en postulant *a priori* leur inexistence. Tous les individus sont censés avoir des opinions politiques sur tous les sujets, *tout le temps*. Et si, en mettant au jour l'effet d'imposition de problématique, la sociologie critique a démontré que tous les individus n'avaient pas en réalité des opinions constituées sur toutes les questions politiques, elle s'est en revanche peu interrogée sur le fait que ces impositions de problématique pouvaient produire des effets différenciés dans le temps.

En ne reposant pas sur une démarche de questionnement, l'enregistrement passif des opinions politiques sur Twitter rend alors davantage saillante cette question des facteurs *conjoncturels*. Pourquoi enregistre-t-on certains jours davantage de messages politiques ? Ces inégalités de publication dans le temps concernent-elles tous les individus, toutes les classes sociales ? Ces inégalités sont-elles déterminées par l'activité du champ politique ou médiatique ? Ces questions sont alors déterminantes si l'on souhaite faire de Twitter un observatoire des opinions politiques.

⁶⁴⁵ Daniel Gaxie, « Retour sur les modes de production des opinions politiques », in Philippe Coulangeon et Julien Duval (dir.) *Trente ans après La Distinction de Pierre Bourdieu*, Paris, La Découverte, 2013.

Dans une première section de chapitre, nous comparerons les volumétries des deux panels et tenterons d'objectiver les facteurs individuels de production de tweets politiques (1). Puis nous verrons dans une deuxième section que la production des panels est très inégale dans le temps, et que cette inégalité concerne toutes les catégories de producteurs (2). Enfin, nous mettrons en relation le volume de production de nos panels représentatifs avec le volume de production de comptes d'agents et d'institutions appartenant aux champs médiatiques et politiques, et tenterons alors de démontrer que l'activité de production de ces champs – l'« offre » d'opinion – détermine pour partie le niveau de la « demande » (3).

1. Les facteurs individuels de production de tweets politiques

Avant de présenter plus en détail les variables sociologiques et politiques qui déterminent le volume de production *individuel* de tweets politiques (1.2), nous allons comparer les volumétries des panels « répondants » et « non répondants » (1.1).

1.1 Comparaison volumétrique de la production des panels « répondants » et « non répondants »

Les individus qui composent les panels « répondants » et « non répondants » ont tous pour point commun d'avoir publié entre le 1^{er} et le 31 mars 2012 au minimum un tweet citant un ou plusieurs candidat(s) qui s'est/se sont présenté(s) au premier tour de l'élection présidentielle française du printemps 2012. Nous avons alors dressé dans le précédent chapitre une sociologie de ces individus à ce moment *t*, sans tenir compte de leur fréquence de publication, passée ou future.

Ce n'est qu'à partir du 1^{er} mars 2013, soit tout juste un an après qu'ils aient commenté l'élection présidentielle et aient été déclarés éligibles au panel, que nous avons pu commencer à enregistrer de manière exhaustive et systématique les tweets publiés par ces individus. Nous avons initialement prévu de démarrer cette phase de collecte le 1^{er} septembre 2012, date à laquelle avaient pris fin la constitution des deux panels et leur programmation dans notre

logiciel. Cependant, en raison de problèmes techniques répétés⁶⁴⁶, nous avons dû finalement retarder le lancement de cette phase de sept mois, et débiter la collecte le 1^{er} mars 2013. Nous avons donc finalement pu archiver et analyser les tweets publiés par les panels « répondants » et « non répondants » durant 11 mois, du 1^{er} mars 2013 au 31 janvier 2014.

Au 1^{er} mars 2013, 8 % des individus du panel « répondants » (soit 50 individus sur 608) avaient supprimé, désactivé ou « laissé à l'abandon »⁶⁴⁷ leur compte Twitter. À cette même date, le « taux de mortalité » du panel « non répondants » était quant à lui deux fois plus élevé : 17 % des comptes (103 comptes sur 620) étaient, un an après leur entrée dans le panel, inexistants ou inactifs. Ce premier résultat n'est pas neutre et met au jour une première différence entre nos deux panels. Les individus qui ont répondu à notre questionnaire témoignent semble-t-il d'un rapport plus investi, plus durable à leur compte Twitter. Mais c'est en analysant statistiquement le nombre de tweets et le nombre de tweets *politiques* publiés que les différences entre les deux panels se font les plus manifestes.

Le nombre de tweets publiés par un individu est une donnée objective, non construite par le chercheur. Dès lors qu'un individu publie un tweet sur son compte, l'information est délivrée automatiquement par Twitter et est enregistrée par notre logiciel. En revanche, le nombre de tweets *politiques* publiés est une donnée construite arbitrairement. C'est à partir d'une recherche par mots-clés, effectuée sur la base totale de tweets archivés par le logiciel, que nous avons pu comptabiliser le nombre de tweets *politiques* publiés. Le choix des mots-clés retenus, et plus largement notre définition de la politique, sont donc déterminants et doivent être ici clairement explicités.

Nous avons fait le choix d'adopter une définition volontairement restreinte de la politique, réduite essentiellement à sa dimension *politics* (à savoir ce qui se réfère à « l'activité politique, c'est-à-dire l'engagement politique et l'exercice de la profession politique »⁶⁴⁸). La recherche par mots-clés est composée de 121 termes politiques différents. Elle comprend des noms propres de femmes et d'hommes politiques (« Hollande », « Valls », « Copé », « Aubry », etc.), des titres de fonction (« ministre », « député », « président », etc.), des noms et des sigles de partis politiques (« eelv », « ump », « ps », « parti socialiste », etc.),

⁶⁴⁶ Le logiciel ne parvenait pas à enregistrer de manière exhaustive et systématique les tweets de l'ensemble des sources du panel.

⁶⁴⁷ Nous qualifions de compte « abandonné » un compte qui n'a publié aucun tweet (politique ou non) durant l'ensemble de la période d'observation, soit du 1^{er} mars 2013 au 31 janvier 2014.

⁶⁴⁸ Jean-Yves Dormagen et Daniel Mouchard, *Introduction à la sociologie politique*, De Boeck, 2010, 3^{ème} édition.

des noms d'institutions politiques (« sénat », « gouvernement », « assemblée », etc. ») ainsi qu'une série de termes relatifs aux idées politiques (« la droite », « république », « démocratie », « communisme », etc.). Précisons que pour être socialement la moins discriminante possible, cette liste de mots-clés comprend également des noms propres mal orthographiés (« holande », « sarkosy », « mélanchon », etc.)⁶⁴⁹.

Les 121 termes retenus ont été définis après une longue phase d'expérimentation. Nous avons en effet vérifié de manière inductive pour chaque terme retenu le nombre de tweets correspondant dans notre base de données. Nous ne pouvons pas prétendre que cette recherche par mots-clés épuise le champ des possibles. Il est tout à fait probable que des tweets issus de notre base de données contiennent des occurrences politiques qui ne figurent pas dans notre liste de mots-clés, et que ces tweets ne soient donc pas comptabilisés comme « politiques ». Cependant, on peut supposer que des tweets se référant à des acteurs politiques qui ne bénéficient pas d'une notoriété suffisante pour figurer dans notre liste sont d'une part peu nombreux, et sont d'autre part publiés par des auteurs particulièrement politisés, et qui ont donc toutes les chances d'avoir publié beaucoup d'autres tweets politiques.

Nous avons alors effectué cette recherche par mots-clés politiques sur nos deux bases de données ; la première est composée de la totalité des tweets publiés par le panel « répondants », la seconde comprend l'ensemble des tweets issus du panel « non répondants ». Le premier résultat qui peut être ici évoqué concerne précisément la taille différente de ces deux bases de données. Avec un nombre moyen de 2 973 tweets publiés par jour (soit une moyenne de 6 tweets publiés par jour et par individu), le panel « non répondants » a publié sur la période un nombre total de tweets supérieur de 25% à celui du panel « répondants » (2 229 tweets par jour, soit 4 tweets par individu et par jour en moyenne).

Ce premier résultat peut donc laisser supposer que les individus qui n'ont pas répondu à notre questionnaire publient davantage de tweets que les individus qui ont accepté de répondre.

⁶⁴⁹ Pour connaître l'intégralité de cette liste, Cf. Annexe 7, p. 495.

Tableau 4.1 Comparaison statistique de la production des deux panels

	Tweets totaux		Tweets politiques	
	Répondants	Non rép.	Répondants	Non rép.
<i>n</i> =	558	517	558	517
Effectif	840 251	1 120 896	81 606	42 940
Minimum	1	1	0	0
Maximum	18 342	79 236	8128	5 689
Moyenne	1 506	2 168	146	83
Médiane	355	322	14	8
Quartile 1	72	43	3	1
Quartile 3	1 521	1 910	74	35
Ecart type	2 815	5 526	569	350
Variance	3 0534 469	7 924 952	323 282	122 617

Cependant, on constate à la lecture de ce tableau que le *nombre médian* de tweets publiés par le panel « répondants » (355 tweets) est légèrement plus élevé que le *nombre médian* de tweets publiés par le panel « non répondants » (322 tweets). En réalité, le *nombre moyen* élevé de tweets publiés par le panel « non répondants » masque une plus grande dispersion statistique. L'écart-type du nombre de tweets publiés par les « non répondants » est ainsi presque deux fois plus élevé que celui des « répondants ». On retrouve donc plus souvent dans le panel « non répondants » des valeurs extrêmes, à savoir des individus qui publient très peu de tweets (le premier quartile est de 43 tweets pour les « non répondants », contre 72 pour les « répondants ») et à l'inverse des individus qui publient un nombre considérable de messages (le troisième quartile est de 1910 tweets pour les individus qui n'ont pas répondu, contre 1521 pour ceux qui ont répondu).

Les plus gros producteurs de tweets du panel « non répondants » présentent un profil relativement proche de celui des « publicateurs de post-it » évoqués dans le précédent chapitre : des individus, souvent adolescents, qui publient tout au long de la journée, *via* notamment leur téléphone mobile, des tweets afin de communiquer à leurs amis leur humeur, leur activité du moment, leurs commentaires sur des programmes télévisés et des retranscriptions sportives, etc. Plusieurs indices laissent suggérer que les titulaires de ces comptes proviennent d'un milieu social plus modeste : préférences et goûts culturels affichés, langage argotique, etc. Cependant, rappelons que ces individus n'ont pas répondu à notre questionnaire, et que l'on ne dispose donc pas de données déclaratives permettant d'objectiver clairement leur milieu social d'appartenance.

Ces « publicateurs de post-it » ont donc moins répondu à notre questionnaire que d'autres catégories d'individus (sans doute car beaucoup n'ont pas encore le droit de vote et ne se sont pas sentis concernés par notre étude ou autorisés à répondre, ou encore parce qu'il est probable que venant d'un milieu plus modeste, ces individus s'intéressent moins à la politique et soient moins réceptifs à la démarche d'un étudiant). C'est donc pour cette raison que l'on retrouve moins de valeurs extrêmes (supérieures) – des très gros producteurs de tweets – au sein du panel « répondants ».

Les individus qui ne publient que très occasionnellement des tweets ont également moins répondu à notre questionnaire, sans doute en raison d'un rapport moins investi au réseau social. Il est ainsi fort probable que les individus qui publient très peu de tweets se connectent également peu souvent au réseau social et n'aient donc pas nécessairement pris connaissance de notre invitation à consulter notre profil et à répondre à notre questionnaire. Cette moins forte propension à répondre à notre enquête expliquerait alors la plus forte dispersion du nombre de tweets publiés par les « non répondants » (en raison d'une plus forte présence de variables extrêmes « inférieures » et « supérieures »).

On retrouve donc en résumé au sein du panel « non répondants » des individus qui tweetent très peu et au contraire d'autres qui tweetent énormément. C'est donc pour cette raison qu'à un niveau agrégé, le panel « non répondants » publie en moyenne davantage de tweets que le panel « répondants ».

Si l'on s'intéresse maintenant au nombre de tweets *politiques* publiés par les deux panels, on observe alors, et c'est là un résultat tout à fait remarquable, le phénomène inverse. Le panel « répondants » a ainsi publié un nombre de tweets politiques pratiquement deux fois plus élevé (en valeur absolue) que le panel « non répondants ». Les individus « répondants » ont en effet publié sur la période 242 tweets politiques par jour en moyenne contre seulement 127 pour les individus « non répondants ».

Les individus qui ont répondu à notre questionnaire publient donc en moyenne moins de tweets que ceux qui n'ont pas répondu, mais publient (en valeur absolue) deux fois plus de tweets politiques. Le nombre médian de tweets politiques publiés, mais également le premier et le troisième quartile, sont également significativement plus élevés parmi les « répondants ».

Et si, contrairement à ce que l'on a observé avec le nombre total de tweets publiés, on constate une plus grande dispersion statistique du nombre de tweets *politiques* cette fois parmi

les « répondants », les facteurs explicatifs de cette dispersion ne sont pas les mêmes. La plus grande dispersion du nombre de tweets politiques publiés par les « répondants » s'explique ainsi uniquement par la présence de valeurs extrêmes « supérieures », et non par la présence simultanée de valeurs extrêmes « inférieures » et « supérieures », comme dans le cas précédemment évoqué du nombre de tweets publiés par les « non répondants ».

En clair, le panel « répondant » comprend davantage de très gros producteurs de tweets politiques et au contraire assez peu de faibles producteurs. À l'inverse, le panel « non répondants » comprend beaucoup de très petits producteurs mais relativement peu de gros producteurs. C'est donc pour cette raison que la dispersion statistique du nombre de tweets politiques publiés est moins importante parmi les « non répondants ».

Il semble donc que le questionnaire que nous avons administré à notre population de producteurs de tweets politiques ait produit un effet particulièrement discriminant. Tout se passe comme si les individus qui publient régulièrement des tweets politiques avaient une plus forte propension à répondre à notre questionnaire. Rappelons que pour répondre à notre questionnaire les individus devaient au préalable s'apercevoir que nous nous étions abonnés à leur compte, puis consulter notre profil et lire notre description de compte (« *Doctorant en science politique, je cherche à dresser un portrait sociologique des réseaux sociaux. Merci de m'aider en répondant à ce questionnaire !* ») qui était suivie du lien URL de la page internet hébergeant le questionnaire. Cette description de profil, que nous pensions alors relativement « neutre », ne l'était en fait pas vraiment. Les termes « doctorant », « science politique », « portrait sociologique », qui nous apparaissent de prime abord – et de manière relativement ethnocentrique – comme courants voire « banals », peuvent être en réalité perçus par certains agents sociaux comme relativement « ésotériques », complexes, ou en tout cas étrangers à leur univers sémantique et donc sociologique. Tout se passe donc comme si les individus qui publient très peu de tweets politiques ne se sentaient pas concernés ou pas autorisés à répondre à un questionnaire d'« un doctorant en science politique ».

Nous avons pourtant vu dans le précédent chapitre qu'il existait relativement peu de différences sociologiques entre les individus qui avaient répondu à notre questionnaire et ceux qui n'avaient pas répondu. Cependant, il est important de rappeler que nous avons pu établir ce constat uniquement à partir des informations socio-démographiques éparses que les « non répondants » indiquent parfois dans leur description de profil ou sur d'autres sites internet et réseaux sociaux. Or tous les individus n'indiquent pas nécessairement ces informations sur

leur profil. Pour rappel, la PCS des individus qui n'ont pas répondu à notre questionnaire a pu être identifiée dans 53 % des cas, le lieu de résidence dans 60 % et l'âge dans seulement 36 % des cas. Nous ne connaissons donc pas le milieu social de presque un « non répondant » sur deux. Si les « non répondants » dont la PCS a pu être identifiée présentent un profil proche des « répondants », rien n'indique en revanche que les « non répondants » dont la PCS n'a pas pu être identifiée présentent un profil social similaire. Nous l'avons dit, de nombreux indices (goûts et préférences culturelles affichés, vocabulaire, etc.) suggèrent le contraire. Comparer les quartiles du nombre de tweets politiques publiés au sein de ces trois populations (les « non répondants » dont la PCS a pu être identifiée, les « non répondants » dont la PCS n'a pas pu être identifiée et les « répondants ») peut alors apporter un élément de preuve supplémentaire, cette fois objectif, de ces différences entre « non répondants ».

Tableau 4.2 Quartiles de la variable « nombre de tweets politiques » selon la population étudiée

Quartiles Nb. De tweets politiques publiés	Non répondants PCS connue (n = 297)	Non répondants PCS inconnue (n = 220)	Non répondants Total (n = 517)	Répondants (n = 558)
Q1	2	1	1	3
Médiane	11	5	8	14
Q3	38	31	35	74

La comparaison du nombre médian de tweets politiques publiés par les « non répondants » dont la PCS a pu être identifiée avec le nombre médian de tweets politiques publiés par les « non répondants » dont la PCS n'a pas pu être identifiée confirme l'hypothèse selon laquelle cette seconde sous-population de « non répondants » présente un profil différent. On constate en effet que le nombre médian de tweets politiques publiés par les « non répondants » dont on connaît la PCS se rapproche davantage de celui des « répondants ». Le nombre médian de tweets publiés par les « non répondants » dont la PCS n'a pas pu être identifiée est quant à lui presque trois fois inférieur à celui des répondants.

Cependant, qu'ils aient indiqué ou non leur profession, on retrouve parmi les « non répondants » assez peu de gros producteurs de tweets politiques, comparativement aux

« répondants ». Contrairement aux deux premiers quartiles, le troisième quartile du nombre de tweets politiques publiés est presque identique au sein de ces deux sous-populations de « non répondants » (respectivement 38 et 31 tweets) et est près de deux fois inférieur à celui des répondants (74 tweets). Les plus gros producteurs de tweets politiques, qui commentent très régulièrement, voire quotidiennement, la politique sur Twitter, ont donc, semblent-ils, majoritairement répondu à notre questionnaire.

Malgré ces différences que l'on peut enregistrer entre nos deux panels, l'amplitude du nombre de tweets politiques publiés est, au sein de notre population totale de producteurs de tweets politiques, très importante. Parmi les « répondants » comme les « non répondants », il existe en effet de très importantes inégalités individuelles en termes de production de tweets politiques.

Nous allons donc maintenant tenter d'identifier les facteurs sociologiques et politiques à l'origine de ces inégalités de publication.

1.2 Les facteurs individuels de production de tweets politiques

Nous avons vu dans le précédent chapitre que les producteurs de tweets politiques – à savoir des individus inscrits sur Twitter qui ont publié durant la séquence présidentielle un ou plusieurs tweets citant le nom des candidats – présentent des propriétés sociologiques et politiques tout à fait singulières et relativement élevées par rapport à la moyenne de la population française. Nous avons néanmoins tenté de démontrer, notamment à partir d'observations qualitatives, qu'il existait certaines différences significatives entre ces individus. Vérifier l'existence de corrélations statistiques entre le nombre de tweets politiques publiés et les variables sociologiques (âge, sexe, PCS, niveau d'études, etc.) et politiques (niveau d'intérêt déclaré pour la politique, pratiques militantes, orientation politique, etc.) de leurs auteurs est une autre façon d'objectiver ces différences.

Afin d'identifier les facteurs sociologiques et politiques qui déterminent le nombre de tweets politiques publiés par un individu (variable dépendante), nous avons réalisé une analyse de régression linéaire.

Les variables indépendantes que nous avons testées sont principalement celles que nous avons mobilisées pour décrire notre population dans le chapitre précédent, et plus

généralement celles mobilisées par les travaux de sociologie politique pour décrire des inégalités de politisation ou de participation politique : le sexe, l'âge, le niveau de diplôme, la PCS, la PCS du père, le lieu d'habitation, le niveau d'intérêt déclaré pour la politique et le fait d'adhérer ou non à un parti politique. Mais nous avons également testé cinq autres variables (continues) issues de l'observation de nos panels : le nombre total de tweets publiés, le nombre d'abonnés, le nombre d'abonnements, dont le nombre d'abonnements à un homme politique et le nombre d'abonnements à un média « politique »⁶⁵⁰. Contrairement aux autres variables testées, ces cinq variables sont spécifiques à notre population et postuler *a priori*, comme nous le faisons ici, qu'elles puissent déterminer le nombre de tweets politiques publiés par un individu reposent sur des hypothèses qui méritent d'être davantage explicitées.

La première variable, le nombre total de tweets publiés, correspond nous l'avons vu à l'ensemble des tweets, politiques ou non, publiés par un individu. Postuler que le nombre *total* de tweets détermine le nombre de tweets *politiques* publiés par un individu peut paraître tautologique de prime abord. Le nombre de tweets politiques est nécessairement inclus dans le nombre total de tweets publiés par un individu. Publier un nombre x de tweets politiques suppose en effet de publier un nombre y de tweets au minimum égal à x . Cependant, la réciproque n'est pas vraie, nous l'avons vu. Publier un grand nombre de tweets n'implique pas nécessairement de publier un grand nombre de tweets *politiques*. On pourra alors se demander dans quelle mesure le nombre total de tweets publiés par un individu sur une période de temps donnée détermine le nombre de tweets *politiques* publiés sur cette même période.

La deuxième variable-test directement issue de nos panels est le nombre d'abonnés d'un individu. Un utilisateur sur Twitter peut être « suivi » par d'autres utilisateurs⁶⁵¹. Ces « abonnés » peuvent être des individus que le titulaire du compte connaît personnellement (amis, membres de la famille, camarades de classe, collègues de travail, etc.). C'est alors l'identité de l'auteur du compte qui suscite la demande d'abonnement et non sa production de tweets (on s'abonne à un individu car c'est un ami, le contenu, la qualité et la fréquence des tweets qu'il publie important peu). Mais un individu peut également s'abonner au compte d'un individu qu'il ne connaît pas. Et dans ce cas, le contenu, la qualité et la fréquence de

⁶⁵⁰ Par « média politique », nous entendons un média (presse écrite ou internet, télévisée et radio) qui publie régulièrement des contenus politiques : *Le Monde*, France 2, France Info, *Rue 89*, *Les Inrockuptibles*, etc.

⁶⁵¹ Cf. introduction.

production du compte peuvent alors avoir une importance, nous avons eu l'occasion de le percevoir lors d'entretiens réalisés avec les membres de nos panels.

[Et toi ça t'arrive de suivre des gens que tu ne connais pas ?] « *Ça m'arrive mais faut vraiment que je trouve un truc très percutant. Si les tweets ou les photos du gars sont nuls, ça m'emmerde vite, ça pollue mon mur et je unfollow [me désabonne] très vite* » (A., un étudiant de 20 ans à l'I.E.P. de Paris).

On peut alors supposer qu'avoir un nombre élevé d'abonnés témoigne d'une part d'un rapport investi au réseau social (présence régulière et active, publication fréquente de tweets, échange avec les autres utilisateurs, etc.). Il est ainsi peu probable qu'un individu consultant très peu son compte Twitter et publiant très peu de tweets soit suivi par un nombre important d'abonnés. Mais on peut également considérer le nombre d'abonnés à un compte comme un indice permettant de mesurer le niveau de capital social d'un individu sur Twitter. Il conviendrait pour être plus précis de dissocier le capital social *réel* (nombre d'individus abonnés que le titulaire du compte connaît « réellement ») et le capital social *virtuel* (abonnés que le titulaire du compte ne connaît pas). Nous n'avons néanmoins pas eu les moyens d'opérer cette distinction. On pourra toutefois se demander si un individu publie davantage de tweets politiques lorsqu'il se sait « suivi » par un nombre important d'abonnés ou à l'inverse si un nombre élevé d'abonnés est la conséquence d'une production importante de tweets.

La troisième variable-test issue de nos panels est le nombre d'abonnements, soit le nombre de comptes auxquels un individu est lui-même abonné. De manière symétrique à ce qui vient d'être dit, un individu peut lui aussi s'abonner à d'autres comptes d'individus (chaque compte Twitter a un nombre x d'abonnés et un nombre y d'abonnements). En effet, nos enquêtés peuvent eux-aussi s'abonner à des comptes d'autres utilisateurs, qu'ils connaissent ou non dans le monde « réel ». Mais ils peuvent également s'abonner à des comptes de médias, d'institutions, de personnalités médiatiques, artistiques, politiques, etc. Contrairement au nombre d'abonnés, nous avons pu distinguer et objectiver ces différentes catégories d'abonnements. Nous avons ainsi isolé les abonnements à des médias politiques⁶⁵² et les abonnements à des hommes politiques.

⁶⁵² Nous avons comptabilisé manuellement pour chacun des 1228 individus panélisés le nombre de médias politiques et le nombre de femmes ou d'hommes politiques auxquels ils étaient abonnés. Pour plus de détails méthodologiques, cf. chapitre 5.

Le nombre d'abonnements à ces comptes peut être alors appréhendé comme un indice mesurant un degré d'exposition et de consommation médiatique mais aussi, particulièrement dans le cas des abonnements à des comptes de femmes ou d'hommes politiques, comme un autre indice permettant de mesurer l'intérêt pour la politique (un intérêt objectivé et non déclaré). On peut en effet supposer que plus un individu est abonné à des comptes de médias et d'hommes politiques, plus il témoigne d'un intérêt et d'une connaissance élevée de la politique, et partant, plus il sera enclin à publier des tweets politiques. Étudier l'existence de corrélations statistiques entre ces variables-test et le nombre de tweets politiques publiés par individu permettra alors de vérifier empiriquement ces hypothèses.

Avant de réaliser notre analyse de régression, nous avons opéré une série de tests statistiques, à l'aide du logiciel SPSS, afin de mesurer une à une la significativité des variables testées (avec principalement des analyses de la variance – ou ANOVA – lorsque les variables explicatives sont nominales, des tests de corrélation de Spearman lorsqu'elles sont ordinales et des tests de corrélation de Pearson, lorsque les variables-test sont quantitatives).

Significativité des variables sociologiques

Parmi les sept variables sociologiques testées (sexe, âge, PCS, PCS du père, lieu d'habitation et niveau de diplôme), seules deux sont significativement corrélées au nombre de tweets politiques publiés : l'âge et la PCS. Avant de développer ces résultats, il convient de rappeler que la population sur laquelle nous avons effectué ces tests statistiques (les « répondants » dont le compte était toujours actif au 1^{er} mars 2013 ; $n = 558$) présente déjà des propriétés sociologiques très singulières, que nous avons détaillées dans le précédent chapitre. Ainsi, affirmer qu'il n'existe pas de corrélation significative entre le sexe et le nombre de tweets politiques publiés n'invalide pas pour autant le résultat selon lequel il existe une corrélation entre le sexe et le fait de produire – ou de ne pas produire – des tweets politiques. On sait ainsi par exemple que les femmes sont statistiquement sous-représentées au sein de notre population de producteurs de tweets politiques par rapport à la population internautes française. En revanche, les femmes qui font partie de notre échantillon ne présentent pas de différence significative avec les hommes en termes de volume de tweets politiques publiés. Dit autrement, si les femmes sont proportionnellement moins nombreuses que les hommes à produire des tweets politiques, celles qui en produisent le font dans les mêmes proportions que les hommes. On peut ainsi faire le même constat concernant le niveau de diplôme,

l'origine sociale (mesurée par la PCS du père) ou encore le lieu d'habitation (Paris/province ou selon la taille de la commune habitée).

En revanche, on enregistre une corrélation significative (de faible intensité) entre l'âge des producteurs et le nombre de tweets politiques publiés ($r = 0,235$, $p = 0,00$).

Tableau 4.3 Test de corrélation de Pearson entre l'âge et le nombre de tweets politiques publiés par les « répondants »

Variabes	$n =$	r	valeur de p	$p < 0,05 =$ *
Nombre de tweets politiques	558	0,235	0,00	*
Âge				

Le test de corrélation de Pearson ($r = 0,235$, $p = 0,00$) indique en effet une association linéaire positive entre ces deux variables. En clair, le nombre de tweets politiques publiés croît en fonction de l'âge. Plus un individu est âgé, plus il a tendance à publier un nombre élevé de tweets politiques. Si nous avons vu que les catégories les plus âgées de la population (dont notamment les plus de 61 ans) étaient nettement sous représentées au sein de notre population (cette classe d'âge représente 4 % des « répondants »), elles sont pourtant celles qui produisent en moyenne le plus de tweets politiques. Dit autrement, les plus de 61 ans sont proportionnellement les moins nombreux à produire des tweets politiques, mais la minorité qui en produit est celle qui en produit le plus en moyenne.

On sait cependant que le test de corrélation de Pearson est extrêmement sensible à la présence de valeurs extrêmes. Or nous avons vu que le nombre de tweets politiques publiés comprenait justement de nombreuses valeurs extrêmes. On peut alors se demander si ce ne sont pas parmi les catégories les plus âgées de notre population que l'on retrouve précisément ces valeurs extrêmes.

L'analyse statistique valide en effet cette hypothèse. Si les catégories les plus âgées de notre population sont celles qui publient en moyenne le plus de tweets politiques, elles sont également celles qui comprennent le plus de valeurs extrêmes, à savoir les plus gros producteurs de tweets politiques. La classe d'âge des plus de 61 ans est ainsi celle qui publie en moyenne le plus de tweets politiques (921 tweets contre 146 tweets pour l'ensemble des

« répondants ») mais également celle qui connaît le plus important écart type (2044 contre 569).

Tableau 4.4 Indices de dispersion et de tendance centrale du nombre de tweets politiques par classes d'âge

	n =	Nb tweets	Moyenne	Médiane	Ecart-type
Moins de 18 ans	43	1089	25	14	34
18 à 25 ans	230	17990	78	13	188
26 à 34 ans	108	8098	75	10	142
35 à 45 ans	103	22271	216	19	691
46 à 60 ans	52	11886	229	48	579
61 ans et plus	22	20272	921	30	2044
Total	558	81606	146	14	569

C'est donc parmi les catégories les plus âgées de notre population que l'on retrouve les très gros producteurs de tweets politiques. Ces individus semblent réunir simultanément trois propriétés remarquables : un âge relativement avancé (par rapport à l'âge moyen des internautes, qui était de 41 ans en 2012⁶⁵³), un fort niveau de politisation et un important capital technique numérique⁶⁵⁴. Les plus de 61 ans sont ainsi ceux qui déclarent le plus s'intéresser « beaucoup » à la politique (86 % contre 37 % des moins de 18 ans et 56 % des « répondants » en moyenne). Et si nous n'avons pas mesuré (par déclaration) le niveau de capital technique des enquêtés, l'observation de leurs pratiques numériques et les propos réunis en entretiens d'enquête semblent confirmer un sens pratique très développé en matière informatique (et c'est sans doute au sein de cette classe d'âge que l'autodidaxie informatique est la plus forte).

Au sein des catégories les plus jeunes de notre panel « répondants », on observe une nette différence entre les moins de 18 ans (qui n'ont pas le droit de vote, mais qui ont publié durant la séquence présidentielle au moins un tweet politique) et les 18-34 ans. Les moins de 18 ans ont publié en moyenne trois fois moins de tweets politiques (25 tweets sur la période,

⁶⁵³ Vincent Gombault, « L'internet de plus en plus prisé, l'internaute de plus en plus mobile », *INSEE Première*, n°1452, juin 2013.

⁶⁵⁴ Pour une définition du capital technique numérique, cf. chapitre 3.

soit un peu plus de deux tweets politiques par mois), que les 18-34 ans (77 tweets sur la période, soit 7 tweets politiques par mois). Et c'est parmi cette classe d'âge que le volume de publication est le plus homogène : l'écart-type du nombre de tweets politiques publiés n'est en effet que de 34 tweets parmi les moins de 18 ans, contre 569 à l'échelle du panel).

En résumé, plus une classe d'âge est « âgée », moins son volume de production est homogène. C'est ainsi parmi les catégories les plus jeunes de notre population que l'étendue du nombre de tweets politiques publiés est la moins forte, et inversement, parmi les plus âgées qu'elle est la plus développée.

Mais l'âge n'est pas la seule variable sociologique significativement corrélée au nombre de tweets politiques publiés. Le volume de production de tweets politiques est en effet également déterminé par la PCS.

Tableau 4.5 ANOVA nombre de tweets publiés et PCS

Variable	PCS	n =	Moyenne	F	valeur de p	$p < 0,05$ = *
Nombre de tweets politiques publiés	Artisans, commerçants, et chefs d'entreprise	88	551,0	2,454	0,000	*
	Cadres et prof. intellectuelles	152	121,8			
	Professions intermédiaires	81	205,5			
	Employés, ouvriers ⁶⁵⁵	42	84,3			
	Retraités	20**	1011,9			
	Autres inactifs (dont étudiants)	255	71,6			
	Total	558	146,2			

**Effectifs faibles

On peut cependant objecter que ces différences en termes de PCS ne font en réalité que refléter les différences en termes d'âge que l'on vient d'évoquer. Les catégories « autres inactifs » concentrent ainsi l'essentiel des moins de 25 ans et la catégorie des « retraités » les plus de 60. Pour contrôler ce biais, deux choix peuvent être alors opérés. Le premier est de mener ce test statistique sur une population qui exclue ces deux catégories. La corrélation entre PCS et nombre de tweets politiques est alors toujours significative bien que de plus

⁶⁵⁵ Afin d'obtenir des effectifs autorisant le traitement statistique, nous avons regroupé les PCS employés et ouvriers.

faible intensité ($n = 283$, $F = 2,526$, $p = 0,048$). Le second choix consiste à reclasser les « retraités » dans leur PCS d'origine, lorsqu'ils étaient encore actifs. Dans cette seconde configuration, la corrélation reste là encore significative ($n = 558$, $F = 3,892$, $p = 0,004$). Enfin, précisons qu'après avoir listé les autres variables significatives de notre modèle, nous réaliserons une régression linéaire, ce qui nous permettra de raisonner *toute chose égale par ailleurs* et donc d'objectiver cette relation de dépendance entre âge et PCS.

Significativité des variables politiques

Parmi les trois variables « politiques » testées (le niveau d'intérêt déclaré pour la politique, le fait d'adhérer ou non à un parti politique et l'orientation politique déclarée), seule l'orientation politique n'a pas d'influence sur le volume de tweets politiques publiés. Se positionner « très à gauche », « plutôt à gauche », « au centre », « plutôt à droite », « très à droite » ou « ni à gauche, ni à droite » ne détermine pas le nombre de tweets politiques publiés ($F = 1,242$, $p = 0,288$). Ce résultat semble infirmer *a priori* la thèse de Christopher Anderson et Aida Paskevicate, selon laquelle les citoyens d'un pays parleraient moins fréquemment de politique lorsqu'ils partagent les mêmes convictions politiques que le gouvernement en place⁶⁵⁶. Alors que la gauche était au pouvoir durant la séquence étudiée (du 1^{er} mars 2013 au 31 janvier 2014), les producteurs de gauche n'ont pas moins publié de tweets politiques que ceux de droite. Il convient néanmoins de rappeler que nous avons analysé la volumétrie des panels lors d'une séquence temporelle différente de celle durant laquelle ils ont été constitués (du 1^{er} au 31 mars 2012). La gauche était alors encore dans l'opposition. Nos résultats ne permettent donc pas de déterminer dans quelle mesure les individus de droite qui ne publiaient pas de tweets politiques lorsque la droite était au pouvoir (et qui ne figurent logiquement pas dans notre panel) « parlent » aujourd'hui de politique sur Twitter, depuis que la gauche est au pouvoir. On peut simplement affirmer que les individus de droite qui « parlaient » de politique même lorsque la droite était au pouvoir (et qui figurent donc dans notre panel) publient aujourd'hui des tweets politiques dans les mêmes proportions que les individus de gauche.

⁶⁵⁶ Christopher J. Anderson et Aida Paskevicate, « Macro- politics and micro- behavior: mainstream politics and the frequency of political discussion in contemporary democracies », in Alan Zuckerman (dir.), *The social logic of politics. Personal networks as contexts for political behaviour*, Philadelphie, Temple University Press, p.228-250, 2005.

On enregistre en revanche une relation significative (de moyenne intensité) entre le niveau d'intérêt déclaré pour la politique et le nombre de tweets politiques publiés sur la période ($\rho = 0,272$, $p = 0,000$). Les « répondants » qui déclarent « beaucoup » s'intéresser à la politique ont publié en moyenne huit fois plus de tweets politiques que ceux qui s'y intéressent « peu » ou « pas du tout » et quatre fois plus que ceux qui s'y intéressent « assez ».

Tableau 4.6 Test de corrélation de Spearman entre l'intérêt déclaré pour la politique et le nombre de tweets politiques publiés

Variable	Groupe	n =	Moyenne	P	valeur de p	$p < 0,05 = *$
Nombre de tweets politiques publiés	Un peu ou pas du tout	72	29,14	0,272	0,000	*
	Assez	171	53,59			
	Beaucoup	315	223,31			
	Total	558	146,25			

Ce résultat ne semble certes guère surprenant. Les individus qui s'intéressent le plus à la politique sont ceux qui publient le plus de messages politiques. Il est intéressant néanmoins de constater que l'intérêt pour la politique, qui est une variable *déclarative*, détermine *objectivement* le nombre de messages politiques publiés. Dans ce cas de figure, les données issues de l'observation corroborent bien les données générées par le biais de l'interrogation.

Enfin, le fait d'adhérer à un parti politique détermine également le nombre de tweets politiques produits (Test $t = 2,256$, ddl = 556, $p = 0,024$). Les militants ont ainsi publié en moyenne sur la période deux fois plus de tweets politiques (225 tweets, soit 20 tweets par mois) que les non militants (109 tweets, soit 10 tweets par mois).

Tableau 4.7 Comparaison du nombre de tweets politiques publiés entre militants et non militants

Variable	Groupe	n =	Moyenne	Test t	valeur de p	$p < 0,05 = *$
Nombre de tweets politiques publiés	Militants	178	225,30	2,256	0,024	*
	Non militants	380	109,22			
	Total	558	146,25			

Là encore, ce résultat n'est pas très surprenant. Nous avons vu dans le précédent chapitre que les auteurs militants publiaient un très grand nombre de tweets politiques, notamment pour communiquer des articles de presse relatifs à leur parti politique, au niveau local ou national, mais également pour « soutenir » leurs leaders politiques ou au contraire « attaquer » le camp adverse, à travers des échanges souvent nombreux et quasi-synchrones avec d'autres utilisateurs. Ces activités génèrent donc un niveau plus élevé de tweets politiques. On aurait même pu supposer que la différence de production entre militants et non militants serait encore plus marquée. Cependant, comme nous l'avons évoqué dans le précédent chapitre, il existe des producteurs qui n'adhèrent pas à des organisations politiques, mais qui mobilisent néanmoins leur compte Twitter comme un véritable instrument « militant ». Nous avons alors qualifié ces producteurs d'« auto-entrepreneurs militants ». Ces individus publient un nombre important de tweets politiques, souvent aussi élevé que celui des militants-adhérents. Cependant, notre test de corrélation étant basé sur la variable déclarative « *adhérez-vous à un parti politique ?* », les tweets de ces « auto-entrepreneurs militants » ont été comptabilisés dans le groupe des « non militants ». Il est alors fort probable que si ces individus avaient été catégorisés comme « militants », le différentiel de volumétrie entre les deux groupes aurait été encore plus prononcé.

Significativité des variables-test issues des panels

Enfin, avant de mener notre analyse de régression, nous avons mesuré la significativité des cinq variables-test issues directement de nos panels : le nombre total de tweets publiés, le nombre d'abonnés, le nombre total d'abonnements, le nombre d'abonnements à des médias politiques et le nombre d'abonnements à des hommes politiques. Contrairement aux autres variables, nous avons pu effectuer ces tests statistiques sur nos deux populations (« répondants » et « non répondants »). En effet, ces variables ne sont pas issues de notre questionnaire mais sont directement générées par les comptes des individus panélisés (elles sont pour certaines immédiatement disponibles, comme le nombre total d'abonnés ou d'abonnements, et pour d'autres calculées par le chercheur, comme le nombre d'abonnements à des médias ou à des hommes politiques⁶⁵⁷).

⁶⁵⁷ Nous développerons ce point plus en détail notre prochain chapitre, consacré aux logiques sociales d'abonnements.

Le nombre total de tweets publiés est la variable la plus fortement corrélée au nombre de tweets politiques produits. Le nombre total de tweets (politiques et non politiques) publiés par les « répondants » ($r = 0,437$, $p = 0,000$) et les « non répondants » ($r = 0,309$, $p = 0,000$) détermine en effet significativement le volume de production de tweets *politiques*.

Tableau 4.8 Test de corrélation de Pearson entre le nombre de tweets et le nombre de tweets politiques

Population	Variable	<i>n</i> =	<i>r</i>	valeur de <i>p</i>	$p < 0,05 = *$
Répondants	Nombre de tweets politiques	558	0,437	0,000	*
	Nombre total de tweets				
Non répondants	Nombre de tweets politiques	517	0,309	0,000	*
	Nombre total de tweets				

Nous garderons donc cette variable dans notre modèle, en postulant que le nombre total de tweets publiés par un individu peut contribuer à « prédire » le volume de production de tweets politiques.

Si l'on s'intéresse maintenant au nombre d'abonnés des membres de nos panels, on enregistre là encore une relation significative, mais de plus faible intensité, avec le nombre de tweets politiques publiés par les « répondants » ($r = 0,159$, $p = 0,000$) et les « non répondants » ($r = 0,087$, $p = 0,000$).

Tableau 4.9 Test de corrélation de Pearson entre le nombre de tweets et le nombre d'abonnés

Population	Variable	<i>n</i> =	<i>r</i>	valeur de <i>p</i>	$p < 0,05 = *$
Répondants	Nombre de tweets politiques	558	0,159	0,000	*
	Nombre d'abonnés				
Non répondants	Nombre de tweets politiques	517	0,087	0,001	*
	Nombre d'abonnés				

Plus un individu a d'abonnés à son compte, plus il publie de tweets politiques. Un nombre élevé d'abonnés peut trahir un rapport plus investi au réseau social (nous avons pu observer que les individus qui bénéficiaient d'un nombre important d'abonnés publiaient très fréquemment des tweets) mais aussi un plus important capital social numérique (composé de relations « réelles » et « virtuelles »). Tout se passe comme si les individus qui se savaient suivis – et potentiellement lus – par un nombre important d'individus avaient davantage tendance à « parler » de politique sur Twitter.

Ce même constat peut être fait concernant le nombre d'abonnements. Là encore, on observe une relation significative avec le nombre de tweets politiques publiés.

Tableau 4.10 Test de corrélation de Pearson entre le nombre de tweets et le nombre d'abonnés

Population	Variable	<i>n</i> =	<i>r</i>	valeur de <i>p</i>	<i>p</i> < 0,05 = *
Répondants	Nombre de tweets politiques	58	0,217	0,000	*
	Nombre d'abonnements				
Non répondants	Nombre de tweets politiques	17	0,140	0,001	*
	Nombre d'abonnements				

Plus un individu est abonné à d'autres comptes Twitter, plus il publie de tweets politiques. Cependant, on peut se demander dans quelle mesure le nombre total de tweets publiés ne constituerait pas une variable « cachée ». On sait en effet que le nombre de tweets politiques est corrélé au nombre total de tweets publiés. Or on enregistre une relation toute aussi significative et même de plus forte intensité entre le nombre d'abonnements et le nombre total de tweets publiés ($r = 0,259$, $p = 0,000$ pour les « répondants »).

En revanche, le nombre d'abonnements à des comptes de médias politiques (comptes que l'on a isolés du nombre total d'abonnements), n'est pas significativement corrélé au nombre total de tweets publiés ($r = -0,30$, $p = 0,482$) mais est significativement corrélé au nombre de tweets politiques publiés ($r = 0,107$, $p = 0,013$). De manière analogue, le nombre d'abonnements à des hommes ou femmes politiques est corrélé au nombre de tweets politiques produits ($r = 0,178$, $p = 0,000$) mais n'est pas corrélé au nombre total de tweets publiés ($r = 0,013$, $p = 0,763$). Être abonné à un grand nombre de comptes politiques augmente

donc les chances de produire un nombre élevé de tweets politiques, mais n'augmente pas les chances de publier des tweets en général.

Ces deux variables sont d'ailleurs significativement corrélées au niveau d'intérêt déclaré pour la politique. Les individus « répondants » qui ont déclaré « beaucoup » s'intéresser à la politique sont abonnés en moyenne à sept fois plus d'hommes politiques (13 abonnements en moyenne) que ceux qui déclarent s'y intéresser « un peu » ou « pas du tout » (2 abonnements en moyenne⁶⁵⁸).

Tableau 4.11 Intérêt déclaré pour la politique et nombre d'abonnements à des médias et hommes politiques

Variable	Groupe	n =	Moyennes	ρ	valeur de p	p < 0,05 = *
Nombre d'abonnements à des médias politiques	Un peu ou pas du tout	72	3	0,321	0,000	*
	Assez	171	6			
	Beaucoup	315	12			
	Total	558	9			
Nombre d'abonnements à des hommes politiques	Un peu ou pas du tout	72	2	0,451	0,000	*
	Assez	171	3			
	Beaucoup	315	13			
	Total	558	8			

Ces résultats confirment donc que le nombre d'abonnements à des comptes de médias ou d'hommes politiques peut être appréhendé comme un indice permettant de mesurer l'intérêt pour la politique, et est donc à ce titre une variable déterminante à prendre compte pour « prédire » le volume de production de tweets politiques.

Au final, nous retiendrons donc huit variables dans notre modèle de régression linéaire : le nombre total de tweets, le nombre d'abonnés, le nombre d'abonnements à des hommes politiques, le nombre d'abonnements à des médias politiques, l'intérêt déclaré pour la politique, le fait d'adhérer à un parti politique (variables politiques), ainsi que l'âge et la PCS (variables sociologiques). Établir une régression linéaire du nombre de tweets politiques

⁶⁵⁸ Notons au passage que même les individus qui déclarent peu ou pas du tout s'intéresser à la politique sont abonnés en moyenne à deux hommes politiques. À la lecture de ce résultat, on peut donc supposer que même les catégories les moins politisées de notre échantillon sont, comparativement à la moyenne des Français, très politisées.

publiés permettra alors de repérer et d'ordonner les variables les plus influentes au sein de notre modèle.

Analyse de la régression linéaire du nombre de tweets politiques publiés.

Parmi ces huit variables, quatre ont été finalement écartées du modèle, la valeur t associée à leur coefficient β n'atteignant pas le seuil de significativité ($p < 0,05$) : le fait d'adhérer à un parti politique, le nombre d'abonnements à des comptes de médias politiques, le nombre d'abonnés et la PCS (rappelons néanmoins que si elles ne contribuent pas significativement au modèle, chacune de ces variables, prise individuellement, était pourtant bien corrélée significativement au nombre de tweets politiques publiés).

Notre modèle se compose donc de quatre variables : le nombre de tweets, l'âge, le nombre d'abonnements à des comptes d'hommes politiques et le niveau d'intérêt déclaré pour la politique (1 : peu ou pas du tout, 2 : assez, 3 : beaucoup). Ce modèle de régression explique alors 28,5% de la variabilité du nombre de tweets politiques publiés ($R^2 = 0,285$).

Tableau 4.12 Analyse de régression linéaire du nombre de tweets politiques publiés par les « répondants »

Variabes dans l'équation	A (coefficients non standardisés)	Erreurs standards	β (coefficients standardisés)
Nombre total de tweets	0,090	0,007	0,445***
Âge	9,794	1,601	0,223***
Nb. abonnements hommes pol.	4,945	1,361	0,137***
Intérêt déclaré	66,736	30,482	0,083*

Note : * = Significativité inférieure à 5 %, ** = Significativité inférieure à 1 %, *** = Significativité inférieure à 0,1 %.

$R^2 = 0,285$

Notre modèle de régression peut donc se résumer à l'équation suivante :

$$Y_{nb\ Tw.\ pol.} = -491,295 + 0,007 nb\ Tw. + 1,601 nb\ ans + 1,361 Nb.\ abo.\ hommes\ pol. + 30,482\ niv.\ intérêt\ pol.$$

À âge, niveau d'intérêt et nombre de tweets constants, être abonné à un homme politique supplémentaire augmente de 1,4 tweets la production totale de messages politiques publiés. Précisons ici que nous nous gardons bien de tomber dans une sorte de « scientisme statistique ». Notre propos se limite simplement à donner du sens et à illustrer cette équation mathématique, dont il convient de rappeler que tous les paramètres ont été définis arbitrairement, en fonction de données empiriques construites (et parfois incomplètes) et d'hypothèses de recherche que nous avons essayé d'explicitier le plus possible. La complexité des logiques sociales qui président à la publication des messages politiques sur un réseau social ne serait bien sûr se réduire à une équation mathématique.

La comparaison des coefficients β des variables de notre modèle peut néanmoins nous aider à hiérarchiser le niveau d'influence des différents facteurs testés sur le niveau de production de tweets politiques. Le nombre total de tweets publiés apparaît alors comme le facteur le plus déterminant. Publier un grand nombre de tweets suppose l'adoption d'une posture active sur le réseau social : se connecter régulièrement et développer une réelle activité sur son compte, notamment en « discutant » régulièrement avec ses abonnés. Cette posture active s'oppose donc à une posture passive, qui consiste pour l'essentiel à « suivre » l'actualité des comptes auxquels on est abonné, sans soi-même publier de tweets. Tout se passe comme si publier régulièrement des tweets était la condition nécessaire (mais pas suffisante) à la production de tweets politiques. Être « bavard » politiquement sur Twitter suppose donc avant tout d'être « bavard » de manière générale sur le réseau social. Cependant, contrairement à l'âge, le nombre total de tweets publiés peut être aussi une variable à expliquer, et non uniquement une variable explicative.

Nous avons alors tenté d'identifier les facteurs qui déterminent le nombre total de tweets publiés, en menant une analyse de régression linéaire en fonction des variables explicatives suivantes : sexe, âge, niveau de diplôme, PCS, PCS du père et lieu de résidence. De ces sept variables, seule la PCS contribue significativement au modèle. Le modèle ne parvient néanmoins à expliquer que 2,7 % de la variabilité du nombre total de tweets publiés. Contrairement au nombre de tweets politiques, les facteurs sociologiques explicatifs du nombre total de tweets publiés par les membres de notre panel restent donc un point aveugle de notre démonstration.

L'âge, le nombre d'abonnements à des hommes politiques et le niveau d'intérêt déclaré pour la politique apparaissent ensuite respectivement comme les facteurs les plus déterminants. En résumé, publier un nombre important de tweets politiques suppose donc de réunir les trois caractéristiques suivantes : adopter une posture active sur le réseau social en publiant régulièrement des tweets, être âgé et être très intéressé par la politique (intérêt que l'on peut mesurer déclarativement mais aussi objectivement, en mesurant le nombre d'abonnements à des hommes politiques).

Nous avons donc voulu démontrer dans cette première section que le nombre de tweets politiques publiés est une variable très inégalement distribuée au sein de notre population. Cette inégale distribution peut d'abord être observée en comparant nos deux panels. C'est alors l'administration de notre questionnaire et ses effets induits qui permettent d'expliquer ces différences de volumétrie. Mais cette inégale distribution du nombre de tweets politiques publiés peut également s'expliquer par des facteurs sociologiques et politiques individuels. Nous avons vu ainsi que l'âge, le niveau d'intérêt déclaré et objectivé pour la politique ainsi que le nombre total de tweets publiés par les producteurs permettaient d'expliquer une part significative de la variabilité du nombre de tweets politiques publiés.

Nous avons alors jusqu'ici analysé le nombre *total* de tweets politiques publiés par les membres de nos panels sur l'ensemble de la période étudiée (du 1^{er} mars 2013 au 31 janvier 2014), sans tenir compte de la distribution de cette production dans le temps. Or si l'on intègre maintenant la variable *temps* à l'analyse, on constate alors de nouvelles disparités dans la production de nos panels.

2. L'inégale distribution du nombre de tweets politiques publiés dans le temps

Nous allons donc maintenant analyser la distribution temporelle du nombre de tweets politiques publiés par nos deux panels.

Nous verrons ainsi que l'inégale distribution du nombre de tweets politiques dans le temps est un phénomène commun aux deux panels, même si les taux quotidiens de production du panel « répondants » sont toujours significativement plus élevés que ceux du panel « non répondants » (2.1). Puis nous verrons que cette inégale distribution dans le temps concerne toutes les catégories de producteurs, et pas uniquement les moins politisées (2.1).

2.1 Une inégale distribution dans le temps

Si le nombre de tweets politiques publiés est très inégal d'un individu à l'autre, il est également très inégal d'un jour à l'autre. Le nombre de tweets politiques, comme le nombre d'individus qui les publient, varie en effet significativement selon la date de publication des messages sur le réseau social.

Contrairement aux sondages qui ne peuvent, essentiellement pour des raisons matérielles, interroger les individus qu'à des échéances plus ou moins espacées dans le temps, notre méthodologie permet de mesurer la production des membres de nos panels *en temps réel*. Chaque tweet recensé par notre logiciel de collecte est horodaté : on connaît ainsi précisément l'année, le mois, le jour, l'heure, la minute et même la seconde de publication de chaque message. Nous avons choisi d'adopter la journée comme unité temporelle (même si cette unité peut être à tout moment substituée à une autre pour les besoins de l'analyse). Sur l'ensemble de la période étudiée – du 1^{er} mars 2013 au 31 janvier 2014, soit 337 jours – nous avons donc enregistré à 337 reprises le *nombre quotidien* de tweets politiques publiés par nos deux panels, mais également *leur taux quotidien* de publication.

Le *nombre quotidien de tweets politiques* correspond à la simple agrégation du nombre de tweets politiques publiés par chacun des membres des panels. Cet indice permet donc de mesurer assez simplement le volume total de production d'un panel donné pour une journée donnée. Cependant, le nombre de messages publiés peut être très inégal d'un individu à

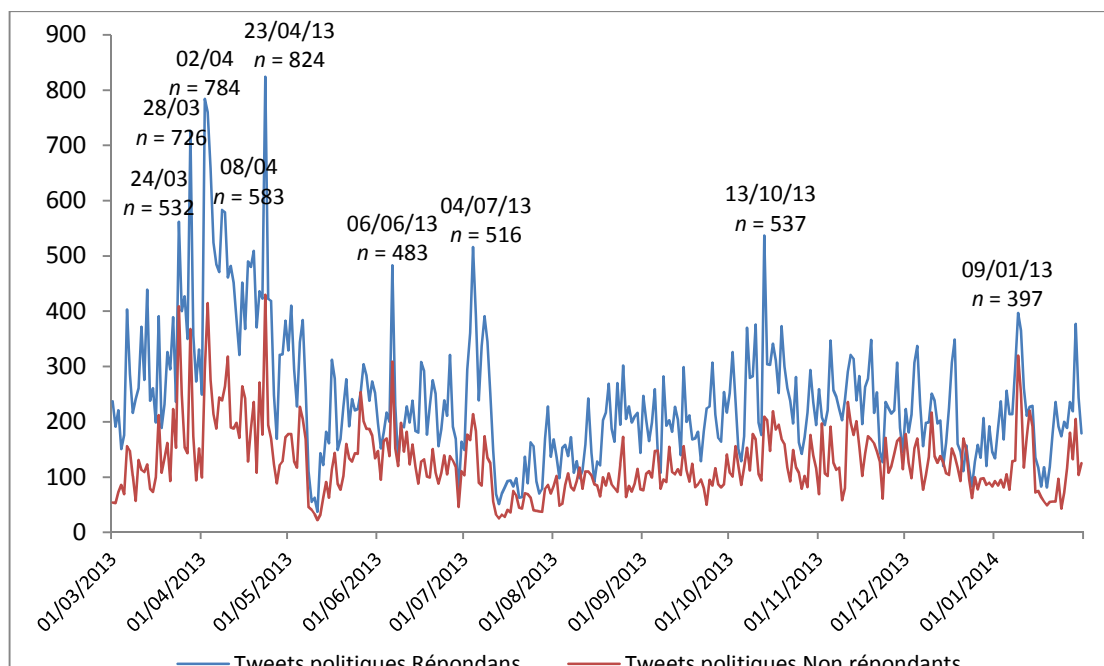
l'autre, même lors d'une seule journée. Afin de contrôler ce biais, mais également afin de pouvoir comparer les productions des deux panels avec une même unité de mesure (rappelons que nos deux panels ne sont pas tout à fait composés du même nombre d'individus), nous avons alors mobilisé un second indice, le *taux quotidien de publication*. Ce taux correspond à la part relative du nombre d'individus ayant publié un ou plusieurs tweets politiques sur une journée donnée par rapport au nombre total d'individus qui composent le panel.

Tableau 4.13 Production quotidienne des panels « répondants » et « non répondants »

	Nombre quotidien de tweets politiques publiés par les panels		Taux de publication politique quotidien des panels	
	Répondants	Non répondants	Répondants	Non répondants
<i>n</i> =	558	517	558	517
Total	81606	42940	485 (87%)	419 (81%)
Minimum	37	22	5%	3%
Maximum	824	430	27%	17%
Moyenne	242	127	11%	8%
Médiane	221	115	11%	8%
Ecart type	120	63	3 points	2 points

Quel que soit l'indice de mesure retenu, on enregistre de fortes disparités volumétriques d'un jour à l'autre. Ainsi, à titre d'exemple, le 11 mai 2013, le panel « répondants » n'a produit que 37 tweets politiques (publiés par 5 % de ses membres) alors que le 23 avril, ce même panel en publiait 824 (27 % de ses membres s'étaient alors exprimés politiquement ce jour-là). Bien qu'il soit moins prononcé, on observe le même phénomène avec le panel « non répondants » (avec 120 tweets contre 63, l'écart-type du nombre journalier de tweets politiques est deux fois plus élevé parmi les « répondants »).

Graphique 4.1 Distribution journalière des tweets politiques publiés par les panels « répondants » et « non répondants » du 01/03/2013 au 31/01/2014



Trois enseignements peuvent être tirés à la lecture de ce graphique. Le premier concerne la symétrie des courbes des deux panels. Le nombre quotidien de tweets politiques publiés par le panel « répondants » varie d'un jour à l'autre de manière relativement identique à celui du panel « non répondants ». Ce résultat est d'autant plus remarquable eu égard aux sociologies et aux modalités de constitution différentes des deux panels. Alors que le panel « répondants » concentre les plus gros producteurs de tweets politiques (et sans doute aussi les plus politisés), son volume et son taux quotidien de production évoluent pourtant de la même manière que ceux des « non répondants ».

Cependant, si l'évolution quotidienne de la production des deux panels suit la même tendance, le panel « répondants » connaît systématiquement – et sans aucune exception – un volume et un taux de production supérieurs au panel « non répondants ». Il s'agit là d'un deuxième enseignement. En moyenne, le taux de production du panel « répondants » est de trois points supérieur à celui du panel « non répondants » (11 % contre 8 %). Ce résultat n'est certes guère surprenant, si l'on se remémore que sur l'ensemble de la période, le panel « répondants » a produit deux fois plus de tweets politiques que le panel « non répondants ».

Cependant, il est intéressant de noter que ces écarts de volumétrie se reproduisent à l'identique chaque jour, nous aurons l'occasion de le démontrer par la suite.

Enfin, le dernier enseignement que l'on peut tirer à la lecture de ce graphique concerne la présence de très nets pics de production. On peut observer d'une part que ces pics, à l'exception de ceux parus au mois d'avril, sont relativement espacés dans le temps et qu'ils sont d'autre part très « brefs » : le volume de tweets chute dans la plupart des cas dès le lendemain de la formation du pic. Nous avons en particulier isolé neuf pics de production et consulté des échantillons de tweets parus ces jours-là. Chacun de ces pics correspond en fait à un « événement » politique, qui a « occupé le haut de l'agenda médiatique ».

Tableau 4.14 Volume et taux de production quotidiens les plus élevés des panels

Date	Nb Tweets REP	% REP	Nb Tweets N REP	% NREP	Événement
23 Avril 2013	824	27	430	16	Adoption définitive de la loi sur le mariage pour tous par l'Assemblée nationale
2 Avril 2013	679	25	415	17	Jérôme Cahuzac avoue avoir fraudé le fisc
28 Mars 2013	726	19	368	16	Interview télévisée de François Hollande au 20H de France 2
8 Avril 2013	583	19	239	12	Publication des déclarations de patrimoine de certains ministres
6 Juin 2013	483	19	309	11	Mort de Clément Méric, un jeune militant d'extrême gauche
24 Mars 2013	562	17	409	15	« Manif pour tous » : 300 000 à 1,4 M de participants + Législative partielle UMP/FN Oise
4 Juillet 2013	516	17	214	10	Comptes de campagne de Nicolas Sarkozy invalidés
9 Janvier 2014	397	15	320	14	<i>Closer</i> révèle la liaison Hollande/Gayet
13 Octobre 2013	537	14	209	9	Premier tour de la primaire PS à Marseille

Tout semble en fait incliner le chercheur à attribuer un rôle causal à l'événement⁶⁵⁹ dans la compréhension de ces pics de production de tweets politiques. On peut définir ici simplement « l'événement » politique comme une surproduction inhabituelle et ponctuelle de biens informationnels politiques, notamment dans les médias télévisés, radiophoniques et de presse écrite (papier ou numérique).

On peut alors se demander quelles sont les catégories de producteurs les plus sensibles à ces « événements » politiques, c'est-à-dire les plus enclines à les commenter ? Sont-ce les plus politisées, qui auraient une réelle appétence pour des « événements » politiques *extraordinaires* ? Ou au contraire les moins politisées, qui ne s'exprimeraient politiquement que lorsque le temps politique se synchronise avec le temps social⁶⁶⁰ ?

2.2 Une inégale distribution temporelle qui concerne toutes les catégories de producteurs

Ces « événements » semblent en fait générer des taux de production de tweets politiques inhabituels (bien qu'inégaux) parmi toutes les catégories de producteurs, des plus « petits » aux plus « gros ».

Tableau 4.15 Production de tweets à l'occasion d'« événements » politiques

Nb de Tw Pol publiés sur la période (01 mars 2013 – 31 janvier 2014)	% du panel (n = 558)	% qui ont commenté au moins un des neuf événements
0 tweet	13	-
1 à 9 tweets	30	25
10 à 50 tweets	26	76
51 à 200 tweets	18	97
Plus de 201 tweets	13	100

Avant de commenter plus avant ces résultats, des précisions doivent être ici apportées concernant les individus qui n'ont publié aucun tweet politique sur l'ensemble de la période,

⁶⁵⁹ Nous nous permettons ici de paraphraser Stéphane Latté qui, dans son article, s'interroge sur les effets mobilisateurs des événements dans les mobilisations collectives. Cf. Stéphane Latté « La "force de l'événement" est-elle un artefact ? », *Revue française de science politique*, 3/ 2012 (Vol. 62), p. 409-432.

⁶⁶⁰ Vincent Goulet, *Médias et classes populaires. Les usages ordinaires des informations*, INA Éditions, Paris, 2010.

soit 13% des individus qui ont répondu à notre questionnaire (administré entre les mois de mai et juillet 2012). Ces individus ont pourtant tous publié des tweets (non politiques) sur la période et ont déjà, par le passé, publié des tweets politiques, au minimum un durant la séquence présidentielle du printemps 2012 (qui était la condition *sine qua non* de leur éligibilité au panel). Malgré cela, ces individus n'ont publié aucun tweet politique, même pour commenter ou évoquer les « événements politiques » de la séquence (01 mars 2013 – 31 janvier 2014). À l'exception de l'âge qui n'apparaît pas comme une variable discriminante, ces individus se différencient des autres producteurs par rapport aux trois autres variables de notre modèle de régression : l'intérêt déclaré pour la politique, l'intérêt objectif (nombre d'abonnements à des comptes de femmes ou d'hommes politiques) et le nombre total de tweets publiés. Les individus qui n'ont pas publié de tweets politiques sur la période sont ainsi significativement moins nombreux que la moyenne à déclarer « beaucoup » s'intéresser à la politique (46 % contre 71 % des individus qui ont produit un ou plusieurs tweets politiques sur la période), sont abonnés à trois fois moins de comptes de femmes ou d'hommes politiques (3 contre 9), et surtout ont publié sur l'ensemble de la période un nombre total de tweets 35 fois inférieur à celui de ceux qui ont publié des tweets politiques (50 tweets en moyenne contre 1724 pour ceux qui ont publié au minimum un tweet politique).

Les individus qui ne se sont pas exprimés politiquement sur Twitter sur l'ensemble de la période présentent donc deux propriétés distinctives : ils sont relativement moins politisés (bien que largement plus politisés que la moyenne des Français⁶⁶¹) et sont de très faibles producteurs de tweets en général. Tout se passe donc comme si être très intéressé par la politique était une condition nécessaire mais non suffisante à la production de tweets politiques. Produire des tweets politiques nécessite également de produire régulièrement des tweets en général, c'est-à-dire d'adopter une posture active sur le réseau social (et ne pas se contenter de consulter passivement et occasionnellement les tweets des comptes auxquels on est abonné).

Seule l'existence d'événements politiques de très forte intensité médiatique (comme les élections présidentielles, qui parviennent à générer un niveau de production de biens politiques médiatiques autrement plus important que l'ensemble des « événements » politiques recensés sur la période) semble parvenir à ce que ces individus s'expriment politiquement sur Twitter. Or aucun événement politique de cette ampleur n'est survenu durant notre période d'étude (il n'y avait notamment aucune échéance électorale).

⁶⁶¹ Cf. chapitre 3.

Près d'un tiers des individus « répondants » ont publié quant à eux moins de 10 tweets politiques sur l'ensemble de la séquence. Un quart de ces « petits producteurs » ont consacré au moins un de leurs tweets politiques à commenter l'un des neuf « événements » politiques recensés plus haut. Ces « événements » politiques génèrent donc des taux de production tout à fait inhabituels parmi ces catégories de « petits producteurs », surtout lorsque ces événements peuvent être interprétés à partir d'une grille de lecture « éthique » et non spécifiquement *politics* (comme le mariage homosexuel, l'affaire Cahuzac ou la liaison de François Hollande avec l'actrice Julie Gayet).

Encadre 4.1 Un petit producteur de tweets politiques réagit à l'affaire Cahuzac

J. est un conducteur routier de 30 ans, titulaire d'un CAP. Il se déclare « assez » intéressé par la politique, il se dit « ni de droite, ni de gauche », n'est abonné à aucun compte d'homme politique et a voté blanc au premier tour de l'élection présidentielle du 22 avril 2012 (un profil donc rare au sein de notre population). Du 1^{er} mars 2013 au 31 janvier 2014, il a publié 842 tweets (soit plus de deux tweets par jour en moyenne) dont seulement trois tweets politiques. Ces trois tweets ont été publiés le même jour, le mercredi 3 avril 2013, soit au lendemain des aveux de fraude fiscale du ministre socialiste du budget, Jérôme Cahuzac.

Tweet 1 : « @J_Cahuzac⁶⁶² vous pouvez revoir votre profil tweeter svp. » [alors qu'il a démissionné depuis le 19 mars 2013, son profil n'a pas encore été actualisé et le présente toujours comme ministre du budget]

Tweet 2 : « RT @PierreMenes: Plus je vois la politique plus j'aime le foot. »

Tweet 3 : « #Cahuzac ça commence à me les casser. Filez lui une corde sérieux! Tu fais le malin, tu tombes dans le ravin escroc à col blanc! »

Nous avons réalisé un entretien semi-directif avec J. le 19 avril 2013, soit 15 jours après qu'il ait publié ses tweets relatifs à l'affaire Cahuzac :

[Et est-ce qu'il vous arrive de parler politique sur Twitter ?] « Euh, ça m'a arrivé, comme euh, y a des semaines là, sur les escrocs en col blanc, moi je les appelle les escrocs en col blanc hein, ceux qui plaçaient de l'argent dans les pays étrangers et qui donnent des leçons aux Français ». [Vous aviez fait un tweet à propos de Cahuzac ?] « Oui j'avais marqué

⁶⁶² Cf. glossaire p. 464, pour comprendre la syntaxe spécifique des tweets.

un truc, j'avais vu un tweet de quelqu'un ou d'un journal je sais plus, ou alors c'est moi-même qui ai tweeté ça, je vais revenir en arrière, je sais plus ce que j'avais marqué exactement [il cherche en même temps sur son téléphone portable] c'est principalement pour le sport [qu'il va sur Twitter] c'est très rare que je tweete, ça m'a agacé en fait, ça m'a agacé de voir ce jour-là voir des gens faire des leçons et qui sont, j'appelle des escrocs en col blanc, donc je me souviens plus ce que j'ai tweeté, la personne concernée quoi, je m'en rappelle plus ce que j'ai marqué exactement, c'est rare, très très rare ». [Et quand vous le faites, c'est pour pousser des coups de gueule ?] « Ben voilà c'est comme un truc comme là ça m'a énervé donc j'avais envie de le dire et je l'ai dit, donc je sais pas si le message a été lu par quelqu'un, peut-être que oui peut-être que non, c'est pas mon problème, je m'en fiche à la limite, qu'y ait quelque chose ou pas, c'était juste pour ouais dire vous faites chier, ouais voilà c'était l'affaire Cahuzac c'est ça [il a retrouvé les tweets en question], voilà je suis revenu dessus j'ai trouvé. Voilà c'est plus pour pousser un coup de gueule que pour féliciter un homme politique, moi vraiment pour que je parle politique il faut que ça soit quelque chose qui m'agace, sinon ça m'intéresse pas trop quoi ».

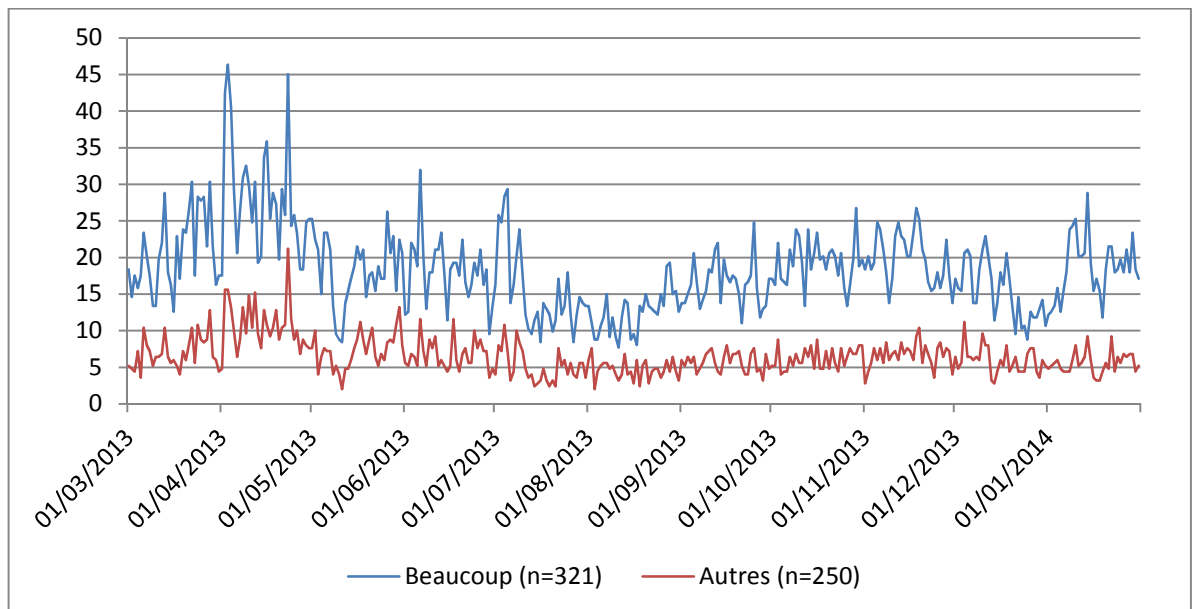
Cet enquêté ne s'exprime donc que très rarement sur la politique sur Twitter, et le fait exclusivement sur le mode de la protestation ou de la prise de parole (*voice*), pour reprendre les termes d'Albert Hirschman⁶⁶³. Il interprète l'affaire Cahuzac à partir d'une grille de lecture plus éthique que politique. S'il s'adresse directement au ministre (Tweet 1) il ne le fait pas tant pour être lu par ce dernier que pour exprimer à « voix haute » son dégoût, tel un automobiliste qui insulterait un autre automobiliste sans descendre de son véhicule. Il est enfin intéressant de constater à quel point la mémoire politique (et la mémoire des tweets politiques publiés) semble fragile et de courte durée. À peine deux semaines après leur publication, J. ne se souvenait déjà plus précisément des tweets qu'il avait publiés à ce sujet. Il pensait également n'avoir jamais publié de tweets sur l'élection présidentielle, ce qui n'était pourtant pas le cas (il a durant la campagne publié quatre tweets politiques). La production de tweets politiques est donc pour cet enquêté une activité anormale (au sens statistique du terme) qui ne se produit que lors de conjonctures politiques de crise ou de très haute intensité.

⁶⁶³ Albert Hirschman, *Exit, voice, loyalty. Défection et prise de parole*, Bruxelles, Éditions de l'Université de Bruxelles, 2011.

Si ces événements politiques génèrent donc des taux de production inhabituellement élevés parmi les petits producteurs de tweets politiques, ces taux restent néanmoins bien moins importants que ceux que l'on peut enregistrer parmi les plus gros producteurs (cf. tableau 4.14). Ainsi, 76 % des individus qui ont publié entre 10 à 50 tweets politiques sur la période ont commenté au moins l'un des « événements » de la période (contre seulement 25 % des individus qui ont publié moins de 10 tweets). Ce taux s'élève même à 100 % pour les individus qui ont publié plus de 200 tweets politiques.

Si l'on observe maintenant l'évolution comparée du taux de production quotidien des « répondants » qui déclarent s'intéresser « beaucoup » à la politique avec celui des autres « répondants », on s'aperçoit que les deux courbes sont là encore symétriques. Autrement dit, à des échelles différentes, les inégalités journalières de production de tweets politiques sont les mêmes parmi les catégories les plus politisées comme les moins politisées (qui restent, rappelons-le encore une fois, largement plus politisées que la moyenne des Français).

Graphique 4.2 Évolution quotidienne du taux de production des « répondants » selon leur intérêt déclaré pour la politique (« beaucoup » VS autres modalités : assez, un peu, pas du tout)



En étudiant la distribution dans le temps du nombre de tweets politiques publiés par nos panels, nous avons donc constaté d'importantes disparités journalières en termes de volume de production, venant s'ajouter aux inégalités individuelles déjà analysées dans la première section de ce chapitre. Nous avons ensuite voulu démontrer que ces inégalités concernaient toutes les catégories de producteurs, et pas uniquement les moins politisées.

Ces inégalités temporelles sont particulièrement perceptibles lorsque se créent des « pics de production » de messages politiques sur Twitter. Nous avons alors observé – sans toutefois le prouver empiriquement – que ces pics se formaient à la suite d'« événements » politiques qui pouvaient eux-mêmes être appréhendés comme des pics de production de biens médiatiques politiques. Exception faite de ces pics, on peut alors maintenant se demander s'il existe une réelle corrélation entre la production de tweets politiques publiés par des comptes appartenant aux champs médiatique et politique et la production de tweets politiques publiés par les membres de nos panels.

3. « Offre » et « demande » d'informations et d'opinions politiques sur Twitter

Dans cette dernière section, nous allons questionner l'existence d'une corrélation statistique entre la production de nos panels représentatifs et la production de comptes d'agents et d'institutions appartenant aux champs médiatique et politique (comptes de médias, de journalistes, d'éditorialistes, de parlementaires, etc.).

Nous appréhenderons le volume de production de tweets politiques des comptes médiatiques et politiques comme un indice mesurant l'« offre » d'opinions et d'informations politiques et le volume de production des panels représentatifs comme un indice mesurant la « demande ». Nous emprunterons donc ici le cadre analytique du marché, mobilisé par Pierre Bourdieu dans *La Distinction* afin de rendre compte des logiques sociales de production et de consommation des opinions politiques⁶⁶⁴.

Dès lors, postuler, comme Daniel Gaxie⁶⁶⁵, que la conjoncture politique peut impacter le volume de production des opinions politiques formulées par des agents sociaux dits

⁶⁶⁴ Pierre Bourdieu, *La Distinction. Critique sociale du jugement*, Paris, Éditions de Minuit, 1979.

⁶⁶⁵ Daniel Gaxie, « Retour sur les modes de production des opinions politiques », *op. cit.*

« ordinaires » (c'est-à-dire n'appartenant pas aux champs politiques et médiatiques) revient à admettre, si l'on s'inscrit dans ce cadre analytique, que l'activité de l'« offre » détermine celle la « demande ».

Dit autrement, on se demandera si les variations de la production de messages politiques de nos panels représentatifs ne sont pas corrélées voire déterminées par les variations de la production des comptes d'agents et d'institution appartenant aux champs médiatiques et politiques. Les agents sociaux sur Twitter produisent-ils davantage de tweets politiques lorsque le niveau de l'offre est élevé ?

Avant de démontrer l'existence d'une telle corrélation entre « offre » et « demande » d'opinions (3.3), nous discuterons le cadre analytique adopté (3.1) et présenterons une méthodologie permettant d'étudier et de mesurer l'« offre » d'opinions et d'informations politiques sur le réseau social (3.2).

3.1 Le marché de l'opinion

Les notions d'offre et de demande d'opinions ont été mobilisées par Pierre Bourdieu dans *La Distinction* afin de rendre compte des logiques sociales de production et de consommation des opinions politiques par les agents sociaux. Pour le sociologue, « *le fait de produire une réponse à un questionnaire sur la politique, comme le fait de voter ou, à un autre niveau de participation, de lire un journal d'opinion ou d'adhérer à un parti, est un cas particulier [comme l'est, nous le verrons, le fait de publier un tweet politique] de rencontre entre une offre et une demande : d'un côté, le champ de production idéologique, univers relativement autonome, où s'élaborent, dans la concurrence et le conflit, les instruments de pensée du monde social objectivement disponibles à un moment donné du temps [cette notion de disponibilité temporellement située, donc conjoncturelle est, pour notre démonstration, capitale] et où se définit du même coup le champ du pensable politiquement, ou, si l'on veut, la problématique légitime ; de l'autre, des agents sociaux occupant des positions différentes dans le champ des rapports de classes et définis par une compétence politique spécifique plus ou moins grande [...]* »⁶⁶⁶. Dès lors il « *n'existe aucun autre moyen de s'approprier complètement [une problématique politique] que de participer réellement à sa production*

⁶⁶⁶ Pierre Bourdieu, *La Distinction*, op. cit., p. 465-466.

dans et par le jeu politique de facto réservé aux professionnels à plein temps (hommes politiques, permanents des appareils – mot adéquat en ce cas – de partis, de journalistes politiques, idéologues professionnels) »⁶⁶⁷.

En relevant cette seconde citation dans son article discutant des modes de production des opinions politiques⁶⁶⁸, Daniel Gaxie pointe une certaine contradiction (qu’il ne présente pas vraiment comme telle) entre cette affirmation d’une relative autonomie de l’offre par rapport à la demande d’opinion, et le *mode de production par expression d’un parti politique*, qui nécessite un haut niveau de compétence politique mais qui ne semble pas être cependant une disposition exclusive des agents appartenant aux champs de production idéologique. Daniel Gaxie nous rappelle ainsi que ce mode de production des opinions politiques est caractérisé par un ensemble de propriétés dont la « *production en première personne... et l’auto-gestion de l’opinion politique... par des petits propriétaires de leurs instruments de production des opinions politiques [Bourdieu, 1979a, p.485] »⁶⁶⁹. Or la production en première personne semble, d’après la première citation, réservée aux « professionnels à plein temps ».*

Pour Daniel Gaxie, « *le problème de savoir si des agents extérieurs aux champs de production sont en mesure de produire des opinions en première personne est une question empirique »⁶⁷⁰. Il se fonde alors sur ses propres observations pour conclure que « les “amateurs” les plus compétents disposent surtout d’instruments de reproduction d’opinions. Ils sont souvent davantage reproducteurs que producteurs d’opinion, mais c’est en première personne qu’ils sont plus ou moins en mesure de s’approprier, de mémoriser et de mobiliser les prises de position qui circulent dans divers lieux de production et de diffusion idéologique. Ce n’est que de manière relativement exceptionnelle, dans les fractions les mieux équipées du public politisé – qui sont aussi les plus proches des milieux politiques – que l’on peut observer la production d’opinions personnelles en première personne par des agents qui sont alors vraiment des petits propriétaires de leurs moyens de production [et on pense ici à certains blogueurs politiques, qui, nous le verrons, figurent également dans le « panel influence » que nous allons par la suite mobiliser afin de mesurer l’offre sur Twitter] »⁶⁷¹.*

⁶⁶⁷ *Ibid.*, p. 504.

⁶⁶⁸ Daniel Gaxie, « Retour sur les modes de production des opinions politiques », *op. cit.*

⁶⁶⁹ *Ibid.*, p. 300.

⁶⁷⁰ *Ibid.*, p.300.

⁶⁷¹ *Ibid.*, p.301.

Nous postulons par la suite que les individus qui composent nos panels « répondants » et « non répondants » – qui, rappelons-le, n'appartiennent pas au champ de production de l'opinion – ne « produisent » pas mais « consomment » des opinions et des informations « produites » par les agents et institutions du champ médiatique et politique.

Pour être tout à fait précis, il convient ici de dissocier *production de tweet* et *production d'opinion*. Les individus qui ont répondu à notre questionnaire sont ainsi tous des « producteurs » de tweets politiques : ils composent et publient eux-mêmes des messages de 140 caractères relatifs à des objets politiques. Pour autant, ils ne sont pas des producteurs d'opinions. Les opinions ou les informations évoquées, commentées ou discutées dans leurs messages ont été produites par des agents appartenant aux champs médiatiques et politiques. En revanche, la *production d'un message politique* sur Twitter rend manifeste *la consommation d'une opinion ou d'une information*. Le réseau social rend en effet apparentes certaines formes de consommation des opinions.

On peut distinguer selon nous trois formes de consommation ostentatoire d'opinions et d'informations sur Twitter : retweeter un message publié par un compte appartenant au champ médiatique ou politique, indiquer dans un tweet publié en première personne un lien vers un site d'information, un site médiatique ou un site politique, ou encore évoquer dans un tweet un sujet inscrit à l'agenda médiatique, sans toutefois référencer l'origine de l'information ou de l'opinion consommée.

Une première forme de consommation ostentatoire consiste donc à retweeter le message d'un compte médiatique ou politique. Un individu sur Twitter peut en effet, nous l'avons déjà évoqué, republier à l'identique sur son compte un message qu'il a lu sur le compte d'un autre utilisateur (auquel il n'est pas forcément abonné). Cet autre utilisateur peut être un agent social situé à l'extérieur mais aussi à l'intérieur des champs de production de l'opinion et de l'information. Sont présents en effet sur Twitter l'ensemble des grands médias nationaux (télévisés, radiophoniques, de presse écrite – papier et numérique) mais également une part non négligeable d'agents qui appartiennent au champ médiatique (journalistes politiques, éditorialistes, rédacteurs en chef, « experts », etc.) ainsi qu'une grande majorité des agents du champ politique (élus, ministres, cadres militants, conseillers, etc.). Il est alors fréquent que les membres de nos panels retweetent des messages publiés par ces agents du champ de production idéologique.

D'après nos estimations⁶⁷², 54 % des tweets politiques publiés par les individus qui composent notre panel « répondants » sont des retweets et 52 % de ces retweets proviennent d'un compte – individuel ou institutionnel – appartenant au champ de production idéologique.

Tableau 4.16 Origine du compte retweeté (n = 576 tweets)

Origine du compte retweeté	%
« Grands » ⁶⁷³ médias	13
Journalistes de grands médias	9
« Petits » médias / blogs	5
Journalistes de petits médias /blogueurs	9
Organisations politiques	5
Hommes politiques	10
Autres	48
Total	100

Une deuxième forme de consommation ostentatoire consiste à publier en « première personne » un tweet où figure un lien URL renvoyant vers un article publié sur le site internet de ces mêmes médias, ou sur les blogs et sites personnels d'éditorialistes, de journalistes politiques, de ministres, d'élus, etc. Toujours d'après nos estimations, un tiers des tweets publiés par les membres de notre panel contiennent un lien hypertexte (36 %). La grande majorité de ces liens renvoient alors à des sites de « grands » médias (42 %).

⁶⁷² Nous avons analysé un échantillon de 1108 tweets (provenant de 334 individus) extrait aléatoirement de la base exhaustive de tweets politiques publiés par le panel « répondants » sur une période allant du 22 février au 22 avril 2013. Nous avons codé manuellement chacun de ces tweets en fonction de 22 variables (pour plus de détails sur la méthodologie et les variables codées, cf. Chapitre 5). Parmi ces variables, il y avait la nature du compte du tweet retweeté.

⁶⁷³ Nous avons opéré une distinction entre « grands » et « petits » médias essentiellement en fonction de critères d'audience, cf. Chapitre 5.

Tableau 4.17 Nature des sites vers lesquels renvoient les liens hypertextes publiés dans les tweets politiques du panel « répondants » (n = 363tweets)

Nature des sites des liens hypertextes	%
Lien vers un « grand » média	43
Lien vers un « petit » média ou un blog	27
Lien vers une organisation politique	8
Lien vers une photo	12
Lien Youtube ou Dailymotion	4
Autres liens	6
Total	100

Enfin, on peut identifier une troisième forme de consommation, plus indirecte. Le tweet publié ne fait ici pas directement référence à un média ou à un agent du champ de production idéologique mais commente une information politique ou exprime une opinion sur un sujet politique inscrit néanmoins à l’agenda médiatique. Ainsi, seul un tweet politique sur dix (publiés par les membres de notre panel « répondants ») ne fait pas référence, de manière directe ou indirecte, à un sujet politique inscrit à l’agenda médiatique.

Tableau 4.18 Objet des tweets politiques publiés par le panel « répondants » (n = 1 108 tweets)

Objet du tweet politique publié	%
Sujet inscrit à l’agenda médiatique	67
Réactions à des déclarations	15
Réactions à des prestations télévisées	7
Autres	11
Total	100

On peut donc en résumé appréhender Twitter comme un véritable marché de l’opinion et de l’information politique où s’observent et s’objectivent des actes de consommation de biens informationnels et d’opinions politiques, avec d’un côté une demande, dont on peut mesurer le niveau de consommation en comptabilisant le nombre de tweets politiques publiés par des panels représentatifs d’individus inscrits sur le réseau social (il s’agit de nos panels « répondants » et « non répondants ») et de l’autre une offre, dont on peut mesurer le niveau

de production en comptabilisant le nombre de tweets politiques publiés par les agents appartenant aux champs médiatiques et politiques.

Se pose alors la question méthodologique des conditions de réalisation de la mesure de cette offre.

3.2 Mesurer l'« offre » d'opinions sur Twitter : le panel « Twitter audience politique »

Afin de mesurer l'« offre » d'opinions politiques sur Twitter, nous mobiliserons le panel « Twitter audience politique » (produit par une équipe de recherche dont nous faisons partie mais dans un cadre distinct de celui de cette recherche doctorale⁶⁷⁴). Avant de détailler la composition de ce panel (3.2.2), nous évoquerons ses conditions de réalisation (3.2.1).

3.2.1 La réalisation du panel « Twitter audience politique »

La première condition nécessaire à une mesure de l'« offre » d'opinions sur Twitter est précisément d'identifier les comptes des agents et des institutions appartenant au champ de production de l'opinion et de l'information. Afin d'identifier ces comptes, deux postures méthodologiques peuvent être adoptées. La première consiste à partir d'une cartographie de ce champ, à en identifier précisément les agents et les positions qu'ils y occupent, et ensuite de vérifier s'ils sont présents ou non sur Twitter. Une telle entreprise apparaît néanmoins particulièrement coûteuse, pour la raison assez simple qu'il n'existe pas à notre connaissance de cartographie « toute faite » et surtout actualisée de ce champ (rappelons que si le réseau social a été créé en 2006, ce n'est véritablement que depuis les années 2011/2012 que les journalistes et les hommes politiques se sont inscrits en nombre sur Twitter). De plus, il y a fort à parier qu'une telle cartographie (érigée à partir d'une problématique de recherche qui ne serait pas la nôtre) n'intégrerait pas des producteurs « indépendants », propriétaires de « leurs petits moyens de production », tels que les blogueurs politiques, qui, pourtant, contribuent eux aussi à créer l'« offre » d'opinions politiques. Une seconde posture imaginable est alors

⁶⁷⁴ Contrairement aux panels représentatifs « répondants » et « non répondants », le panel « Twitter audience politique » n'a pas été réalisé dans le cadre de cette recherche doctorale. Nous tenons ici à remercier l'ensemble des personnes qui ont contribué à sa réalisation : Emmanuelle Rimbault, Jean-Yves Dormagen, Marie Neihouser, Rémi Andris, Geoffrey Della Valentina, Émeline Rochefeuille et Lysne Fröydis.

d'adopter une démarche essentiellement inductive et de partir d'une liste exhaustive d'individus inscrits sur le réseau social et ayant déjà publié un tweet politique (telle que la liste dont nous nous sommes servis pour échantillonner aléatoirement les individus à qui nous avons administré notre questionnaire) en hiérarchisant cette liste en fonction de critères objectifs permettant de mesurer l'audience de ces comptes.

C'est cette seconde posture méthodologique qui a été adoptée pour réaliser le panel « Twitter audience politique ». Nous avons hiérarchisé une liste exhaustive de plusieurs dizaines de milliers de comptes d'utilisateurs (ayant publié un ou plusieurs tweets politiques) en fonction d'un indice tenant compte à la fois du *nombre d'abonnés* à ces comptes et du *nombre moyen de reprises* (retweets) des messages politiques publiés par ces comptes. Le nombre d'abonnés constitue selon nous un indice permettant de mesurer l'audience *potentielle* des comptes Twitter. Un individu abonné à un compte x recevra dans sa *time-line* (un espace d'édition où défilent en temps réel et selon un ordre antéchronologique les tweets publiés par les comptes auxquels un individu est abonné) l'ensemble des tweets publiés par ce compte x . Cependant rien ne nous garantit que l'abonné du compte x lise effectivement dans sa *time-line* les tweets publiés par ce compte. C'est pour cette raison que l'on qualifie de « potentielle » l'audience mesurée par le nombre d'abonnés. Le nombre moyen de retweets d'un message politique constitue en revanche un indice permettant de mesurer une *audience effective*, ou une *consommation active* de tweets politiques. Retweeter un tweet publié par un autre utilisateur implique en effet *a minima* d'avoir lu ce tweet, voire de le trouver « intéressant ». On peut donc considérer le nombre moyen de retweets d'un tweet politique comme un indice permettant de mesurer l'audience effective d'un compte Twitter. Plus les tweets d'un compte sont repris (retweetés) par un nombre élevé d'utilisateurs, plus l'audience effective de ce compte sera élevée.

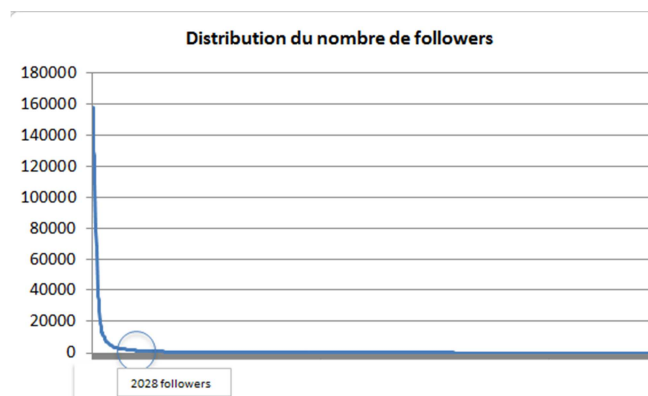
Nous avons donc construit un indice synthétique à partir de ces deux informations (nombre d'abonnés et nombre moyen de retweets politiques), calculé selon la formule suivante :

$Y_{\text{audience}} = \text{Nb. Abonnés} * (\sum \text{Nb. Retweets de tweets politiques} / \sum \text{Nb. Tweets politiques publiés}).$

Nous avons alors classé notre base exhaustive de 147 421 comptes en fonction de cet indice Y_{audience} . L'enjeu était ensuite de délimiter un seuil minimal d'audience en dessous duquel un compte ne pouvait être considéré faisant partie – ou gravitant autour – du champ

de production idéologique. Cette délimitation s'est faite là encore de manière relativement inductive. Nous avons en effet pu observer que la distribution de notre population en fonction de cette variable Y_{audience} prenait la forme d'une courbe logarithmique. C'est à partir d'un nombre équivalent à 2 028 abonnés que « s'envole » cette courbe. Seuls 1041 comptes (soit 0,7 % de la population) étaient situés au-dessus de ce seuil. Le panel « Twitter audience politique » est donc au final composé de ces 1041 comptes.

Graphique 4.3 Distribution du nombre d'abonnés aux comptes producteurs de tweets politiques



Source : Panel « Twitter audience politique »

L'ensemble de ces comptes a été alors « codé » manuellement en fonction d'une grille comprenant neuf variables : la catégorie du compte (média, agent du champ médiatique, agent du champ politique, organisation, personnalité publique), la sous-catégorie (avec une déclinaison de modalités pour chaque catégorie, comme par exemple, pour la catégorie « agents du champ médiatique » : éditorialiste, directeur de la rédaction, chef de service, présentateur, chroniqueur, journaliste), l'indicateur de position dans le champ (qui mesure la position occupée dans le champ au moment précis de l'analyse, selon des critères propres à chaque champ), le plus haut indicateur de position (dans les champs médiatique et surtout politique, la montée des « échelons » n'est pas linéaire dans le temps), le parti politique

auquel l'agent est affilié (lorsque disponible), l'orientation politique de l'agent titulaire du compte et enfin le caractère politique ou non politique de sa production de tweets⁶⁷⁵.

Cette dernière variable mérite d'être ici plus longuement explicitée. Il peut en effet sembler paradoxal, voire hasardeux, de qualifier de « non politique » un compte qui publie (même occasionnellement) des tweets politiques, et qui sont de surcroît relayés par un nombre significatif d'utilisateurs. Par « non politique », nous entendons un compte dont les tweets politiques ne constituent qu'une fraction très marginale de sa production totale de messages. Nous allons le voir, ces comptes « non politiques » sont souvent ceux d'humoristes, d'artistes, de candidats de télé-réalité, de présentateurs d'émissions télévisées de divertissement ou encore de magazines thématiques (la presse féminine, sportive, etc.). Ces comptes ne publient que très occasionnellement des tweets politiques. Nous avons néanmoins fait le choix de les garder dans notre panel, car, comme le soulignait à juste titre Pierre Bourdieu à propos des journaux omnibus : « *l'effet proprement politique d'un journal [on pourrait ajouter d'un compte Twitter] ne se mesure pas à l'orientation politique de son discours proprement politique telle qu'elle peut être appréciée dans le champ de production idéologique et moins encore au nombre de centimètres carrés consacrés en propre à la politique [ou le pourcentage de tweets politiques dans la production totale de tweets], mais à la relation que les lecteurs entretiennent avec le journal, dont ils peuvent ignorer le message politique et dont l'action politique la plus importante peut être de n'accorder aucune importance à la politique* »⁶⁷⁶. C'est donc pour cette raison que ces comptes ont été conservés dans le panel.

Après avoir codé manuellement chacun des comptes du panel, nous les avons ensuite programmé dans le logiciel de collecte *AMI Opinion Tracker*, qui a archivé la totalité des tweets (politiques ou non) publiés par ces comptes. Avant de comparer cette production avec celle des panels « répondants » et « non répondants », nous allons rapidement évoquer la composition sociologique de ce panel « Twitter audience politique ».

⁶⁷⁵ Nous ne mentionnons pas ici les deux dernières variables (le nombre de vidéos consacrées au titulaire du compte Twitter sur les sites Youtube et Dailymotion), censées mesurer la notoriété de l'agent dans le monde « réel » en raison d'une trop grande imprécision méthodologique.

⁶⁷⁶ Pierre Bourdieu, *op. cit.*, p. 515.

3.2.1 La composition du panel « Twitter audience politique »

Le résultat sans doute le plus frappant lorsqu'on analyse la composition du panel « Twitter audience politique » est la prédominance des comptes issus du champ médiatique.

Tableau 4.19 Catégories des 1041 et des 200 comptes du panel « Twitter audience politique » connaissant la plus forte audience sur le réseau

Catégories	% (n = 1041)	% (n = 200)
Agents du champ médiatique	31,0	21,0
Médias	18,4	35,0
Agents du champ politique	14,1	14,5
Personnalités publiques	6,0	9,5
Organisations	5,2	6,0
Autres (producteurs indépendants)	25,3	14,0
Total	100,0	100,0

Source : Panel « Twitter audience politique », 2014

La moitié des comptes (producteurs de tweets politiques) qui bénéficient de la plus forte audience sur Twitter sont des comptes de médias (31 %) ou d'agents appartenant au champ médiatique (18 %). Ce taux s'élève même à 56 % si l'on recentre l'analyse sur les 200 comptes les plus « influents ». Dans cette même configuration, plus d'un tiers des comptes du panel sont alors des comptes de médias (35 %). Sur Twitter, la place des médias dans la production de l'opinion et de l'information politique est donc toujours centrale. Alors que, contrairement aux médias « traditionnels », les coûts de production, d'édition et de distribution sont quasi-nuls sur Twitter (et plus généralement sur internet), les petits producteurs « indépendants » d'opinions et d'informations politiques (blogueurs, journalistes de « petits » médias – fanzine, site internet, radio libre, etc.) restent néanmoins minoritaires au sein du panel « Twitter audience politique ». Ils ne représentent en effet qu'un compte sur quatre, et même un compte sur dix si l'on recentre l'analyse sur les 200 comptes les plus « influents ».

Ces résultats rendent compte néanmoins d'une relative « nouveauté » concernant le champ médiatique : une tendance, analogue à ce qu'a pu connaître le champ politique, à une

personnification, à une incarnation du champ (notamment de la presse écrite) dans des « personnalités » ; et avec pour corollaire une forme d'attachement et de délégation, de la part du public, à ces personnages « médiatiques ». Cette forme de délégation, voire cette *fides implicita*, peut même être parfois plus forte que celle qui lie le public au média pour lequel travaillent ces « personnalités ». Ainsi alors que Jean-Jacques Bourdin, journaliste et interviewer politique, figure à la 8^{ème} place de notre panel (273 401 abonnés sur Twitter au 6 mars 2013⁶⁷⁷), le compte Twitter de RMC, la radio sur laquelle il réalise son émission politique, ne figure qu'à la 204^{ème} place (37 092 abonnés). Dans le domaine de la presse écrite, Edwy Plenel (256 926 abonnés), président du site *Mediapart*, ou Christophe Barbier (263 802 abonnés), directeur de la rédaction de l'hebdomadaire *L'Express*, comptabilisent à eux seuls un nombre d'abonnés égal à 50% de celui de leurs journaux respectifs (531 063 abonnés pour *Mediapart*, 479 887 pour *L'Express*). Twitter permet donc de rendre saillante cette tendance à la personnification du champ.

Hormis cette prédominance du champ médiatique, un second résultat qui peut être mis en lumière concerne la présence non négligeable (10 % des 200 comptes les plus « influents ») de comptes de personnalités publiques (artistes, humoristes, chanteurs, candidats de télé réalité, etc.) et plus généralement de comptes que nous avons qualifiés de « non politiques » (32 % des comptes du panel et 34 % des 200 les plus « influents »). Nous l'avons dit, nous qualifions de « non politique » un compte dont les tweets politiques ne constituent qu'une part marginale de la production totale de messages. Ces comptes d'artistes ou ces comptes « non politiques » comptabilisent un nombre très élevé d'abonnés avec un profil sociologique très différent de celui des comptes de grands médias politiques (si nous n'avons pas eu les moyens de valider empiriquement cette hypothèse, nous pourrions apporter néanmoins certains éléments d'éclairage dans le prochain chapitre consacré aux logiques sociales et politiques d'abonnements) et leurs tweets politiques, certes peu nombreux, bénéficient d'un écho souvent considérable. Ainsi, figure à la 23^{ème} place du classement « Twitter audience politique », Youssoupha (@youssouphamusik), un rappeur français. Avec un nombre de 268 895 abonnés, il bénéficie d'une audience potentielle sur Twitter relativement équivalente à celle de Christophe Barbier, directeur de la rédaction de *L'Express* (263 802 abonnés). Une courte lecture des descriptions et des photographies de profil de ses abonnés laisse rapidement entrevoir un public sociologiquement très différent de celui de Christophe Barbier (beaucoup plus jeune et « populaire »). Et c'est cette différence de

⁶⁷⁷ Source : www.twitter.fr

composition sociologique du public qui explique sans doute qu'à nombre d'abonnés similaire, les tweets politiques de ce rappeur (tweets politiques certes très peu nombreux, il n'en a ainsi publié que quatre en 2012) soient légèrement plus retweetés en moyenne que ceux de Christophe Barbier. Il y a en effet fort à parier que les adolescents aient une plus forte propension à relayer des messages de leurs « idoles ».

Tableau 4.20 Les vingt comptes du panel « Twitter audience politique » bénéficiant de la plus forte audience sur le réseau

Rang	Nom du compte	Catégorie du compte	Sous-Catégorie	Caractère politique
1	@stephaneguillon	Personnalité publique	Humoriste	Politique
2	@Dailymotion	Média	Pureplayer	Politique
3	@lemondefr	Média	Presse	Politique
4	@ruquierofficiel	Agent médiatique	Présentateur TV/radio	Politique
5	@SophiaAram	Agent médiatique	Chroniqueur	Politique
6	@NicolasBedos1	Personnalité publique	Humoriste	Politique
7	@lequipe	Média	Presse	Non Politique
8	@JJBourdin_RMC	Agent médiatique	Présentateur	Politique
9	@LANDEYves	Autre	Twittos	Politique
10	@20minutes	Média	Presse	Politique
11	@liberation_info	Média	Presse	Politique
12	@bernardpivot1	Agent médiatique	Chroniqueur	Politique
13	@jmaphatie	Agent médiatique	Chef de service	Politique
14	@youssouphamusik	Personnalité publique	Artiste	Politique
15	@C_Barbier	Agent médiatique	Directeur de rédaction	Politique
16	@B_Giabiconi	Personnalité publique	Artiste	Non Politique
17	@lesinrocks	Média	Presse	Politique
18	@nk_m	Homme politique	Député	Politique
19	@jf_cope	Homme politique	Président Parti	Politique
20	@mouloudachour	Agent médiatique	Chroniqueur	Politique

Source : Panel « Twitter audience politique », 2014

Nous allons maintenant analyser la production de ce panel « Twitter audience politique » – et mesurer ainsi le niveau de l'« offre » d'opinions et d'informations politiques – et la comparer à la production de notre panel « répondants » – et mesurer ainsi le niveau de consommation de la « demande ».

3.3 Une corrélation entre « offre » et « demande » d'opinions et d'informations politiques

Afin de réaliser cette analyse comparée des deux panels, nous avons dû choisir une plage temporelle différente de celle que nous avons retenue dans la deuxième section de ce chapitre. Nous avons en effet comparé la production de nos deux panels sur une période de trois mois, du 1^{er} avril 2014 au 31 mai 2014⁶⁷⁸. Contrairement à la précédente période analysée (qui s'étendait du 1^{er} mars 2013 au 31 janvier 2014), cette séquence temporelle a été marquée par une forte politisation de l'actualité, avec notamment deux campagnes électorales (les élections municipales du 23 et 30 mars 2014 et l'élection européenne du 25 mai 2014) et la nomination d'un nouveau Premier ministre, Manuel Valls, le 31 mars 2014.

Notre principale interrogation est alors de savoir si les membres de notre panel « répondants » produisent davantage de messages politiques sur le réseau social lorsque le niveau de l'offre d'informations et d'opinions politiques est élevé ; niveau de l'offre que l'on peut mesurer sur Twitter en comptabilisant le nombre de tweets politiques publiés par les comptes d'agents et d'institutions appartenant au champ de production de l'opinion (le panel « Twitter audience politique »). Notre hypothèse est donc que le niveau de l'offre d'opinions et d'informations politiques détermine le niveau de la demande.

L'analyse statistique valide cette hypothèse. On enregistre en effet une corrélation significative et de très forte intensité ($r = 0,862$, $p = 0,000$) entre le nombre quotidien de tweets publiés par le panel « répondants » et le nombre quotidien de tweets publiés par le panel « Twitter audience politique ».

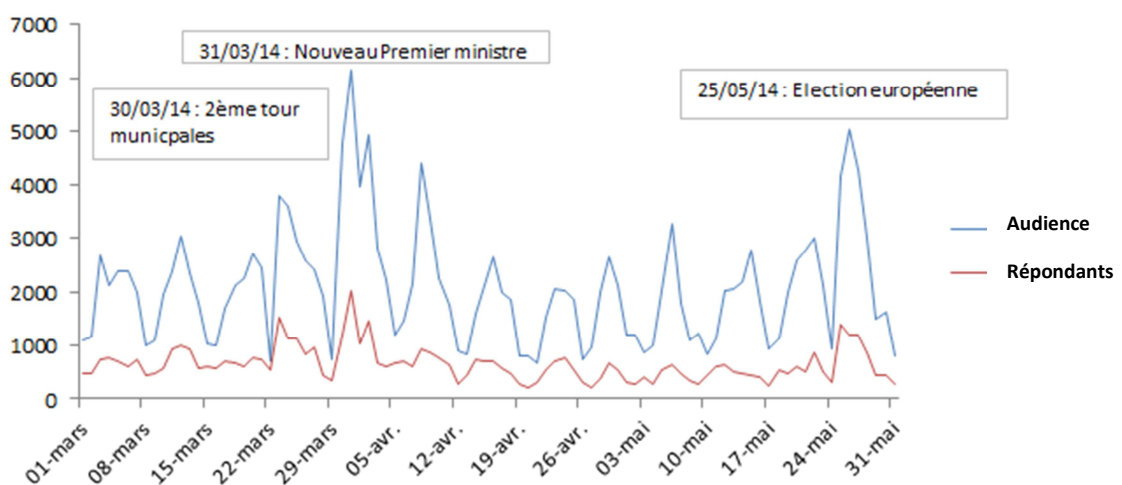
Tableau 4.21 Corrélation entre le nombre de tweets politiques publiés par les panels « Répondants » et « Audience politique » (n = 92 jours)

Variables	n =	r	valeur de p	p < 0,05 = *
Nombre de tweets Panel Répondants	92	0,862	0,000	*
Nombre de tweets Panel « Twitter audience politique »				

⁶⁷⁸ La programmation du panel « Twitter audience politique » dans le logiciel *Ami Opinion Tracker* s'est terminée en février 2014. La collecte des tweets n'étant pas rétroactive, nous n'avons donc pu comptabiliser la production de ce panel qu'à partir de mars 2014.

L'évolution quotidienne de la volumétrie des deux panels suit donc la même tendance. Le nombre de tweets politiques évolue de la même manière au sein des deux panels, à savoir parmi des individus très compétents politiquement mais extérieurs au champ de production de l'opinion (le panel « répondants », la « demande ») et parmi des individus qui appartiennent à ce champ (le panel « Twitter audience politique », l'« offre »).

Graphique 4.4 Évolution quotidienne des tweets politiques publiés par les panels « Répondants » et « Audience politique » du 01/03/2014 au 31/05/2014



Source : Panel « Twitter audience politique », 2014

Les pics de production de tweets politiques du panel « répondants » correspondent systématiquement à des pics de production du panel « Twitter audience politique ». Ces pics trahissent un niveau élevé de l'offre d'informations et d'opinions politiques disponibles sur le marché. Nous l'avons dit, Twitter synthétise l'offre de biens politiques disponible sur le marché de l'opinion et de l'information à un temps t . Dans leur grande majorité, les médias et les journalistes ne produisent pas des biens informationnels spécifiquement *pour* Twitter, mais relaient sur le réseau social leurs productions parues sur leurs supports respectifs (télévision, radio, presse écrite, presse numérique). Dès lors, constater un nombre élevé de tweets politiques publiés par le panel « Twitter audience politique » permet de postuler que l'offre de biens politiques disponibles sur le marché médiatique (télévision, radio, presse) connaît un niveau élevé. À la lecture de ce graphique, on constate donc bien que l'activité de la « demande » est corrélée à celle de l'« offre ».

Constater que le *nombre total* de tweets politiques publiés par le panel « répondants » est statistiquement corrélé à celui du panel « Twitter audience politique » ne permet pas en revanche d'affirmer que les biens consommés par la « demande » sont les mêmes que les biens produits par l'« offre ». Pour le dire plus simplement, on peut en effet se demander si les « répondants » et les « influents » parlent des mêmes sujets, ou plutôt des mêmes objets politiques, dans leurs messages. Pour répondre à cette question, nous avons comptabilisé, à l'aide de notre logiciel, les occurrences politiques les plus fréquentes au sein de la production de tweets politiques des deux panels.

Tableau 4.22 Occurrences politiques les plus courantes contenues dans les tweets politiques publiés par les panels « répondants » et « audience politique » du 01/03/2014 au 31/05/2014

PANEL AUDIENCE		PANEL REpondANTS	
<i>Occurrences politiques</i>	<i>Nombre de tweets</i>	<i>Occurrences politiques</i>	<i>Nombre de tweets</i>
Manuel Valls	13015	Manuel Valls	1083
François Hollande	9652	François Hollande	1076
Nicolas Sarkozy	3162	Jean-Luc Mélenchon	601
Jean-François Copé	3044	Jean-François Copé	398
Arnaud Montebourg	2528	Nicolas Sarkozy	330
Marine le Pen	1788	Arnaud Montebourg	200
Jean-Luc Mélenchon	1174	Marine le Pen	181
Michel Sapin	1111	Christiane Taubira	133
Christiane Taubira	1076	Michel Sapin	120
Jean-Marc Ayrault	1019	François Fillon	89
Européennes 2014	813	Henri Guaino	87
François Fillon	754	Jean-Marc Ayrault	67
Henri Guaino	624	Laurent Fabius	55
Laurent Fabius	595	Européennes 2014	53

On peut alors constater que non seulement les « répondants » et les « influents » évoquent les mêmes objets politiques dans leurs tweets, mais en « parlent » de plus quasiment dans les mêmes proportions. L'ordre des occurrences politiques les plus fréquemment citées est en effet sensiblement le même dans la production des deux panels (on peut d'ailleurs noter que ces occurrences sont très largement des noms propres d'hommes et de femmes politiques : ce résultat confirme donc bien une tendance à une très forte personnification du

champ politique sur Twitter). Les individus situés à l'extérieur ou à la périphérie du champ de production de l'opinion « parlent » donc des mêmes sujets politiques que les individus appartenant à ce champ.

En résumé, on s'aperçoit donc bien que les inégalités volumétriques de production du panel « répondants » sont fortement déterminées par les variations conjoncturelles de la production du champ de production de l'opinion et de l'information politique. On peut donc affirmer que le niveau de l'« offre » d'opinions et d'informations politiques sur Twitter détermine le niveau de la « demande ».

Conclusion du chapitre 4

L'objectif de ce chapitre était donc d'identifier les facteurs sociologiques et conjoncturels à l'origine des disparités de production de tweets politiques constatées à l'intérieur de – et entre – les panels « répondants » et « non répondants ».

À un niveau individuel, les inégalités volumétriques – très importantes – constatées entre les différents membres de nos panels s'expliquent en fonction de quatre principaux facteurs : la propension à produire un grand nombre de tweets (nombre total de tweets publiés), l'âge (les producteurs les plus âgés sont à la fois moins bien représentés dans notre échantillon mais également les plus gros producteurs de tweets politiques), l'intérêt *déclaré* et l'intérêt *objectif* pour la politique (mesuré en fonction du nombre d'abonnements à des comptes de femmes ou d'hommes politiques).

À un niveau diachronique, nous avons vu que les inégalités volumétriques quotidiennes de production des panels « répondants » et « non répondants » étaient fortement déterminées par la conjoncture macro-politique, qui pouvait être elle-même mesurée en analysant la production politique des comptes d'institutions et d'agents issus des champs médiatiques et politiques. Nous avons alors identifié ces comptes (à l'aide du panel « Twitter audience politique », produit en dehors du cadre de cette recherche doctorale), puis mesuré et comparé leur production de tweets politiques avec celle du panel « répondants ». Nous avons

ainsi enregistré une corrélation significative – et de très forte intensité – entre la production des panels « répondants » et « Twitter audience politique ».

Nous avons alors mobilisé le cadre analytique du marché, proposé par Pierre Bourdieu dans *La Distinction*, afin d’appréhender les facteurs conjoncturels à l’origine des disparités journalistiques de production de tweets politiques. La production des comptes issus du champ de production de l’opinion peut être alors considérée comme un indice mesurant « l’offre » d’opinions politiques et la production des panels « répondants » et « non répondants » comme un indice mesurant la consommation de la « demande ». La corrélation très forte mesurée entre les productions des panels «répondants » et « Twitter audience politique » nous a donc permis de conclure que le niveau de l’« offre » d’opinions sur Twitter détermine le niveau de la « demande ».

Néanmoins, on ne peut se contenter de mobiliser ces seuls résultats volumétriques si l’on souhaite démontrer que l’offre détermine la demande de tweets politiques. Si l’on décide d’appréhender analytiquement Twitter comme un marché de l’opinion, il convient alors de s’interroger plus finement sur les logiques sociales et politiques de consommation des tweets politiques et surtout d’exposer très clairement les différents types de biens échangés sur ce marché. En effet, si nous avons jusqu’ici évoqué, de manière relativement indifférenciée, les tweets politiques « consommés » comme des « opinions politiques » ou des « informations politiques », il convient désormais de dissocier clairement ces différents types de « biens » et partant de se demander dans quelle mesure ces « biens » peuvent être appréhendés – ou non – comme des marqueurs (directs ou indirects) d’opinion. Il s’agira là de l’objectif du chapitre 5.

CHAPITRE 5 – LES LOGIQUES SOCIALES ET POLITIQUES DE CONSOMMATION DES BIENS POLITIQUES SUR TWITTER

L'analyse comparée de la volumétrie des panels « répondants » et « audience politique » nous a conduit à appréhender Twitter comme un « marché » où se rencontrent une « offre » et une « demande » d'opinions et d'informations politiques. D'un côté, une offre composée d'environ un millier de comptes de médias et d'agents appartenant au champ de production idéologique. De l'autre, une « demande » émanant d'une centaine de milliers de comptes d'individus situés à l'extérieur de ce champ et dont nos panels « répondants » et « non répondants » constituent un échantillon représentatif.

Dans ce chapitre, nous allons analyser ce marché au prisme de la « demande » – c'est-à-dire de nos panels « répondants » et « non répondants » – en interrogeant plus particulièrement les logiques sociales et politiques de consommation des opinions et des informations politiques sur le réseau social. Si l'on appréhende – analytiquement – Twitter comme un marché, il convient alors de s'interroger sur les logiques qui orientent les individus – la « demande » – à consommer certains « biens » plutôt que d'autres. On peut ainsi se demander si ces logiques de consommation sont les mêmes que celles analysées par Pierre Bourdieu dans *La Distinction* à propos des choix – déclarés – en matière de lecture de journaux ou d'hebdomadaires marqués politiquement (titres de presse qui ont tous aujourd'hui un compte Twitter)⁶⁷⁹. Malgré la gratuité et l'immédiateté des informations disponibles sur Twitter, les individus « de gauche » s'abonnent-ils à des médias politiquement positionnés à gauche ? Les individus les plus diplômés retweetent-ils les mêmes messages que les individus les moins diplômés ?

Afin de répondre à ces questions, nous avons analysé trois types de données : les abonnements des membres de nos panels à des comptes de médias, de journalistes, d'hommes politiques, etc., leurs retweets politiques (lorsque l'auteur du message original appartient au champ de production idéologique) et les messages contenant des liens vers des sites de presse.

Nous avons mis en relation ces choix en matière de consommation d'opinions et d'informations politiques avec les propriétés sociologiques et politiques des membres de nos panels. Contrairement à la méthode employée dans *La Distinction*, les informations obtenues

⁶⁷⁹ Pierre Bourdieu, *La Distinction. Critique sociale du jugement*, op. cit., p. 515.

en matière de consommation médiatique ne sont pas ici déclarées par les enquêtés, mais enregistrées « passivement » par le chercheur. Le biais déclaratif, particulièrement problématique dans ce type d'étude – il suffit pour s'en convaincre de comparer le taux de Français déclarant regarder la chaîne de télévision *ARTE* avec son audience réelle – sera donc écarté de l'analyse.

Au-delà de cette neutralisation du biais déclaratif, notre analyse se différencie également sur un second point : les actes de consommation des opinions se trouvent, sur Twitter, précisément matérialisés, et peuvent alors être examinés « dans le détail » par le chercheur. Notre méthodologie permet en effet d'étudier les articles de presse ou les vidéos que les membres de nos panels choisissent ou non de diffuser auprès de leurs abonnés – après les avoir eux-mêmes « consommés » – en les retweetant ou en les relayant dans leurs tweets via des liens hypertextes par exemple⁶⁸⁰). Sur Twitter, le chercheur peut donc analyser « dans le détail » les opinions et informations (ostensiblement) consommées. Et contrairement aux travaux de la sociologie de la réception⁶⁸¹, ces informations peuvent être obtenues sans interagir avec les enquêtés, en dehors de toute situation de laboratoire. Le chercheur n'interroge pas en effet les enquêtés sur les articles qu'ils ont lus ou qu'ils les ont intéressés ou marqués, mais se contente d'observer les retweets que les enquêtés choisissent eux-mêmes de publier.

Cependant, dès lors que l'on analyse dans le détail les « biens » consommés sur ce marché de l'opinion, se pose la question suivante : ces différents types de « biens » sont-ils à proprement parler des opinions ? Titre d'article de presse commenté – ou non – par le lecteur, opinion d'un éditorialiste discutée ou relayée telle quelle, mot d'esprit d'un artiste à propos d'un homme politique ou blague potache publiée « en première personne » : sous l'appellation générique de « tweets politiques », se dissimule en réalité toute une variété de « biens » politiques de différentes natures, oscillants entre l'information et l'opinion. Nous tenterons alors d'esquisser une typologie des ces différents « biens », et de déterminer dans quelle mesure ils peuvent être appréhendés – ou non – comme des opinions ou des marqueurs d'opinion. Peut-on considérer qu'un individu que l'on sait de droite et qui retweete un éditorial du *Figaro* prenant parti contre le mariage homosexuel se positionne lui-même contre

⁶⁸⁰ Cf. glossaire p. 464.

⁶⁸¹ Voir notamment Brigitte Le Grignou, « Les études de réception : un mauvais objet ? », in Didier Georgakakis et Jean-Michel Utard (dir.), *Les Sciences des médias. Jalons pour une histoire politique*, Paris, 2001, p. 179-194 ou Brigitte Le Grignou, *Du côté du public : Usages et réceptions de la télévision*, Paris, Economica, 2003.

ce projet de loi ? À travers cet exemple, se trouve alors posé le problème de l'aspect non nécessairement formel et manifeste des opinions politiques circulant sur Twitter.

Après avoir analysé les logiques sociales et politiques d'abonnement et de consommation directe des « biens » politiques sur Twitter, *via* notamment l'examen des retweets et des liens hypertexte publiés par les membres de nos panels (1), nous nous demanderons dans quelle mesure ces différents types de « biens » consommés – directement ou indirectement – constituent des marqueurs d'opinion (2).

1. Les logiques sociales et politiques de consommation d'informations et d'opinions politiques

Sur Twitter, trois types d'information peuvent être mobilisés afin d'analyser la consommation d'opinions et d'informations politiques. La nature des comptes auxquels les individus sont abonnés sur le réseau social constitue un premier type d'information (1.1). Cependant, être abonné à un compte n'implique pas nécessairement de consulter – et donc de « consommer » – la production de ce compte. En revanche, retweeter un message publié par un journaliste ou un homme politique ou encore relayer dans un tweet un article de presse constituent des indices matériels d'une consommation *effective* et *directe* de biens politiques (1.2).

1.1 Les logiques sociales et politiques d'abonnement

Dès les premières étapes de leur inscription sur Twitter, les nouveaux utilisateurs sont invités à s'abonner à des comptes d'autres utilisateurs, déjà inscrits sur le réseau social. S'abonner à un compte d'utilisateur permet à un inscrit de recevoir les tweets publiés par cet utilisateur dans sa *timeline*, un espace d'édition où défilent en temps réel et selon un ordre

antéchronologique les tweets publiés par l'ensemble des comptes auxquels il est abonné. Un même individu peut ainsi s'abonner jusqu'à 2000 comptes différents⁶⁸².

Les individus peuvent schématiquement s'abonner à trois types de compte :

- des comptes d'individus concrètement connus et fréquentés, appartenant à l'univers de l'interconnaissance : amis, membres de la famille, (anciens) camarades de classe, collègues de travail, militants de section, etc.

- des comptes de « particuliers » qu'ils ne connaissent pas « réellement », en dehors de Twitter, mais dont ils ont pu trouver le profil ou les tweets intéressants, au hasard de leur navigation sur le réseau social.

- des comptes de médias (télévisés, radiophoniques, de presse écrite ou numérique), d'institutions politiques (partis, ministères, associations de défense d'une cause, etc.) ou non politiques (entreprises, clubs sportifs, etc.), ou encore de « personnalités » bénéficiant d'une notoriété plus ou moins importante (journalistes, hommes politiques, chanteurs, acteurs, humoristes, etc.).

C'est à cette troisième catégorie d'abonnements que nous allons ici nous intéresser. Par analogie aux recherches prenant pour objet les pratiques de lecture des titres de presse, les pratiques d'abonnements à des comptes de médias (ou d'agents appartenant au champ de production idéologique) peuvent être analytiquement appréhendées comme *une forme de consommation d'informations et d'opinions politiques*.

Cela étant dit, il convient d'emblée de préciser qu'il existe de nombreuses différences entre la lecture d'un journal papier et la consultation de tweets publiés par un compte de journal sur Twitter ; ces deux pratiques ne peuvent en aucun cas être considérées comme parfaitement similaires.

Comme l'a observé Guillaume Le Saulnier, internet « *redéfinit les régimes d'appropriation et les modes de lecture des supports d'actualité [et] représente un terrain propice pour la lecture extensive, [...] par le maillage de documents hétérogènes [qu'il permet] à la faveur de liens hypertextes* »⁶⁸³. Les liens hypertextes constituent précisément la base architecturale de la consommation d'informations sur Twitter : contraints par la taille des

⁶⁸² Ce pallier peut être dépassé à condition que l'individu ait lui-même en retour dépassé un certain seuil d'abonnés.

⁶⁸³ Guillaume Le Saulnier, « La lecture de la presse en ligne. L'appropriation des contenus d'actualité au défi de la technique », in Josiane Jouët et Rémy Rieffel (dir.), *S'informer à l'ère numérique*, Rennes, Presses Universitaires de Rennes, 2013, p.69.

140 caractères, les tweets publiés par les comptes de journaux se limitent à énoncer le titre de l'article et le lien hypertexte conduisant vers le site internet du journal. L'auteur observe alors que ces liens hypertextes conditionnent une « navigation » oscillant entre la « lecture didactique » (« *les liens hypertextes rencontrés au cours de la lecture sont autant d'invitations à approfondir les informations et à enrichir les connaissances* ») et le « survol erratique ». En effet, la navigation hypertextuelle « *suppose une capacité à relier entre elles des informations décontextualisées. Le lecteur qui progresse de lien en lien est obligé, à chaque fois, de reconstruire mentalement les éléments contextuels permettant la compréhension des textes* »⁶⁸⁴. Or cette capacité de contextualisation est inégalement distribuée dans la population et est déterminée par le niveau de capital scolaire : dépourvus de cette capacité, les individus les moins diplômés sont alors contraints à un « survol erratique » de l'actualité.

Au-delà de ces nouvelles modalités de consommation, une autre différence fondamentale entre la lecture de la presse écrite et la consultation de tweets « médiatiques » peut être ici soulignée. Elle concerne l'accessibilité, l'immédiateté et la gratuité des biens informationnels consommés sur Twitter. En effet, contrairement à l'achat d'un titre de presse, qui comporte des coûts de différentes natures (un coût financier, de plus en plus important, mais aussi un coût temporel de « déplacement » jusqu'à un kiosque de presse par exemple), l'abonnement à un compte de média sur Twitter est gratuit et immédiat. En moins d'une minute, un individu peut s'abonner et recevoir aussitôt dans sa *timeline* la production de l'ensemble des titres de la presse française ou internationale⁶⁸⁵.

Eu égard à cette neutralisation des coûts, à la multiplication de l'offre et aux nouvelles modalités de consommation des informations sur internet, on peut alors se demander dans quelle mesure les logiques sociales et politiques de consommation des « biens » médiatiques

⁶⁸⁴ *Ibidem.*

⁶⁸⁵ Il convient de préciser que la consultation de la presse en ligne n'est pas venue se substituer aux autres formes de consommation médiatiques préexistantes (télévision, presse écrite et radio). Ainsi, l'enquête de Fabien Granjon et Aurélien le Foulgoc démontre que la population qui consomme régulièrement des contenus d'actualité sur internet est celle qui consomme également le plus d'actualité sur les médias traditionnels (cf. Fabien Granjon et Aurélien le Foulgoc, « Les usages sociaux de l'actualité. L'expérience médiatique des internautes », *Réseaux*, n°160-161, 2010, p. 225-253). Viviane Le Hay et al. ont quant à eux identifié quatre catégories de mode de consommation de l'information politique, associées chacune à un profil sociologique type : une première centrée sur la télévision comme moyen d'information prioritaire des classes populaires, une deuxième centrée sur internet comme support privilégié par les moins de 34 ans qui ont un rapport distant à la politique, une troisième combinant presse écrite et télévision auprès des populations les plus âgées et enfin une quatrième combinant presse, internet et radio, privilégiée par les individus les plus diplômés et les plus aisés. Cf. Viviane Le Hay, Thierry Vedel et Flora Chanvril, « Usages des médias et politique : une écologie des pratiques informationnelles », *Réseaux*, n°170, 2011, p. 45-73. Pour une synthèse de ces travaux, cf. Josiane Jouët et Rémy Rieffel, *S'informer à l'ère numérique*, op. cit.

sont, aujourd'hui, sur Twitter, les mêmes que celles observées par Pierre Bourdieu dans les années 1970 à propos de la lecture des titres de presse écrite. Les propriétés politiques et sociologiques des individus déterminent-elles toujours leurs choix en matière de consommation médiatique ?

Afin de répondre à cette question, nous avons recensé et analysé, du 1^{er} au 20 mars 2013⁶⁸⁶, l'ensemble des comptes auxquels étaient abonnés les 1 148 individus qui composaient à ces dates⁶⁸⁷ nos panels « répondants » ($n = 558$) et « non répondants » ($n = 590$). Pour ce faire, nous avons relevé « manuellement »⁶⁸⁸ la liste exhaustive des comptes auxquels étaient abonnés chacun des membres de nos panels, soit, concernant le panel « répondants »⁶⁸⁹, un total de 197 377 abonnements (soit 354 abonnements en moyenne par individu). Derrière ce total de près de 200 000 abonnements, on comptabilise un nombre de 106 632 *comptes uniques*. Plusieurs individus peuvent être en effet abonnés à un même compte. Ce chiffre total de 197 377 abonnements comprend donc à la fois des abonnements de type « connaissances interpersonnelles », « connaissances virtuelles » et des abonnements à des médias, des institutions ou des personnalités publiques.

Afin d'isoler ce troisième type d'abonnements – celui qui nous intéresse directement ici – nous avons étudié la distribution du nombre de « répondants » abonnés à chacun de ces 106 632 comptes uniques et avons alors exclu les comptes qui n'avaient qu'un seul abonné (au sein de notre panel « répondants »). Ces comptes « à un abonné » n'étaient en effet pas des comptes de médias ou de personnalités publiques, mais des comptes d'individus appartenant au réseau d'interconnaissances des membres du panel. Ces comptes concentrent 41 % des abonnements totaux du panel. Nous avons également exclu les comptes qui avaient moins de 10 abonnés (toujours au sein de notre panel). Ces comptes étaient soit, eux aussi, des comptes de connaissance interpersonnelle, soit des comptes de blogs ou de « très petits » médias. Pour des raisons essentiellement pratiques, nous avons également décidé d'exclure ces comptes (ils représentent 38 % des abonnements totaux).

⁶⁸⁶ Le nombre d'abonnements est une donnée qui, comme le nombre de tweets ou le nombre d'abonnés, évolue en temps réel. Il convient donc de préciser que l'ensemble des résultats exposés ici ont été obtenus à ce moment t , soit un an avant la rédaction de ce chapitre.

⁶⁸⁷ Depuis la création du panel jusqu'à cette date, 80 individus ont supprimé ou désactivé leur compte Twitter.

⁶⁸⁸ Contrairement aux données empiriques présentées dans le chapitre 4, l'ensemble des données que nous mobiliserons dans ce chapitre ont été construites « manuellement », c'est-à-dire sans avoir recours à notre logiciel de veille, de collecte et d'archivage des tweets, et ce pour une raison assez simple : le logiciel ne contient aucun outil permettant de traiter les données d'abonnements et d'abonnés.

⁶⁸⁹ On enregistre un résultat relativement similaire concernant les « non répondants » : 191 546 abonnements (325 abonnements en moyenne par individu) à 116 991 de comptes uniques.

Finalement, nous avons conservé une liste de 1 855 comptes (qui concentrent 21 % des abonnements totaux des « répondants »). Ces comptes sont donc uniquement des comptes de médias, d'institutions politiques, de journalistes, d'hommes politiques ou encore de personnalités publiques.

*Tableau 5.1 Comptes auxquels les membres des panels sont le plus souvent abonnés
(les chiffres des colonnes Taux REP et Taux NREP sont en %)*

Class. REP	Nom du compte	Taux REP	Taux NREP	Class. NREP	Nb. total abonnés	Catégorie du compte
1	@lemondefr	39*	32	1	1 392 698	Média presse écrite
2	@Rue89	35	23	3**	432 641	Média presse online
3	@stephaneguillon	35	23	4	382 753	Humoriste politique
4	@fhollande	32	26	2	535 354	Homme politique
5	@liberation_info	27	20	5	494 817	Média presse écrite
6	@mediapart	27	18	6	304 473	Média presse online
7	@Maitre_Eolas	26	16	13	100 853	Blogueur
8	@humourdedroite	25	15	24	184 913	Compte humoristique
9	@valtrier	25	15	22	214 941	Homme politique
10	@BarackObama	24	18	9	28 990 226 ⁶⁹⁰	Homme politique
11	@afpfr	23	14	27	147 824	Média
12	@kyank	23	15	21	300 925	Humoriste non politique
13	@lesinrocks	23	16	15	341 302	Média presse écrite
14	@bernardpivot1	22	15	19	109 357	Journaliste non politique
15	@edwyplenel	21	15	23	138 210	Journaliste politique

Source : www.twitter.com

* 39 % des individus qui composent le panel « répondants » sont abonnés au compte @lemonde.fr

** Le compte @Rue89 est le troisième compte auxquels les « non répondants » sont le plus abonnés.

Lorsque l'on analyse les quinze comptes auxquels les membres du panel « répondants » sont le plus souvent abonnés, un premier résultat peut être rapidement mis en avant : la quasi-totalité de ces comptes appartiennent au champ de production idéologique. Ce résultat semble d'autant plus remarquable si on le compare aux comptes bénéficiant du plus grand nombre d'abonnés sur l'ensemble du réseau social. Sur les 10 comptes les plus « populaires », on ne retrouve qu'un seul homme politique (Barack Obama, président des

⁶⁹⁰ Contrairement aux autres comptes présents dans ce tableau, les abonnés au compte de Barack Obama proviennent de plusieurs pays, dont majoritairement les États-Unis.

États-Unis). Les autres titulaires de ces comptes sont quasi-exclusivement des chanteuses états-uniennes appartenant à l'industrie mondiale du disque.

Tableau 5.2 Comptes bénéficiant du plus important nombre d'abonnés sur Twitter (au 05/11/2013)

N°	Nom du compte	Catégorie du compte	Nombres abonnés
1	Katy Perry	Chanteur	46 599 517
2	Justin Bieber	Chanteur	46 546 451
3	Lady Gaga	Chanteur	40 412 893
4	Barack Obama	Homme politique	39 339 531
5	Taylor Swift	Chanteur	36 305 668
6	Britney Spears	Chanteur	33 794 310
7	Rihanna	Chanteur	32 529 938
8	Justin Timberlake	Chanteur	28 057 996
9	Jennifer Lopez	Chanteur	25 204 627
10	Ellen Degeneres	Animatrice TV	23 373 224

Source : www.twitter.com

Il convient néanmoins d'apporter à ces résultats une précision de taille : le nombre d'abonnés est ici calculé au niveau mondial, sans qu'il ne soit possible⁶⁹¹ de dissocier les abonnés selon leur nationalité. Cependant, le taux de pénétration du réseau social dans la population et la composition sociologique des inscrits semblent relativement similaire d'un pays à l'autre, notamment entre la France et les États-Unis⁶⁹². De plus, si l'on examine plus spécifiquement des comptes d'artiste populaires français, leur nombre d'abonnés est là aussi très sensiblement supérieur à celui des comptes auxquels les membres de nos panels sont le plus souvent abonnés. Pour ne citer qu'un seul exemple, le rappeur français Booba est « suivi » par 1 511 085 abonnés⁶⁹³, un score bien plus élevé que la majorité des comptes auxquels sont abonnés nos enquêtés (cf. tableau 5.1).

⁶⁹¹ Du moins à notre connaissance et à ce jour.

⁶⁹² D'après une étude du *Pew Research Center*, en 2013, 16 % des Américains déclareraient être « des utilisateurs de Twitter », un taux très proche de celui mesuré par l'IFOP en France en 2013 également (16 %). De plus, la composition sociale des utilisateurs américains et français semble également relativement proche, avec une nette surreprésentation des plus jeunes et des plus diplômés. Cf. *Pew Research Center* « Twitter News Consumers: Young, Mobile and Educated », novembre 2013 et IFOP, « Observatoire des Réseaux sociaux », 2013.

⁶⁹³ Source : www.twitter.com le 12 mars 2014.

La nature des abonnements des membres de nos panels est une preuve supplémentaire de la très forte politisation des producteurs de tweets politiques et de leur sentiment d'appartenance au « pays légal ». En effet, « *le fait de lire un journal national, et surtout un des grands journaux légitimes, comme le Figaro ou Le Monde, est une façon parmi d'autres (comme le fait d'écrire dans les journaux ou d'écrire aux journaux, [ou de publier des tweets politiques contenant des liens vers ces journaux]), de manifester que l'on se sent membres du pays légal, c'est-à-dire en droit et en devoir de participer à la politique, d'exercer vraiment ses droits de citoyen* »⁶⁹⁴.

Comparaison des abonnements des panels

La comparaison des abonnements des « répondants » et des « non répondants » peut également apporter une preuve de la – légère – différence de politisation entre ces deux panels. Nous l'avons vu dans le chapitre précédent, les « non répondants » produisent significativement moins de tweets politiques que les individus qui ont répondu à notre questionnaire, et plusieurs indices suggèrent que certains de ces « non répondants », et surtout ceux qui n'ont pas indiqué d'informations sociodémographiques sur leur profil, proviennent d'un milieu social plus modeste. La nature de leurs abonnements peut nous en apporter une preuve objective.

Les « répondants » et les « non répondants » sont abonnés à 958 comptes communs. Ces comptes, qui comptent parmi le plus d'abonnés dans chacun des panels, sont très majoritairement des comptes de médias, de journalistes, d'hommes politiques, etc. Les comptes auxquels sont le plus souvent abonnés les « non répondants » sont donc les mêmes que ceux des « répondants » et appartiennent, dans leur grande majorité, au champ de production idéologique. Le panel « non répondants » est donc également très politisé et la majorité des individus qui le composent se sentent également membres du « pays légal ».

Cependant, si l'on regarde maintenant les abonnements exclusifs de chaque panel, c'est-à-dire les comptes auxquels ne sont abonnés les membres que d'un seul des deux panels⁶⁹⁵, des différences significatives se font jour. Parmi les comptes « exclusifs » auxquels sont abonnés les « non répondants » (et auxquels aucun « répondant » n'est donc abonné), la très grande majorité sont des comptes de footballeurs et d'artistes de variété et de rap,

⁶⁹⁴ Pierre Bourdieu, *La Distinction, op. cit.*, p. 518.

⁶⁹⁵ 5 % des « répondants » peuvent être abonnés à un compte x alors que ce même compte x ne compte aucun abonné parmi les « non répondants », et inversement.

appartenant à la culture populaire (ou à la culture non-légitime). Ainsi, 5,2 % des « non répondants » sont abonnés au compte de @Sopranopsy4, un rappeur français populaire dont le compte Twitter totalise un nombre de 1 203 693 abonnés⁶⁹⁶. Pourtant, aucun des « répondants » n'est abonné à ce compte. Sur les 10 comptes exclusifs du panel « répondants », trois sont ainsi des comptes de rappeurs, cinq des comptes de chanteurs de variété ou de hip hop et un de footballeur (Christiano Ronaldo, qui a près de 25 millions d'abonnés sur Twitter).

Tableau 5.3 Comptes « exclusifs » auxquels sont abonnés les panels

Abonnements exclusifs NREP	Nature du compte	Taux NREP	Abonnements exclusifs REP	Nature du compte	Taux REP
@Sopranopsy4	Rappeur	5,2	@madmoiZelle	Blog de mode	5,0
@B_Giabiconi	Chanteur	4,7	@koztoujours	Blog politique	4,7
@BrunoMars	Chanteur	4,7	@LaFranceapeur	Humour politique	4,5
@NICKIMINAJ	Chanteur	4,6	@Ideboissieu	Journaliste politique	4,3
@wizkhalifa	Rappeur	4,4	@EnfantsDroits	Association cause	4,1
@aliciakeys	Chanteur	4,2	@a__k	Blogueur dessinateur	4,1
@Beyonce	Chanteur	4,2	@HuffingtonPost	Média presse online	4,1
@Drake	Rappeur	4,2	@labdemocrate	Blogs militants	4,1
@Cristiano	Footballeur	4,1	@LibeDesintox	Média presse écrite	4,1
@amel_bent	Chanteur	3,9	@linsurge_	Blogueur politique	4,1

À l'inverse, les comptes auxquels sont exclusivement abonnés les « répondants » appartiennent souvent au champ de production idéologique, et sont dans leur majorité des petits producteurs d'opinions (pas nécessairement politiques) indépendants, blogueurs ou « twittos ».

Ces résultats confirment donc qu'une minorité de « non répondants » présente un profil sociologique sensiblement différent de celui des « répondants », un profil plus « populaire », comme en atteste leurs préférences culturelles objectivées par leurs choix d'abonnements. Cependant ces individus ne constituent qu'une minorité ; la majorité des « non répondants » sont abonnés, eux aussi, à des comptes appartenant au champ de production idéologique.

⁶⁹⁶ Source : www.twitter.com, le 13 mars 2014.

Composition des abonnements

Si l'on s'intéresse maintenant aux 500 comptes auxquels les membres du panel « répondants »⁶⁹⁷ sont le plus souvent abonnés, on se rend compte que la composition de cet échantillon est relativement similaire à celle du panel « Twitter audience politique ». Précisons toutefois que les modalités méthodologiques de sélection des comptes qui composent le panel « Twitter audience politique » ne sont pas les mêmes que celles de cet échantillon de 500 comptes. Ne figurent en effet dans le panel « Twitter audience politique » que des comptes qui publient, même très occasionnellement, des tweets politiques. Ce qui n'est pas nécessairement le cas des comptes auxquels sont abonnés les « répondants » (même si une très grande majorité de ces comptes appartient au panel « Twitter audience politique » et publie donc, à des fréquences diverses, des tweets politiques).

Tableau 5.4 Comparaison des comptes du panel « Twitter audience politique » et des abonnements des « répondants »

Catégories de compte	Panel Twitter audience politique	Abo. Panel Répondants
Agents du champ médiatique	31	22,6
Médias	18,4	19,4
Agents du champ politique	14,1	16,8
Personnalités publiques	6,0	16,4
Organisations	5,2	7,0
Autres (producteurs indépendants)	25,3	16,4
Total	100	100

Source : Panel « Twitter audience politique », 2013

Cette différence méthodologique explique la plus grande proportion de « personnalités publiques » parmi les comptes auxquels sont abonnés les « répondants » : 16 % de ces comptes sont ceux de personnalités (artistes, sportifs, humoristes, etc.) contre seulement 6 % des 1041 comptes producteurs de tweets politiques bénéficiant de la plus forte audience sur l'ensemble du réseau social.

⁶⁹⁷ Afin de ne pas « saturer » notre démonstration, nous ne mobiliserons dorénavant que les données issues du panel « répondants ».

Hormis cette forte représentation de « personnalités publiques », on retrouve dans cet échantillon, comme au sein du panel « Twitter audience politique », une prédominance de comptes médiatiques (un compte sur cinq), d'agents issus du champ médiatique (23 % de l'échantillon), des comptes de femmes et d'hommes politiques (17 %) et des comptes de petits producteurs indépendants – blogueurs, twittos – (16 %) et d'organisations, politiques ou non (7 %).

Nous allons maintenant étudier plus spécifiquement les logiques sociales et politiques d'abonnement à deux catégories de comptes : les comptes de médias (1.1.1) et les comptes de femmes et d'hommes politiques (1.1.2).

1.1.1 Les abonnements à des comptes de médias

Les comptes de médias sur Twitter peuvent être classés selon trois principaux critères : leur caractère politique ou non politique (en fonction de la place relative accordée à la politique au sens étroit de *politics*), leur orientation politique (selon une dichotomie binaire gauche/droite) et enfin leur support originel (presse écrite, radio, télévision ou *pure player*⁶⁹⁸).

74 % des comptes de médias auxquels sont abonnés les membres de notre panel sont des médias que nous qualifions de « politiques » : *Le Monde*, *Le Nouvel Observateur*, *Rue89*, *BFMTV*, etc. L'actualité politique occupe en effet une place tout à fait centrale dans la production de ces médias. Ce résultat confirme la propension de nos enquêtés à s'abonner à des comptes dont les auteurs appartiennent au champ de production idéologique. Cela étant dit, les médias « non politiques » auxquels sont abonnés les « répondants » ne sont pas pour autant des titres totalement dépolitisés. Ces médias, pour la plupart thématiques, proposent souvent une grille de lecture politique ou sociologique de leur objet. On peut citer ici par exemple le magazine thématique *So Foot* qui propose une analyse sociale et culturelle du football (4 % des « répondants » y sont abonnés), ou encore le magazine féminin *Causette*, qui adopte une lecture politique des sujets habituellement traités par la presse féminine (également 4 % des « répondants » y sont abonnés). On remarquera d'ailleurs, à travers ces deux exemples, que ces magazines sont destinés à un public socialement très typé, diplômé et disposant d'un important capital culturel, nous aurons l'occasion d'y revenir.

⁶⁹⁸ Un *pureplayer* est un titre de presse n'ayant d'existence que sur internet, comme par exemple *Rue89*, *slate.fr* ou encore *Mediapart*.

Parmi les médias dont l'orientation politique peut être clairement établie⁶⁹⁹ (soit seulement 27 % des médias politiques examinés), les deux tiers sont des médias de gauche, comme principalement *Rue89* (35 % des enquêtés y sont abonnés), *Libération* (27 %) ou encore *Mediapart* (27% également) ; contre seulement un tiers de médias de droite, dont principalement *Le Figaro* (14 %), *Atlantico* (8 %) et *Le Point* (6 %).

Ces résultats nous rappellent la composition politique très déséquilibrée de notre panel et la très nette surreprésentation de l'électorat de gauche ; rappelons que 56 % des individus qui ont répondu à notre questionnaire se situent « très » (19 %) ou « plutôt » (37 %) à gauche, contre seulement 23 % se situant « très » (4 %) ou « plutôt » (19 %) à droite. On peut donc logiquement supposer que la nette sur-représentation de l'électorat de gauche explique des taux d'abonnement plus élevés à des médias positionnés à gauche.

Cette hypothèse peut être vérifiée empiriquement en analysant la distribution des abonnements en fonction de l'orientation politique déclarée des « répondants ».

Tableau 5.5 Taux d'abonnement (en %) des « répondants » à des comptes de médias en fonction de leur orientation politique déclarée (n = 558)

Orientation abonné	<i>Le Monde</i>	<i>Libération</i>	<i>Mediapart</i>	BFMTV	<i>Le Figaro</i>
Très à gauche (n = 106)	45*	36	44	15	11
Plutôt à gauche (n = 206)	39	34	30	15	9
Au centre (n = 56)	40	26	17	21	17
Plutôt à droite (n = 106)	39	17	16	22	26
Très à droite (n = 22)	35	4	17	17	30
Ne se positionne pas (n = 62)	31	16	18	16	6
Total Panel (n = 558)	39	27	27	17	14

* 45% des « répondants » se déclarant être « très à gauche » sont abonnés au compte Twitter du *Monde*

Alors que l'abonnement à un compte Twitter est gratuit et immédiat, on observe, à la lecture de ce tableau, que les médias conservent sur Twitter un lectorat politiquement très typé et surtout très similaire à celui de leur support d'origine. Dit autrement, sur Twitter, malgré les nouvelles modalités de consommation de l'information que nous avons décrites plus haut

⁶⁹⁹ L'orientation politique des médias n'a pas été établie sur critères totalement objectifs. Nous nous sommes basé avant tout sur des présupposés communément partagés (*Libération*, journal de gauche, *Le Figaro*, quotidien de droite, *Le Monde* « neutre », etc.).

(gratuité, immédiateté, abondance de l'offre, etc.), l'orientation politique des individus détermine toujours très clairement leurs choix en matière d'abonnements à des médias politiques. Ce résultat, qui n'est peut-être pas surprenant, est pourtant, à notre connaissance, inédit. En effet, si le nombre d'abonnés des comptes Twitter est une donnée souvent mobilisée (car très facilement accessible), la composition sociologique et politique de ces « abonnés » ou « followers » est en revanche un point aveugle des travaux de recherche. Notre méthodologie de panélisation sociologique des inscrits sur le réseau social permet alors, par un renversement méthodologique⁷⁰⁰, de qualifier politiquement et socialement l'audience de ces comptes.

On constate ainsi que seulement 11 % des individus qui se déclarent « très à gauche » sont abonnés au compte du *Figaro* alors qu'ils sont 36 % à être abonnés à celui de *Libération*. Et inversement, alors que 30 % des individus qui se déclarent « très à droite »⁷⁰¹ sont abonnés au *Figaro*, seuls 4 % sont abonnés à *Libération*. Les logiques d'abonnement aux comptes de médias politiques sont donc bien déterminées politiquement.

Cependant, si on analyse maintenant la totalité des comptes de médias politiques auxquels sont abonnés les individus sur Twitter, on se rend compte d'une part que la majorité des enquêtés sont abonnés à plusieurs médias politiques, et que ces médias sont d'autre part politiquement hétérogènes.

Les membres de notre panel sont en effet abonnés à un nombre important de médias politiques sur Twitter, et semblent donc consommer bien davantage de biens médiatiques en ligne que hors-ligne. En moyenne, les producteurs de tweets politiques sont ainsi abonnés à 10 médias politiques sur Twitter. La gratuité et l'immédiateté de l'acte d'abonnement sur le réseau social semble donc augmenter sensiblement le nombre de biens médiatiques consommés. Tout comme le nombre de tweets politiques publiés, le nombre d'abonnements à des médias politiques varie significativement en fonction de variables sociologiques et politiques, telles que l'intérêt déclaré pour la politique ($F = 28,784$, $r = 0,000$) ou plus encore le fait d'adhérer à un parti politique ($F = 32,089$, $r = 0,000$). Les individus qui s'intéressent

⁷⁰⁰ Pour être tout à fait précis, nous n'analysons pas l'orientation politique des abonnés des comptes de médias mais nous analysons les abonnements des individus composant un échantillon représentatif dont on connaît les caractéristiques politiques et sociologiques (comme les études d'audiences de *Médiamétrie* par exemple, réalisées à partir d'un échantillon représentatif de téléspectateurs). Cependant, une réserve peut être formulée : notre échantillon est représentatif des individus qui publient des tweets politiques, mais pas de ceux qui n'en publient pas. Or cette deuxième catégorie d'utilisateurs s'abonne elle aussi à des comptes de médias. Or on ne sait pas si leur logiques d'abonnements sont les mêmes que celles des producteurs de tweets politiques.

⁷⁰¹ Les effectifs faibles de cette sous-catégorie ($n = 22$) appellent bien sûr à une grande prudence interprétative.

« beaucoup » à la politique sont ainsi en moyenne abonnés à 13 médias politiques contre 4 pour ceux qui ne s’y intéressent « peu » ou « pas du tout » ; les militants sont quant à eux abonnés à un nombre de médias politiques deux fois supérieur (14 abonnements en moyenne) à celui des non-militants (7 abonnements). En revanche, l’orientation politique déclarée ne détermine pas significativement le nombre de médias politiques consommés ($F = 2,490$, $r = 0,084$). Être de gauche ou de droite n’implique pas de consommer plus – ou moins – de biens médiatiques politiques.

Si elle ne détermine pas le nombre *total* de médias politiques auxquels sont abonnés les producteurs de tweets politiques, l’orientation politique des enquêtés détermine en revanche le nombre d’abonnements spécifiques à des médias *de gauche* ($F = 16,312$, $r = 0,000$) et *de droite* ($F = 8,974$, $r = 0,000$).

Tableau 5.6 Nombre moyen d’abonnements à des médias politiques selon l’orientation politique de l’enquêté

Orientation enquêté	Nb médias politiques ⁷⁰²	Nb médias de gauche	Nb médias de droite
À gauche ($n = 314$)	10,3	3,5	0,6
À droite ($n = 180$)	9,2	1,8	1,2
Ne se positionne pas ($n = 64$)	6,4	1,6	0,6
Total ($n = 558$)	9,5	2,8	0,8

Qu’ils soient de droite ou de gauche, les producteurs de tweets politiques sont donc abonnés à des médias politiques qui ne sont pas exclusivement associés à leur tendance politique (rappelons néanmoins que seuls les individus les plus politisés, tels que le sont les membres de notre panel, sont en capacité d’identifier la ligne politique des journaux qu’ils consomment). Ce résultat rend compte encore une fois de la très forte politisation des membres de notre panel. Certains d’entre eux justifient ainsi le fait d’être abonnés à des journaux de différentes tendances politiques comme la preuve d’un exercice effectif de leur devoir civique, conforme au modèle idéalisé du citoyen rationnel, parfaitement informé et capable ainsi de se faire « sa propre opinion » et d’agir politiquement *en toute connaissance de cause*. À la lecture de ce tableau, on remarque cependant que ce sont les individus

⁷⁰² Le nombre de médias politiques comprend le nombre d’abonnements à des médias de gauche et de droite mais aussi à des médias « neutres » (*Le Monde*, *BFMTV*, etc.), ce qui explique que la somme médias de gauche + médias de droite ne soit pas égale au nombre de médias politiques.

positionnés à gauche qui ont, en moyenne, le plus grand nombre d'abonnements à des médias de gauche (3,5 abonnements en moyenne contre 1,8 pour les individus positionnés à droite), et inversement les individus de droite qui ont, en moyenne, le plus grand nombre d'abonnements à des médias de droite (1,2 abonnements, soit deux fois plus que les individus de gauche, avec 0,6 abonnements).

On peut remarquer également que les individus positionnés à gauche sont abonnés en moyenne à six fois plus de médias de gauche que de droite. En revanche, la réciproque n'est pas exacte. Les individus positionnés à droite sont abonnés en moyenne à plus de médias de gauche (1,8 abonnements) que de droite (1,2 abonnements), même si, nous l'avons dit, ce sont eux qui ont comparativement le plus grand nombre d'abonnements à des médias de droite (1,2 contre 0,6 pour les individus de gauche). Si ce résultat peut sembler surprenant, voire paradoxal, il peut en réalité s'expliquer par la structure globale de l'offre de médias politiques, et la plus faible présence de médias que l'on positionne communément à droite. Dit autrement, si les individus de droite sont abonnés à plus de titres de gauche que de droite, c'est peut être avant tout parce que, parmi l'offre totale de titres disponibles, les titres de droite sont significativement moins nombreux que ceux de gauche, et que sur Twitter, l'immédiateté et la gratuité de l'information conduisent les individus les plus politisés, qu'ils soient de droite ou de gauche, à consommer beaucoup de titres politiques.

Au-delà du caractère et de l'orientation politique des médias auxquels sont abonnés les individus sur Twitter, une autre grille de lecture possible consiste à analyser le support originel de ces comptes médiatiques (presse écrite, radio, télévision ou presse exclusivement en ligne) et questionner leur légitimité dans la hiérarchie des biens culturels. On peut en effet se demander dans quelle mesure internet impacte ou non la hiérarchie des biens médiatiques socialement *légitimes*. Les individus bénéficiant d'un important capital culturel consomment-ils toujours d'avantage sur Twitter des titres de presse écrite plutôt que des médias télévisés ? Les producteurs les moins diplômés ont-ils au contraire davantage tendance à s'abonner à des comptes de médias télévisés pour s'informer sur Twitter ? Où s'inscrivent les sites *pure player* dans cette hiérarchie culturelle des biens médiatiques ?

Avant de répondre à ces questions, quelques éléments généraux de contexte peuvent être ici rapidement exposés. Parmi les 195 comptes de médias auxquels sont le plus souvent abonnés les membres du panel « répondants », 39 % sont des titres de presse n'ayant d'existence que sur internet (des titres « *pure players* »), 34 % des titres de presse écrite, 21 %

sont des comptes de médias télévisés et seulement 6 % des comptes de radio. On peut donc remarquer d'une part la faiblesse relative des médias radio-télévisés dans la consommation médiatique des membres de notre panel sur Twitter, et constater d'autre part la prédominance des *pure players* (que l'on peut également qualifier de titres de presse exclusivement *online*). Ainsi, nous avons vu plus haut (cf. tableau 5.1) que le compte Twitter de *Rue89*, avec un taux d'abonnés, au sein de notre panel de 35 %, était le deuxième compte, derrière celui du *Monde*, auxquels nos enquêtés étaient, tous types de comptes confondus, le plus souvent abonnés ; ou encore que le compte du site *Mediapart*, avec un taux d'abonnés de 27 %, était le sixième compte comptabilisant le plus d'abonnés parmi les « répondants ».

Tableau 5.7 Support originel des comptes de médias auxquels les « répondants » sont le plus souvent abonnés

Support originel	Part relative dans l'échantillon (n = 195)
<i>Pure player</i>	39 %
Presse écrite	34 %
Radio	6 %
Télévision	21 %
Total	100 %

Ces résultats semblent donc confirmer l'hypothèse évoquée dans le chapitre 3 selon laquelle les producteurs de tweets politiques présentent un profil proche de celui des usagers en « tâche de fond » décrits par Jean-Baptiste Comby⁷⁰³. Ces individus, issus des classes supérieures, s'informent notamment *via* Twitter et consultent de nombreux *pure players*, contrairement aux « usagers conventionnels », qui « visitent quasi exclusivement LeMonde.fr, Libération.fr, LeFigaro.fr, c'est-à-dire des sites des grands médias et notamment des quotidiens disposant au préalable d'une place légitime dans le champ de la presse écrite. La plupart [des « usagers conventionnels »] ignorent les *pure players* et ceux qui en connaissent, n'évoquent que les plus réputés d'entre eux : Médiapart, Rue 89, Slate ». Pour l'auteur, ces différences entre ces deux catégories d'individus s'expliqueraient comme suit :

⁷⁰³ Jean-Baptiste Comby *et al.*, « Les appropriations différenciées de l'information politique en ligne au sein des catégories sociales supérieures », art. cit.

« *Le parcours social ascendant* [des usagers en « tâche de fond », qui, contrairement aux usagers « conventionnels », seraient des « nouveaux venus » au sein des classes supérieures] *pourrait également expliquer leur moindre attachement aux médias « traditionnels ».* Ne peut-on pas en effet émettre l'hypothèse que, socialisés dans des univers où les normes de la légitimité culturelle et médiatique étaient sans doute moins manifestes que dans les milieux sociaux « supérieurs », ils soient plus disposés à percevoir et à reconnaître les profits de distinction liés aux nouveautés technologiques ? Leurs critères de la bonne information ne tiendraient plus tant aux lieux de sa production qu'à sa valeur d'immédiateté ou d'altérité. À l'inverse, il semblerait que les usagers « conventionnels » perpétuent des rapports traditionnels aux médias qui ont été valorisés au sein de leurs groupes de pairs. Pour eux, les entreprises reconnues et légitimes de fabrication et de diffusion de l'information constitueraient toujours une garantie forte de sa qualité. »⁷⁰⁴

Afin de tester cette hypothèse sur notre propre population, nous avons examiné précisément à quel type de support chaque individu était en *majorité* abonné⁷⁰⁵, puis avons croisé cette variable (qui se décline en 4 modalités : « est abonné majoritairement à des *pure players* », « est abonné majoritairement à des titre de presse écrite », « est abonné majoritairement à des comptes de TV » et « est abonné majoritairement à des comptes de radio ») avec la PCS des « répondants » (nous avons exclu la catégorie « agriculteurs, artisans, commerçants et chefs d'entreprise » en raison d'effectifs trop faibles, $n = 5$).

⁷⁰⁴ Art. cit.

⁷⁰⁵ On peut ainsi établir que l'individu x1 est abonné à 100 comptes de médias, et que sur ces 100 comptes, 60 sont des comptes de presse écrite, 30 de presse online, 7 de télévision et 3 de radio.

Tableau 5.8 Origine des comptes médiatiques majoritairement privilégiés selon la PCS (en %)

PCS	Presse	Online	TV	Radio	Total
Cadres du privé (<i>n</i> = 64)	47	42	6*	5	100
Cadres du public (<i>n</i> = 50)	58	38	4	0	100
Professions intermédiaires du privé (<i>n</i> = 29) ⁷⁰⁶	28	66	7	0	100
Professions intermédiaires du public (<i>n</i> = 29)	48	45	7	0	100
Employés et ouvriers (<i>n</i> = 27)	35	39	26	0	100
Retraités et inactifs (<i>n</i> = 24)	42	38	13	8	100
Étudiants (<i>n</i> = 146)	55	26	18	1	100
Lycéens (<i>n</i> = 26)	42	27	31	0	100
Total (<i>n</i> = 395) ⁷⁰⁷	48	37	14	2	100

** 6 % des cadres du privé comptent, parmi l'ensemble de leurs abonnements à des comptes de média, une majorité d'abonnements à des comptes de radio

À la lecture de ce tableau, on constate d'abord que pour une majorité de nos enquêtés, la presse écrite reste le support majoritairement privilégié pour s'informer sur Twitter. Un « répondant » sur deux comptabilise en effet, parmi l'ensemble de ses abonnements à des comptes de média, une majorité de titres de presse écrite.

On peut ensuite remarquer que la position sociale des enquêtés (mesurée par la PCS) impacte sensiblement les types de comptes médiatiques majoritairement privilégiés pour s'informer sur le réseau social. Toutefois, avant de développer plus avant ce résultat, il convient de rappeler ici le caractère sociologiquement atypique de notre population. Comme nous l'avons évoqué dans notre troisième chapitre, les ouvriers, employés, commerçants et retraités qui composent notre panel présentent des propriétés tout à fait singulières, et ne sont pas représentatifs de leurs catégories sociales d'appartenance. Ils sont ainsi notamment plus informés et surtout plus politisés. Les différences entre catégories sociales que nous allons analyser par la suite ne sont donc pas extrapolables à l'ensemble de la population française, et concernent donc uniquement notre population de producteurs de tweets politiques.

D'une manière générale, on remarque ainsi que plus on s'élève dans la hiérarchie sociale et plus on a tendance à privilégier (majoritairement) la presse écrite ; au contraire, plus

⁷⁰⁶ Ces effectifs faibles invitent à une grande prudence dans l'interprétation des résultats. Cependant, nous avons tout de même choisi de ne pas agréger les cadres du privé avec ceux du public (*idem* pour les professions intermédiaires) en raison des différences privé/public qui nous semblent particulièrement structurantes.

⁷⁰⁷ À noter que 96 individus ont un nombre d'abonnements (à un ou plusieurs types de média) similaire. Exemple : l'individu x1 a 5 abonnements à des médias *radio* et 5 abonnements à des médias TV. Afin de ne pas complexifier l'analyse en introduisant des combinaisons de supports majoritaires, nous avons momentanément exclu ces individus. Enfin, précisons que 62 individus (11% du panel) ne sont abonnés à aucun compte de média.

on descend dans la hiérarchie sociale et plus on a tendance à privilégier une combinaison presse *online*/télévision (même si la télévision n'est majoritaire pour aucune PCS). On pourrait ainsi schématiquement résumer ces résultats de la manière suivante : cadres (presse écrite++ ; presse *online*+), professions intermédiaires (presse *online*++, presse écrite+), classes inférieures⁷⁰⁸ (presse écrite+ ; presse *online*+ ; TV+).

Les retraités et les étudiants se situent quant à eux dans une sorte « d'entre deux » : comme les classes supérieures, leurs abonnements médiatiques sont majoritairement orientés vers des titres de presse écrite et, comme les classes inférieures, les comptes de médias télévisés y occupent une place non négligeable.

Enfin, au sein même du groupe des cadres – mais aussi et peut-être surtout au sein de celui des professions intermédiaires – on peut observer des différences sensibles (même si, rappelons-le, les effectifs sont très faibles et invitent donc à une certaine prudence interprétative) entre les salariés du public et du privé, qui trahit en fait une différence en terme de structure de capital, avec du côté du public, une prédominance du capital culturel sur le capital économique et, inversement, du côté du privé, une prédominance du capital économique sur le capital culturel. On remarque ainsi par exemple que les professions intermédiaires du privé ont davantage tendance à orienter leurs abonnements vers la presse *online* tandis que ceux du public privilégient la presse écrite.

Tout se passe comme si plus les producteurs de tweets politiques avaient un niveau élevé de capital culturel, et plus ils avaient tendance à privilégier, comme les usages « conventionnels » décrits par Comby, « *les entreprises reconnues et légitimes de fabrication et de diffusion de l'information* », à savoir les titres de la presse écrite « traditionnelle ».

Deux objections peuvent être néanmoins formulées à ce résultat. La première est que, s'informant *via* Twitter, l'ensemble de nos enquêtés développent un rapport à l'information qui, par définition, se rapproche davantage de celui des usagers en « tâche de fond ». La seconde objection, plus fondamentale, concerne la différence au final relativement ténue entre les comptes de titres de presse écrite et les comptes de presse *online*.

On peut en effet se demander si, plutôt que d'opposer, sur une échelle de légitimité culturelle, la presse écrite et la presse *online*, il ne conviendrait pas davantage de hiérarchiser,

⁷⁰⁸ Rappelons encore une fois qu'en raison de la faiblesse des effectifs, ces résultats invitent à la plus grande prudence. Les résultats esquissés dans ces paragraphes doivent être davantage conçus comme des hypothèses et des pistes de réflexion à approfondir à l'avenir que comme des résultats « définitifs ».

au sein même de chaque support (presse écrite, presse *online*, télévision et radio), la légitimité culturelle des différents titres. Dit autrement, on peut se demander si l'opposition, au sein du champ de la presse écrite, entre la presse à sensation (de type *Le Parisien*) et la presse de réflexion (de type *Le Monde*) ne pourrait pas s'appliquer aux différents titres de la presse exclusivement *online*, entre par exemple *Mediapart* et le blog de Jean-Marc Morandini.

N'ayant pas ici les moyens de mener une entreprise d'une telle envergure, nous nous contenterons d'étudier les taux d'abonnements par PCS à différents comptes médiatiques, trois plutôt situés « en haut » de l'échelle de légitimité culturelle (*Mediapart* pour la presse *online*, *France inter* pour la radio et *ARTE* pour la télévision), et deux « en bas de l'échelle » (*BFMTV* pour la télévision et *20 minutes* pour la presse écrite).

Tableau 5.9 Taux d'abonnement (en %) à des médias par PCS

PCS	<i>Mediapart</i>	<i>20 minutes</i>	<i>ARTE</i>	<i>BFMTV</i>	<i>France inter</i>
Cadres privé (<i>n</i> = 88)	25*	21	14	16	22
Cadres public (<i>n</i> = 68)	39	14	22	16	30
Prof. inter. privé (<i>n</i> = 40)	26	23	13	14	15
Prof. inter. Public (<i>n</i> = 46)	47	8	23	8	33
Employés/ ouvriers (<i>n</i> = 42)	17	12	7	24	15
Retraités et inactifs (<i>n</i> = 32)	35	13	10	3	13
Étudiants (<i>n</i> = 197)	27	22	17	24	23
Lycéens (<i>n</i> = 45)	5	5	7	7	2
Total (<i>n</i> = 558)	27	17	15	17	21

*25 % des cadres du privé sont abonnés au compte de *Mediapart*

On remarque alors que la position sociale des producteurs de tweets politiques (et notamment leur volume et leur structure de capital) détermine assez nettement leurs choix en matière d'abonnements à des biens médiatiques sur Twitter. Sur le réseau social comme ailleurs, les choix en matière de consommation médiatique ne sont en effet pas dus au hasard mais sont au contraire bien orientés socialement.

Comme nous avons commencé à l'entrevoir à propos de la nature des supports médiatiques privilégiés, la dichotomie public/privé – ou capital culturel/capital économique – semble particulièrement opérante concernant la distribution des choix en matière

d'abonnements à des comptes médiatiques sur Twitter. Le taux d'abonnement des cadres du public à *Mediapart* est ainsi de 14 points supérieur (39 %) à celui des cadres du privé (25 %). Le même phénomène est observable concernant les professions intermédiaires, avec un différentiel encore plus marqué : 19 points séparent le taux d'abonnement du public (47 %) par rapport au privé (17 %). Sur un bien médiatique au contraire délégitimé culturellement, comme le compte du journal gratuit *20 minutes*, le taux d'abonnement des cadres du privé est à l'inverse sensiblement plus élevé (21 %) que celui des cadres du public (14 %).

Malgré l'inflation de l'offre (en termes de supports – avec l'apparition de la presse *online* – mais aussi au sein de chaque support – multiplication des chaînes de télévision, etc.) et les nouvelles modalités de consommation de l'information sur Twitter (gratuité, immédiateté), les choix en matière d'abonnements à des comptes médiatiques restent toujours déterminés socialement (en fonction notamment du volume et de la structure du capital culturel et économique) et politiquement.

Nous allons maintenant voir si ces mêmes logiques s'appliquent à un autre type d'abonnement, celui-ci relativement inédit : l'abonnement à des comptes de femmes et d'hommes politiques.

1.1.2 Les logiques d'abonnement aux comptes de femmes et d'hommes politiques

Contrairement au fait de s'abonner à un compte de média – qui peut s'apparenter simplement à une nouvelle forme de consommation médiatique – s'abonner au compte Twitter d'une femme ou d'un homme politique constitue une pratique sociale relativement « inédite ». Il n'existe pas en effet selon nous d'équivalent fonctionnel ou symbolique à cette pratique dans la « sphère hors-ligne ». S'abonner au compte d'un homme politique – avec pour conséquence de recevoir dans sa *timeline* les messages publiés

(in)directement⁷⁰⁹ par cet homme politique – ne peut être assimilé à une forme de militantisme ou à une relation classique d'élus à administrés. Les significations que les individus attribuent à cette pratique sociale sont d'une part plurielles et sont d'autre part déterminées par leur position sociale, leur niveau de politisation et plus largement leur rapport à la politique.

C'est avant tout l'investissement des réseaux sociaux par les professionnels de la représentation politique qui a permis la création de cette nouvelle forme de communication. Twitter, et plus largement internet, constituent en effet pour ces professionnels un nouvel outil de communication qui autorise une communication « directe » avec les « administrés » (ou les « citoyens »), c'est-à-dire une communication affranchie de la médiation des appareils partisans et surtout de celle des journalistes et des médias. Sur Twitter, un homme politique peut en effet publier seul et sans intermédiaire des messages qui seront potentiellement lus et relayés par des dizaines voire des centaines de milliers d'individus, dont des journalistes, qui pourront ainsi rapporter les propos de l'homme politique sans que celui-ci ait eu besoin d'être – ou de se faire – invité des plateaux télévisés.

L'arrivée des hommes politiques sur Twitter – et avec pour corollaire l'apparition d'un nouveau public : leurs « abonnés » – participe d'un phénomène en réalité plus vaste, à savoir celui de l'« *affaiblissement relatif du crédit accordé aux partis politiques et la tendance au renforcement de la personification des camps politiques [qui ont favorisé] la multiplication d'attachements et de comportements de délégation à des personnages politiques de premier rang, souvent des dirigeants de partis [et plus largement des professionnels qui ont investi le champ de la télévision]*⁷¹⁰. La question est alors de savoir si l'on peut considérer les abonnements à ces comptes de femmes et d'hommes politiques comme une nouvelle forme de délégation politique ?

La réponse apportée à cette question est en réalité tributaire du type de données empiriques mobilisées. Si l'on s'en tient aux significations fournies par les enquêtés, telles qu'on peut les appréhender durant des entretiens qualitatifs, on serait alors tenté de répondre que l'abonnement et la consultation de comptes d'hommes politiques trahissent moins une forme de délégation qu'une source supplémentaire d'information politique.

⁷⁰⁹ Certains comptes ne sont pas gérés directement par l'homme politique (comme le compte de Jean-Luc Mélenchon, dont les tweets sont rédigés et publiés par un subordonné).

⁷¹⁰ Daniel Gaxie, « Retour sur les modes de production des opinions politiques », art. cit.

Les individus les plus politisés de notre échantillon déclarent ainsi être abonnés à des comptes d'hommes et de femmes politiques qui ne sont pas nécessairement issus de leur bord politique et avec qui ils ne partagent pas les mêmes idées. Certains légitiment cette volonté affichée de « pluralisme » de manière relativement analogue au fait d'être abonné à des journaux de bords politiques opposés, comme la preuve d'un devoir civique consistant à se tenir informer en confrontant des sources contradictoires, afin de se faire « sa propre opinion ». Ainsi Pour C., une retraitée de 66 ans, militante catholique de gauche, s'abonner à des hommes politiques qui ne sont pas de son bord politique est un moyen de ne pas « s'enfermer dans une seule pensée politique » :

« Alors dans les gens auxquels je suis abonnée, j'ai quand même aussi pas mal d'hommes politiques, de gauche ou de droite à qui je me suis abonnée [Vous les avez sélectionnés comment ?] Ben je les ai sélectionnés euh un peu au fur et à mesure de mes intérêts, de mes préférences politiques quand même aussi, euh voilà, vous voyez après les premiers sur qui je tombe y a Jean-Marc Ayrault [elle est au PS], euh je suis abonnée à comment elle s'appelle, Najat Belkacem, je suis abonnée à Cécile Duflot, voilà dans l'ordre où je les trouve [elle consulte en même temps son compte Twitter], après je sais qu'y a des gens de la droite, mais c'est un peu volontaire hein, de même que je veux pas m'enfermer dans un réseau catho je veux pas non plus m'enfermer dans une seule pensée politique. Je veux en tout cas savoir, bon j'ai mes choix bien sûr, mais je ne veux pas n'écouter que ceux qui pensent comme moi, voilà, c'est un choix ».

A., étudiant de 20 ans à Science Po. Paris, de gauche, fortement politisé mais sans engagement militant, déclare également consulter des comptes de femmes et d'hommes politiques qui ne sont pas de son bord politique (c'est-à-dire de droite et d'extrême-droite), tout en mettant clairement à distance les messages délivrés par ces personnalités. Cet enquêté n'accorde en effet *a priori* aucune crédibilité aux propos de ces personnalités auxquelles il est pourtant abonné. Le « jeu » consiste alors, pour cet individu très politisé – qui développe une posture distinctive de mise à l'écart d'une pratique « populiste et démagogique » de la politique, par opposition à une conception rationnelle voire technocratique – à s'informer des « bêtises » qui peuvent être dites par les hommes politiques sur Twitter avant tout pour s'en « amuser » voire s'en « moquer » :

[Et tu suis des hommes politiques ?] J'en suis quelque uns mais on va dire que la qualité de leurs tweets laisse à désirer, j'ai beaucoup suivi Nadine Morano, qui est vraiment

très très drôle, généralement ils sont plus drôles à droite qu'à gauche [Tu ne suis donc pas forcément ceux de ton bord politique ?] Non, ben, pff, on va dire que les trois quart du temps c'est... Le problème c'est que généralement y a beaucoup de conneries qui sont dites, à droite, aussi un peu à gauche, donc c'est plutôt les phrases rigolotes des hommes politiques qui m'intéressent [Donc tu me disais, Nadine Morano, y en d'autres ?] Un moment je me demande si y a pas Estrosi qui a sorti des trucs pas mal, bon après sinon oui, Copé, très drôle, Marine Le Pen est très intéressante à suivre aussi, je sais pas si c'est Marine Le Pen ou le Front national que je suis, parce que généralement tout de suite ça tombe dans le cliché, c'est 140 caractères donc faut faire choc, faut faire direct, oui Nadine Morano a fait quand même un record de tweets en décembre, en disant qu'elle regrettait les films sur la natalité, parce que les français ne se reproduisaient plus assez ou un truc comme ça, c'était très drôle.

Selon des modalités et des intérêts différents, ces enquêtés accordent à cette pratique sociale d'abonnement et de consultation de comptes d'hommes politiques une fonction essentiellement informative. D'après l'enquête de Frédérick Bastien et Gersende Blanchard⁷¹¹, cette fonction d'information arriverait également en tête des motivations qui pousseraient les internautes à consulter les profils des hommes politiques sur internet. D'après les résultats d'une étude par questionnaire administrée à un échantillon de 827 internautes engagés dans la campagne présidentielle française du printemps 2012 (dont essentiellement, à l'instar de notre population, des militants et des individus très politisés), 53 % des internautes déclarent consulter les profils des candidats⁷¹² afin de « compléter l'information qu'ils possèdent déjà », 49 % afin d'« être informés sans délais sur la campagne », ou encore 37 % afin d'« obtenir de l'information plus fiable que celle des médias traditionnels ».

Si l'on s'en tient donc aux significations données par les internautes, s'abonner et consulter le compte d'un homme politique sur Twitter ne relèverait pas tant d'une forme de délégation ou d'approbation politique que d'une nouvelle façon de s'informer, directement à la « source ». C'est donc parce qu'ils attribueraient une fonction plus informative – voire « journalistique » – que militante à cette pratique sociale que les enquêtés s'abonneraient presque indifféremment à des hommes politiques de gauche ou de droite.

⁷¹¹ Frédérick Bastien et Gersende Blanchard, « Les internautes face à la communication électorale à l'ère des campagnes postmodernes », in Philippe J. Maarek (dir.) *Présidentielle 2012, une communication politique bien singulière*, Paris, L'Harmattan, 2013, p. 145.

⁷¹² En réponse à la question suivante : « Les internautes consultent les sites web, les pages Facebook et les fils Twitter des candidats et des partis politiques pour diverses raisons. En ce qui vous concerne personnellement, quelles sont les raisons les plus importantes parmi les suivantes ? ».

Si elle n'invalide pas cette hypothèse, l'analyse statistique la nuance néanmoins quelque peu. En effet, lorsque l'on analyse les abonnements des membres de notre panel « répondants », il apparaît alors clairement que l'orientation politique déclarée des individus oriente leur pratique d'abonnement vers des hommes politiques de leur propre bord politique. On enregistre en effet une corrélation significative entre l'orientation politique des enquêtés et le nombre d'abonnements à des hommes politiques de gauche ($F = 7,900$; $r = 0,000$) et de droite ($F = 16,266$, $r = 000$).

Tableau 5.10 Nombre d'abonnement moyen à des comptes d'hommes politiques

Orientation enquêté	Nb. abonnements HP de Gauche	Nb. abonnements HP de Droite
Gauche ($n = 314$)	7*	2
Centre ($n = 53$)	4	3
Droite ($n = 127$)	3	6
Ne se positionne pas ($n = 62$)	1	1
Total ($n = 556$)	5	3

* Les individus « de gauche » sont abonnés en moyenne à 7 comptes d'hommes politiques de gauche

Les individus positionnés à gauche sont bien abonnés également à des hommes politiques de droite. Néanmoins, ils sont en moyenne abonnés à quatre fois plus de comptes d'hommes politiques de leur propre bord politique (7 abonnements en moyenne à des hommes politiques de gauche) que du bord opposé (2 abonnements à des hommes politiques de droite). Et contrairement à ce que l'on a observé avec les abonnements à des médias politiques, la réciproque est également vraie à droite. Les individus de droite sont en moyenne abonnés à deux fois plus d'hommes politiques de droite (6 comptes en moyenne) que de gauche (3 comptes en moyenne).

S'ils invalident donc l'hypothèse d'une consommation politiquement désintéressée et uniquement informative des comptes d'hommes politiques, ces résultats n'invalident pas – ni d'ailleurs ne valident – l'hypothèse selon laquelle être abonné à un homme politique trahirait moins un rapport de délégation qu'une volonté d'information (volonté d'information dont on sait par ailleurs qu'elle est socialement et politiquement sélective).

L'analyse statistique des abonnements des producteurs de tweets politiques a donc mis en lumière que la consommation d'informations et d'opinions politiques sur Twitter reposait sur des logiques sociales et politiques très structurantes et relativement similaires à celles qu'avait observées Pierre Bourdieu dans *La Distinction*. Ces résultats confirment une nouvelle fois notre hypothèse selon laquelle les pratiques sociales en ligne sont régies par le même système de dispositions durables – l'habitus – que les pratiques « hors-ligne ».

Cependant, le fait de s'abonner à un compte Twitter, action gratuite et immédiate, ne saurait suffire à présumer d'une consommation *effective* des contenus publiés par ce compte. Étant abonnés à des dizaines de médias en moyenne, les producteurs de tweets politiques s'exposent alors à recevoir des centaines voire des milliers de tweets par heure dans leur *time-line*. Rien d'indique qu'un abonné ne lise alors, même occasionnellement, les tweets des médias auxquels il est abonné.

Deux autres données peuvent alors être mobilisées afin de saisir de manière plus fine la consommation cette fois *effective* des producteurs de tweets politiques : les retweets et les liens hypertextes.

1.2 La consommation effective de biens politiques

Contrairement au fait de s'abonner à un compte médiatique, retweeter un message ou publier un lien vers un article de presse sont des pratiques sociales qui rendent compte d'une consommation *effective* – et non potentielle – de biens informationnels. Si les significations que les inscrits accordent à ces deux pratiques sont diverses, on peut en revanche affirmer qu'elles supposent *a minima* que leurs auteurs aient lus – ou, dans le cas des liens vers des articles de presse, très brièvement parcourus – les informations relayées sur leurs comptes.

Nous avons analysé ces pratiques de consommation *effectives* en tentant de répondre à la question suivante : ces pratiques sont-elles régies par les mêmes logiques sociologiques que la pratique d'abonnement aux comptes médiatiques et politiques (1.2.1) ou connaissent-elles au contraire certaines spécificités (1.2.2) ?

1.2.1 Des logiques de consommation très similaires à la pratique d'abonnement

Afin d'analyser la fréquence de ces pratiques de consommation *effective* et le taux de producteurs qui y ont recours, nous avons comptabilisé, à partir d'une base exhaustive de tweets⁷¹³ publiés par le panel « répondants » (sur une plage de temps donnée, du 1^{er} mars 2013 au 31 janvier 2014⁷¹⁴), le nombre de (re)tweets ayant comme auteur original un compte de média, de journaliste ou d'homme politique, ainsi que le nombre de tweets contenant un lien hypertexte vers un article publié sur un site d'information politique. Pour chaque catégorie (média, journaliste et homme politique), nous n'avons retenu que les comptes auxquels au minimum 10 % des membres du panel étaient abonnés⁷¹⁵.

On peut ainsi d'abord constater qu'au sein de notre population de producteurs de tweets politiques, retweeter un message publié par un média, un journaliste ou un homme politique, ou relayer un article de presse, sont des pratiques sociales moins répandues que le fait de s'abonner aux comptes de ces différents agents.

Tableau 5.11 Taux et nombre de retweets moyens provenant de médias, de journalistes et d'hommes politiques

Compte à l'origine du RT	Taux Panel (n=558)	Nombre moyen de tweets	Taux d'abonnement
Médias politiques (RT)	61%*	22**	89%
Médias politiques (lien http)	63%	35	
Hommes politiques	31%	7	60%
Journalistes	22%	4	49%

* 61 % des membres du panel « répondants » ont déjà retweeté sur la période un ou plusieurs tweets publiés par des comptes médiatiques.

** Toujours sur la période, ces 61 % de « répondants » ont publié en moyenne 22 retweets de messages publiés par des comptes médiatiques.

⁷¹³ Les retweets et articles de presse analysés dans cette section ne sont donc pas nécessairement « politiques », ils ne contiennent pas forcément des occurrences de type *politics*, cf. chapitre 4.

⁷¹⁴ Soit la même plage de temps que celle étudiée dans le chapitre 4 afin d'analyser le volume et la fréquence de production des panels. Une période, rappelons-le, marquée par aucun événement politique majeur (tel que des élections nationales).

⁷¹⁵ Ce seuil des 10 % a été instauré pour des raisons exclusivement techniques. Il aurait été particulièrement fastidieux de comptabiliser « manuellement » l'ensemble des retweets ayant pour auteur original l'un des 1600 comptes auxquels nos panels étaient abonnés. L'instauration de cette limite rend donc non exhaustive notre analyse. On peut néanmoins supposer que les individus retweetant exclusivement des tweets publiés par des comptes auxquels moins de 10 % du panel est abonné sont relativement rares. Au final, nous avons donc analysé les retweets provenant d'une liste de 25 comptes de médias, 21 comptes d'hommes politiques et 7 comptes de journalistes politiques.

Ainsi, si près de 90 % des membres du panel « répondants » sont abonnés à un compte de média, seuls environ 60 % ont déjà retweeté, sur la période étudiée, un tweet publié par l'un de ces comptes. Ce même résultat peut être observé concernant les hommes politiques. Seul un « répondant » sur trois (31 %) a déjà retweeté un homme politique sur la période, alors qu'ils sont près du double (60 %) à être abonnés à leurs comptes Twitter.

Les individus qui publient des retweets⁷¹⁶ le font, de plus, relativement peu souvent. Les enquêtés qui retweetent des comptes médiatiques (61 % du panel) ne le font ainsi que deux fois par mois en moyenne. Du 1^{er} mars 2013 au 31 janvier 2014, cette catégorie de « répondants » n'a publié en effet que 22 retweets sur la période.

On peut noter à ce propos que les producteurs de tweets politiques semblent privilégier la publication de liens hypertexte (trois tweets par individu et par mois en moyenne) à la pratique du retweet (deux tweets par individu et par mois) lorsqu'ils entreprennent de « diffuser » les informations qu'ils souhaitent « faire partager ». Ce résultat n'est pas anodin et mérite qu'on explicite davantage ici les différences existantes entre ces deux pratiques d'apparence très similaires. Retweeter un tweet publié par un compte de média implique de reproduire à l'identique sur son propre compte Twitter le tweet publié par ce média. Ces tweets se limitent, dans leur grande majorité, à mentionner le titre ou le sujet de l'article ainsi que le lien hypertexte renvoyant vers le site où est hébergé l'article. Les individus qui choisissent de retweeter ces messages n'ont alors aucune capacité expressive supplémentaire⁷¹⁷ : ils ne peuvent ainsi ni commenter, ni contextualiser l'information relayée. En revanche, lorsqu'ils choisissent de publier eux-mêmes un lien hypertexte vers un article de presse, les individus inscrits sur Twitter peuvent commenter l'information qu'ils souhaitent « diffuser »⁷¹⁸. Il s'agit là d'une première différence entre ces deux pratiques. Tandis que la pratique du retweet s'apparente à une consommation « brute » d'un bien informationnel, la publication d'un lien hypertexte rend souvent compte d'une consommation contextualisée en « première personne » de l'information relayée. Au-delà de ces capacités expressives différenciées, une seconde différence entre ces deux pratiques concerne la nature du support privilégié par les individus pour s'informer. Retweeter un message implique d'avoir lu ce

⁷¹⁶ Dorénavant, pour faciliter la lecture, nous préférons la formule « publier des retweets » à celle plus complète de « publier des retweets dont l'auteur original est un homme politique un journaliste ou un média ».

⁷¹⁷ Cette affirmation mérite toutefois d'être nuancée, en ce sens que les individus ont la possibilité de commenter un message retweeté. Seulement, contraint par la taille des 140 caractères, le commentaire associé au message original ne peut être que très bref et se limite le plus souvent qu'à un seul mot ou hashtag.

⁷¹⁸ Les liens hypertextes, qui comprennent souvent de très nombreux caractères, trouvent une forme « compressée » et « raccourcie » sur Twitter, ce qui permet de ne pas dépasser trop facilement la limite des 140 caractères.

tweet directement sur Twitter, dans sa *timeline* ou sur le compte du média qui l'a publié. En revanche, diffuser un article de presse par l'intermédiaire d'un lien hypertexte implique potentiellement que l'auteur du tweet ait pris connaissance de l'article directement sur le site internet et non sur Twitter (il suffit simplement pour l'individu de copier/coller l'adresse URL qui s'affiche sur son navigateur). Le plus grand nombre de tweets avec lien hypertexte (comparativement au nombre de retweets de comptes médiatiques) peut donc sans doute être appréhendé comme le signe d'une plus forte propension de nos enquêtés à consommer les biens médiatiques sur d'autres supports que Twitter.

On peut également remarquer que la pratique du retweet est encore plus rare concernant les messages publiés par des hommes politiques. Sur la période étudiée, le nombre de retweets moyen par individu est ainsi de moins d'un retweet par mois (0,6 retweet). Relayer en première personne des messages d'hommes politiques sur Twitter est donc une pratique relativement rare, y compris parmi les producteurs de tweets politiques, une population pourtant très fortement politisée. Cette pratique varie cependant significativement selon l'intérêt déclaré pour la politique ($F = 8,807$, $r = 0,000$) ou encore le fait d'adhérer ou non à un parti ($F = 7,661$, $r = 0,006$). Les individus qui déclarent « beaucoup » s'intéresser à la politique ont ainsi publié 8 retweets d'hommes politiques sur la période, contre seulement 3 pour ceux qui s'y intéressent « assez », « un peu » ou « pas du tout ». Les militants de parti politique ont publié quant à eux un nombre de retweets d'hommes politiques supérieur de 25 % à celui des non-militants (8 tweets contre 6).

De manière analogue à ce que l'on a observé concernant les pratiques d'abonnement, on constate que les hommes politiques dont les enquêtés choisissent de relayer les messages appartiennent le plus souvent à leur bord politique. On enregistre en effet une corrélation significative entre l'orientation politique des enquêtés et leur nombre de retweets d'hommes politiques de gauche ($F = 2,943$, $r = 0,054$) et de droite ($F = 8,212$, $r = 0,000$). Les individus positionnés à droite ont ainsi retweeté en moyenne trois tweets publiés par des hommes politiques de droite et un tweet publié par des hommes politiques de gauche. Et inversement, les individus positionnés à gauche ont publié en moyenne deux retweets d'hommes politiques de gauche et un retweet d'homme politique de droite. Retweeter des messages d'hommes politiques est donc une pratique politiquement située.

Ces mêmes logiques politiques orientent également le choix des comptes cette fois *médiatiques* que les individus sur Twitter choisissent de retweeter. On observe là encore une

corrélation significative entre l'orientation politique des enquêtés et le nombre de retweets de médias de droite ($F = 4,043$, $r = 018$). Les individus orientés à droite ont ainsi publié en moyenne trois retweets de médias de droite (du type le *Figaro*, le *Point*, etc.) contre un retweet de média de gauche (*Libération*, *Médiapart*, etc.).

Ces pratiques de consommation *effectives* de biens médiatiques sur Twitter obéissent également à des logiques sociologiques. La position sociale des membres de notre panel détermine en effet non seulement le volume *total* d'articles relayés (*via* la publication de liens hypertextes) mais également le nombre d'articles provenant de médias de gauche et de droite.

Tableau 5.12 Nombre de tweets publiés contenant des liens vers des sites d'information en ligne par PCS

PCS	Nombre de liens vers des médias	Nb de liens vers des médias de gauche	Nb de liens vers des médias de droite
Cadres du privé ($n = 87$) ⁷¹⁹	16*	3	5
Cadres du public ($n = 68$)	51	9	5
Prof. interm. privé ($n = 39$)	14	4	4
Prof. interm. public ($n = 42$)	34	16	5
Employés/ouvriers ($n = 42$)	14	3	7
Retraités et inactifs ($n = 32$)	78	22	21
Étudiants ($n = 199$)	11	3	2
Lycéens ($n = 45$)	3	1	1
Total général ($n = 554$)	22	6	5
F =	2,659	3,562	3,755
r =	0,007	0,000	0,000

* Les cadres du privé ont publié sur la période en moyenne 16 tweets contenant un lien vers le site d'un média

Comme on avait commencé à l'entrevoir avec les abonnements aux comptes médiatiques, on constate une très nette différence entre les salariés – cadres ou professions intermédiaires – du privé et du public. Les cadres du secteur public ont en moyenne publié trois fois plus de liens vers des sites médiatiques (51 tweets en moyenne sur la période, soit 5,1 tweets par mois) que les cadres des entreprises (16 tweets en moyenne, soit 1,6 tweet par mois). Les cadres du public constituent donc, avec les retraités, les plus importants

⁷¹⁹ Encore une fois, malgré des effectifs faibles, nous avons fait le choix de ne pas réunir les différentes sous-catégories de cadres et de professions intermédiaires en raison des différences tout à fait significatives et structurantes entre, pour le dire vite, le public et le privé.

consommateurs *effectifs* de biens informationnels. Ces individus présentent un profil proche des *consommateurs ostentatoires* d'informations politiques décrits dans notre troisième chapitre⁷²⁰ : relayer des articles de presse constitue pour eux une pratique sociale distinctive qu'ils légitiment en mobilisant des arguments normatifs (ils revendiquent ainsi une volonté « civique » de « faire partager » des informations jugées « importantes »).

On remarque cependant que les cadres du public, contrairement aux retraités, s'orientent surtout vers des biens informationnels produits par des médias de gauche (16 tweets en moyenne vers des médias de gauche, contre seulement 5 vers des médias de droite). Les retraités au contraire publient autant de liens vers des médias de gauche (22 tweets) que de droite (21 tweets). Si les effectifs trop faibles de cette dernière catégorie nous contraignent à ne pas dissocier les retraités issus du secteur privé et ceux du secteur public, il y a fort à parier que les retraités du secteur public s'orientent eux aussi majoritairement vers des médias de gauche.

On peut enfin observer que les étudiants (11 tweets en moyenne), et plus encore les lycéens (3 tweets), publient très peu de liens vers des articles de presse. Les lycéens semblent ainsi mobiliser Twitter selon un usage très proche de celui du modèle « post-it » (décrit lui aussi dans notre chapitre 3), à savoir comme un outil permettant de se livrer à « *une narration de soi continue, contextuelle et renouvelée* »⁷²¹, souvent à travers des discussions quasi-synchrones avec leurs abonnés. La publication d'articles de presse ne s'inscrit alors pas dans cette économie de pratiques « adolescentes ».

En résumé, les pratiques de consommation *effectives* de biens politiques sur Twitter, bien que relativement rares, semblent répondre aux mêmes logiques politiques et sociologiques que la pratique de consommation *potentielle* (que l'on peut qualifier également de « passive ») qu'est l'abonnement.

Cependant, nous allons voir maintenant, en examinant plus précisément les comptes les plus retweetés et la distribution dans le temps de leurs retweets que ces pratiques obéissent également à des logiques qui leur sont spécifiques.

⁷²⁰ Cf. chapitre 3.

⁷²¹ Dominique Cardon, « Le design de la visibilité », art. cit., p. 118.

1.2.2 Les logiques spécifiques des pratiques sociales de consommation effectives de biens politiques

Une autre manière d'analyser les spécificités de ces pratiques de consommation *effectives* consiste en effet à comparer précisément l'ordre des comptes les plus retweetés avec celui des comptes bénéficiant du plus grand nombre d'abonnés au sein de notre panel.

Tableau 5.13 Taux (en %) et nombre de retweets moyens provenant de comptes de médias, de journalistes et d'hommes politiques

	Rang Retweet	Nom du compte	Taux d'abonnement	Rang Abonnement	Taux de Retweet	Nb de RT moyen par individu
RT MEDIAS	1	lemondefr	39	1	29	13
	2	afpfr	23	5	23	5
	3	itele	11	27	20	7
	4	BFMTV	17	11	18	5
	5	liberation_info	27	3	16	4
RT JOURNALISTES	1	edwyplenel	21	1	13	3
	2	jmaphatie	21	2	8	3
	3	pierrehasaki	13	4	6	1
	4	quatremr	11	5	4	3
	5	nicodomenach	17	3	3	2
RT HOMMES POL.	1	nk_m	20	3	9	6
	2	JLMelenchon	19	8	8	4
	3	CecileDuflot	19	4	6	2
	4	najatvb	17	10	5	3
	5	jeanmarcayrault	19	6	4	2

On observe alors de nombreuses différences entre les comptes qui comptent le plus d'abonnés et ceux qui sont le plus retweetés. Concernant les médias d'information, l'ordre des comptes les plus « consommés » n'est ainsi pas le même selon la pratique sociale étudiée (abonnements, retweets ou publication de liens hypertextes).

Néanmoins, avant d'examiner ces différences, une exception de taille doit être ici soulignée : il s'agit du rôle prépondérant joué par le compte du quotidien *Le Monde*. Quelle que soit la pratique étudiée, le compte du quotidien est systématiquement le plus

« consommé ». *Le Monde* est le compte auxquels les membres de notre panel sont le plus abonnés (39 %), celui également dont les tweets sont les plus retweetés (29 % des « répondants » ont déjà retweeté l'un des tweets du compte sur la période) et les articles de son site internet les plus diffusés (39 % ont déjà publié un tweet contenant un lien vers son site internet). Sur Twitter également, *Le Monde* semble donc conserver sa position dominante dans le champ de la presse écrite et plus largement dans le champ médiatique. Cette position « hégémonique » s'explique sans doute également par le caractère socialement et politiquement peu clivant du titre sur Twitter⁷²². Il s'agit en effet de l'un des seuls comptes médiatiques qui bénéficie d'un taux de retweets relativement similaire entre les cadres du privé (41 % ont déjà publié un lien vers le site du *Monde*) et ceux du public (46 %) ou encore entre les individus de gauche (37 %) et ceux de droite (36 %). Il n'y a ainsi que parmi les employés et ouvriers que les articles du titre ne sont pas les plus diffusés (29 % ont déjà publié un lien vers *Le Monde* alors que ce taux est de 36 % pour *Le Parisien*).

Hormis cette exception notable, l'ordre des médias les plus « consommés » diffère donc selon la pratique étudiée. On peut ainsi observer que les comptes des chaînes de télévision continue (*BFMTV* et *Itélé*) sont parmi les comptes de médias les plus retweetés (avec respectivement 18 % et 20 % des « répondants » qui ont retweeté au moins une fois l'un de ces comptes durant la période étudiée) et ce alors que leurs nombres d'abonnés au sein de notre panel sont proportionnellement bien moins importants que beaucoup d'autres comptes (avec respectivement 17 % et 11 % d'abonnés). Ce résultat peut sembler alors doublement surprenant.

D'une part car ces deux titres ne bénéficient pas, nous l'avons vu, d'une importante légitimité culturelle. Il peut alors sembler curieux que les « répondants », qui disposent d'un important capital culturel, choisissent de consommer *ostensiblement* des informations produites par ces médias peu légitimes. Étudier la composition sociologique de la sous-population qui a retweeté ces comptes télévisés permet de mieux comprendre ce paradoxe. On remarque en effet que les catégories supérieures ont relativement peu retweeté ces comptes comparativement aux comptes culturellement plus légitimes : les cadres du public ne sont ainsi que 12 % à avoir retweeté *BFMTV* (6 points en-dessous de la moyenne du panel) alors

⁷²² Un caractère socialement peu clivant *uniquement* au sein de cette population très fortement politisée, dont toutes les catégories sociales, mêmes les plus dépourvues en capitaux (économiques ou culturels) sont anormalement (au sens statistique) politisées par rapport à la moyenne des Français. *Le Monde*, en dehors de notre population très spécifique, reste donc un titre très discriminant socialement, privilégié par (et destiné) aux classes sociales supérieures et aux individus les plus diplômés.

qu'ils sont 31 % à avoir retweeté *Le Monde* ou encore 18 % à avoir retweeté le compte de *Libération*. À l'inverse, les employés et ouvriers (24 %), les retraités (22 %), les étudiants (23 %) et les lycéens (20 %) ont significativement plus retweeté ces comptes que la moyenne du panel.

Reste cependant une seconde contradiction : pourquoi ces titres connaissent-ils un nombre de consommateurs *effectifs* plus important que leur nombre d'abonnés⁷²³, ou, dans le cas de *BFMTV*, des taux similaires alors que beaucoup d'autres comptes ont un taux de consommateurs *effectifs* bien moins important que d'abonnés (le compte de *Rue89* comptabilise ainsi par exemple 35 % d'abonnés alors que seuls 13 % des membres du panel ont déjà retweeté l'un des messages publiés par le compte) ?

Sans pouvoir apporter d'éléments de réponse empiriquement étayés à cette question, on peut néanmoins supposer l'existence d'un décalage entre *les perceptions* qu'ont les agents sociaux des titres auxquels ils s'abonnent et *leurs dispositions réelles* à les consommer. Les individus de notre panel ont été ainsi par exemple nombreux à s'abonner au compte de *Rue89*, croyant être sans doute « attirés » par la ligne éditoriale du titre (un site exclusivement en ligne, avec une ligne de gauche, destiné à un public relativement jeune et diplômé). Pourtant, dans les faits, ces mêmes individus ne consomment pas ou peu les articles produits par ce titre et sont, semble-t-il, davantage enclins à consommer des biens informationnels moins légitimes.

Enfin, précisons que ces médias télévisés relaient souvent sur leur compte Twitter des émissions politiques qu'ils diffusent en direct sur leur antenne. Or le commentaire des émissions télévisées génère habituellement un niveau plus élevé de réactions sur le réseau social, ce qui pourrait donc également contribuer à expliquer le fait que les tweets publiés par ces comptes soient plus souvent retweetés. Derrière ce constat se pose alors une question : peut-on considérer un retweet d'un média politique comme une opinion ou un marqueur d'opinion (de la part de l'individu qui relaie le message) ?

Pour répondre à cette question, nous allons maintenant proposer une typologie des différents types de tweets politiques que l'on peut observer sur Twitter, et nous demander dans quelle mesure ces « biens », ces consommations *indirectes* formalisées, peuvent être considérées comme des opinions ou des marqueurs d'opinion.

⁷²³ Le taux d'abonnés du compte d'*Itélé* (11 %) est en effet presque deux fois inférieur au taux de répondants l'ayant déjà retweeté (20 %).

2. Formes de consommation *indirecte* et marqueurs d'opinion

Retweeter un message publié par un compte de média ou un homme politique ne constitue en effet qu'une forme de consommation parmi d'autres d'informations ou d'opinions politiques, même s'il s'agit de la forme de consommation la plus manifeste. Au-delà de ces biens *ostensiblement* consommés, il existe une grande variété de formes et de modalités d'expressions politiques sur Twitter qui oscillent entre l'information et l'opinion.

Nous allons maintenant précisément analyser ces autres formes de consommation, plus *indirectes*, en proposant l'esquisse d'une *typologie*, ou plutôt d'une grille de lecture multicritères *des tweets politiques* (2.1). Nous nous demanderons alors si ces différents types de biens politiques constituent ou non des opinions ou des marqueurs d'opinions. Si nous avons jusqu'ici abordé de manière relativement indifférenciée les tweets d'information et les tweets d'opinion, notre typologie nous invitera à clairement distinguer ces deux types de biens. Nous avancerons alors, conformément au modèle théorique que nous avons esquissé plus tôt⁷²⁴, que sur Twitter, ne s'échangent pas seulement des opinions, mais également des *biens nécessaires à la production d'opinions*. Dès lors, connaître d'une part les propriétés sociologiques et politiques des agents, et objectiver d'autre part les biens informationnels qu'ils choisissent de consommer peut permettre au chercheur d'identifier les opinions des agents sur un sujet donné, sans que ces opinions ne soient nécessairement formalisées (2.2).

2.1 Une grille de lecture multicritères des tweets politiques

Plutôt qu'une typologie, c'est une grille de lecture construite à partir de différents critères d'analyse que nous proposons ici de mobiliser afin de rendre compte de la diversité – tant formelle que fonctionnelle – des tweets politiques publiés par les membres de nos panels, à savoir des productions écrites d'au maximum 140 caractères contenant certaines occurrences politiques⁷²⁵.

Les tweets politiques peuvent être ainsi analysés selon douze principales variables, non exclusives les unes des autres : le format du tweet (tweet ou retweet, présence ou non de

⁷²⁴ Cf. chapitre 2.

⁷²⁵ Cf. chapitre 3.

lien hypertexte), le destinataire du tweet, l'intégration du tweet au sein d'une conversation, le contexte d'émission du tweet, le niveau de langage, le niveau d'orthographe, le registre et le type de discours, le niveau de technicité des termes politiques mobilisés, la nature politique de l'objet sur lequel porte le message, le caractère manifeste ou latent de l'opinion véhiculée par le message, la tonalité de cette opinion ainsi que son mode de production. Chaque tweet politique peut être ainsi classé selon chacun de ces critères⁷²⁶.

Ces différentes variables peuvent être regroupées autour de quatre grands principes classificatoires : le contexte d'émission du tweet, ses caractéristiques formelles, la nature politique de l'objet sur lequel porte le message et le format du tweet. C'est ce dernier principe classificatoire que nous retiendrons afin de répondre à la problématique qui nous intéresse directement ici : dans quelle mesure les tweets politiques peuvent-ils être appréhendés comme des opinions ou des biens nécessaires à la production d'opinions ? Auparavant, nous présenterons brièvement les trois autres principes classificatoires que nous venons d'énumérer, ce afin de rendre compte de la diversité du matériau étudié.

2.1.1 Les principes classificatoires des tweets politiques non retenus pour l'analyse

Précisons en préambule que les résultats exposés ici ne seront abordés que très superficiellement. L'enjeu est simplement de démontrer qu'il n'existe pas un principe classificatoire *unique* des tweets politiques, et que la diversité de ce matériau d'étude ne pourrait être réduite à une seule typologie.

⁷²⁶ Nous avons appliqué cette grille d'analyse à un échantillon de 1 108 tweets politiques, provenant de 435 utilisateurs uniques, soit 78 % du panel. Ce corpus a été constitué selon une double logique : dans un premier temps, nous n'avons retenu qu'un tweet politique (tiré au sort) par individu, afin que soient représentés dans le corpus tous les individus qui ont publié au moins un tweet politique durant la période étudiée (22 février au 22 avril 2013). Nous avons ensuite sélectionné aléatoirement un échantillon de tweets politiques ($n = 673$) à partir de la production totale du panel. Cette sélection aléatoire a permis de rendre compte des fortes inégalités de publication entre producteurs. Avant de dévoiler plus en détail nos résultats, il convient de préciser que le corpus étudié a été constitué durant une séquence temporelle relativement réduite – deux mois – mais qui a été marquée par deux « événements politiques » notables : l'adoption du projet de loi relatif au mariage homosexuel et l'affaire Cahuzac. Précisons enfin que certains des résultats exposés ici ne seront abordés que très superficiellement. L'enjeu n'est pas tant de proposer une analyse la plus détaillée possible mais simplement de démontrer que la diversité de ce matériau d'étude – les tweets politiques – ne serait se résumer à une typologie unifactorielle.

2.1.1.1 Classer les tweets politiques en fonction de leur contexte d'émission

Un premier principe de classification possible est de s'intéresser aux différents contextes dans lesquels sont émis ces messages. Trois variables peuvent être alors mobilisées afin de classer les tweets politiques en fonction de leur contexte d'émission : le destinataire du tweet, l'intégration du tweet dans une conversation et le type de *stimulus* ayant provoqué la production du tweet.

Une première variable à prendre en considération peut être en effet *le destinataire du tweet* politique publié. Contrairement aux opinions enregistrées par sondage, les tweets politiques, comme n'importe quelles productions verbales – écrites ou orales – peuvent être adressés à une tierce personne. C'est ainsi le cas d'un cinquième des tweets politiques étudiés (22 %).

Tableau 5.14 Destinataires des tweets politiques publiés (n = 1 108 tweets)

Catégorie	Définition	%	Exemple
Aucun destinataire apparent	Le tweet n'est pas directement adressé à un tiers.	78	« <i>L'humour de #Hollande envers #Sarkozy laisse vraiment à désirer... #Hollande va trop loin ! #mauvaise blague</i> » (homme, 23 ans, étudiant, de droite)
Public de lecteurs, d'abonnés	L'auteur du tweet prend à parti ses lecteurs, sans directement les nommer	3	« <i>Raffarin fait un point crise/chômage.. On lui explique que tous les ministres ne sont pas mobilisés sur le #mariagepourtous ? #directSenat</i> » (homme, 25 ans, étudiant, de gauche)
Médias, journalistes	Le tweet est adressé à un média ou à un journaliste.	4	« <i>merci @1213dimanche d'avoir donné la parole a @bayrou</i> » (femme, 56 ans, ancienne cadre financier, militante Modem)
Partis et hommes politiques	Le tweet est adressé à un parti politique ou un homme politique.	5	« <i>@cestrosi je suis socialiste mais je trouve que votre réaction est plutôt responsable vu le contexte politique actuel</i> » (homme, 40 ans, affréteur, de gauche)
Individus	Le tweet est adressé à une connaissance personnelle.	10	« <i>@lionbiensage Et moi, dire que j'ai cru au PQ pour défendre la gauche. On apprend de nos erreurs, j'imagine.</i> (homme, 30 ans, technicien en informatique, de gauche)

Cette tierce personne peut être un autre utilisateur de Twitter : un membre issu de son réseau d'interconnaissances (10 % des tweets étudiés), un journaliste ou un média (4 %), ou encore un parti ou un homme politique (5 % des tweets).

L'auteur du tweet peut ainsi prendre à partie un autre utilisateur dans le but de recevoir une réponse de sa part, il s'agit alors d'un appel à « entamer » une conversation. Mais l'auteur du message peut également apostropher une personnalité, notamment politique, sans attendre réellement de retour de sa part. L'auteur cherchera simplement à faire savoir à son réseau d'abonnés ce qu'il pense de cette personnalité, le « message » qu'il tient à lui adresser. Analyser la relation qui unit l'auteur du tweet au tiers qu'il prend à partie permet de mieux comprendre le *verbatim* étudié.

Si le destinataire, ou même un autre utilisateur, décide de répondre à l'auteur du tweet, se crée alors *une conversation par tweets interposés*, entre deux ou plusieurs individus. 49 % des tweets politiques étudiés sur la période donnent lieu ou s'insèrent dans des conversations, parfois quasi-synchrones, entre deux ou plusieurs utilisateurs – trois en moyenne. Dans un cas sur deux, la production de tweets politiques est donc une pratique collective et non individuelle⁷²⁷. En moyenne, les conversations sur Twitter se composent de 8 tweets échangés.

34 % des tweets politiques publiés provoquent un échange entre plusieurs utilisateurs et 15 % sont des réponses aux tweets d'autres utilisateurs. Les individus qui prennent part à la discussion, leur position sociale, leur orientation politique, etc. sont alors autant de variables à prendre en compte afin de mieux saisir le contexte *situationnel* dans lequel est publié le tweet politique étudié.

Au-delà du contexte situationnel *numérique* dans lequel est émis le tweet politique, on peut également classer ces messages en fonction du *stimulus* qui a provoqué leur publication. On peut ainsi constater que 67 % des tweets politiques publiés sur la séquence sont des réactions à un sujet inscrit à l'agenda médiatique, 15 % sont des réactions à des propos de personnalités – notamment politiques – et 7 % des commentaires d'émissions télévisées. Seuls 11 % des tweets étudiés n'abordent pas directement un sujet politique d'actualité.

⁷²⁷ Précisons que ce résultat a été pour nous surprenant : cette réalité empirique est en effet difficilement saisissable *via* notre logiciel. Les tweets indexés apparaissent en effet comme des productions unitaires et isolées, la configuration de notre logiciel ne permet pas de restituer graphiquement les tweets collectés dans les conversations desquelles ils sont issus. Nous avons nous-mêmes mis du temps à prendre conscience de cette réalité empirique. La configuration graphique du logiciel correspondait en fait à nos propres représentations mentales, représentations concevant les opinions comme des productions isolées et individuelles.

Tableau 5.15 Contexte d'émission du tweet (n= 1 108tweets)

Catégorie	Définition	%	Exemple
Sujet inscrit à l'agenda médiatique	Le tweet aborde un sujet inscrit à l'agenda médiatique	67	« RT@itele La côté de confiance de F.#Hollande perd 5 pts en mars à 20% de "satisfaits", vs. 74% de mécontents (+9 pts) (Yougov / @itele / @LeHuffPost) » (Homme, 64 ans, retraité militaire, de gauche)
Réactions à des déclarations dans (ou rapportées par) des médias	Commentaires et réactions à des déclarations d'hommes politiques ou de personnalités publiques	15	« Drogue:Valls annonce une amende payable sur place pour les consommateurs http://dlvr.it/3B0Wyr il ne veut tjrs pas prendre sa carte UMP? <3 » (femme, 20 ans, étudiante, de droite)
Prestations télévisées	Tweets publiés durant ou à la suite d'émissions télévisées	7	« Ca fait plusieurs que je vois #Lefol en debat et c un dur à cuire !!! Etonnant que le PS ne l'ait pas plus exploité par le passé » (homme, 35 ans, professeur du secondaire, militant MoDem)
Hors-actualité immédiate	Tweet dont le sujet n'est pas inscrit à l'agenda médiatique	11	« @metallique53 pour côtoyer des Jeunes communistes presque tout les jours je peux te dire que le pcf n'est pas stalinien... » (homme, 20 ans, étudiant en sociologie, militant PG)

2.1.1.2 Classer les tweets politiques en fonction de leurs qualités formelles

Au-delà de leur contexte d'émission, les tweets politiques peuvent être également classés en fonction de variables relatives à leur *qualité formelle*. Contrairement aux réponses de questionnaire ou aux *verbatim* enregistrés en entretien, les tweets politiques sont en effet des productions écrites, rédigées directement par leurs auteurs. Évaluer les tweets politiques au prisme de leur qualité formelle peut alors apporter au chercheur certaines indications quant à la position sociale de leur auteur ou certaines de ses intentions expressives.

Le niveau d'orthographe et le niveau de langage peuvent ainsi constituer deux indices permettant de mesurer – ou plutôt d' « intuitionner » – le capital scolaire et l'origine sociale des auteurs des tweets étudiés⁷²⁸. Dans le cas des tweets politiques, et plus particulièrement dans le cas de ceux publiés par les individus qui ont répondu à notre questionnaire, ces deux variables sont en réalité très peu discriminantes : 97 % des tweets étudiés ne comprennent aucune faute d'orthographe et 86 % sont rédigés dans un style courant ou soutenu. Ces

⁷²⁸ Cf. chapitre 2.

résultats confirment donc encore une fois le niveau d'étude et le milieu social relativement élevés de notre population⁷²⁹.

Le type et le registre de discours mobilisés peuvent quant à eux nous renseigner sur l'*intention* que l'auteur a souhaité donner au message qu'il a publié sur son compte et délivré à ses abonnés. Nous avons opéré notre propre classification – qui amalgame types et registres de discours – à partir d'observations empiriques. Cette variable se décline selon six modalités :

- *Des tweets politiques d'injonction* : l'auteur du tweet enjoint, par exemple, un homme politique à démissionner. Ces tweets représentent seulement 2% de la production totale de tweets politiques et sont le plus souvent publiés par des militants politiques (pour un exemple, cf. tableau 5.17 ci-dessous).

- *Des tweets polémiques ou critiques* : l'auteur du tweet entend dénoncer un « adversaire » ou un projet politique qu'il désapprouve. Ces messages représentent 29 % des tweets politiques étudiés.

- *Des tweets de soutien* : l'auteur entend au contraire soutenir un projet ou un homme politique. Ces messages ne représentent que 5 % des tweets politiques et sont quasi-exclusivement publiés par des auteurs militants.

- *Des tweets informatifs* : l'auteur du tweet délivre ou relaie une information sur un ton neutre. Ces messages représentent un quart de la production totale de tweets politiques étudiés.

- *Des tweets argumentatifs* : l'auteur du tweet défend une idée à l'appui d'un « argument ». La taille réduite (140 caractères) des tweets politiques rend bien sûr quelque peu abusif l'emploi du terme « argumentatif ». Ces messages représentent 14 % des tweets étudiés.

- *Des tweets ironiques, humoristiques* : un tweet politique sur quatre (25 %) est ironique ou humoristique. L'humour et l'ironie sont en effet deux registres couramment employés par les individus sur Twitter – et plus généralement sur les réseaux sociaux – pour commenter des sujets politiques ou donner des avis et des opinions politiques. Il est intéressant de remarquer que toutes les catégories de producteurs, des moins aux plus politisées, mobilisent ce registre. Parler ironiquement ou humoristiquement de la politique ne témoigne pas nécessairement d'une forme de

⁷²⁹ Précisons néanmoins que les chances, au sens statistique, de commettre des fautes d'orthographe sont réduites avec des messages de 140 caractères.

mise à distance voire de défiance vis-à-vis de la politique. Le maniement d'un discours ironique sur la politique peut au contraire parfois requérir un haut niveau de compétence politique⁷³⁰.

Tableau 5.16 Registre de discours des tweets politiques (n= 1 108tweets)

Catégorie	%	Exemple
Injonction	2	« Hollande, balaie devant ta porte ! - Il faut balayer devant sa porte avant de donner des leçons aux autres http://hsene.over-blog.com/article-hollande-balaie-devant-ta-porte-115834112.html ... » (homme, 65 ans, cadre retraité, de droite)
Polémique, critique	29	« Quelle attitude Mr Wauquiez. Ce n'est pas digne. #MotsCroisés »(homme, 27 ans, externe en médecine, de gauche)
Ironie, humour	25	« France-Espagne : Hollande parie sur un "nul"... Comme tous ceux qui ont parié sur lui ! » (homme, 42 ans, consultant en marketing, de droite)
Informatif	26	« La loi sur l'emploi divise la gauche en profondeur http://www.mediapart.fr/journal/france/260313/la-loi-sur-lemploi-divise-la-gauche-en-profondeur ... via @mediapart » (femme, 62 ans, secrétaire retraitée, militante LO)
Soutien	5	« @karlitozero Un grand merci pour votre lettre à Frigide Barjot! »(femme, 32 ans, illustratrice, sympathisante MoDem)
Argumentatif	14	« RT@ausalle "Choc de simplification pour les entreprises" ? la même que pour les créateurs avec l #autoentrepreneur ? Chiche? #Hollande @Hnovelli » (homme, 52 ans, juriste, militant Nouveau Centre)

À ce titre, la « technicité » des termes politiques mobilisés par les membres de notre panel peut être une autre variable intéressante à mobiliser afin d'évaluer – certes très partiellement – le degré de compétence politique de leurs auteurs. On peut ainsi classer, à partir d'observations empiriques, les termes politiques mobilisés dans les tweets étudiés en fonction de leur degré de proximité au discours politique dominant. À l'échelle de notre corpus, les résultats sont là encore assez homogènes et rendent compte de la très forte politisation de notre population : 19 % des tweets emploient des termes politiques « techniques » (« ANI », « RGPP », « CESE », etc.), 78 % des termes « usuels » (c'est-à-dire des termes couramment mobilisés dans des médias *omnibus* pour traiter des sujets politiques, tels que « parlement », « gouvernement », « ministre », « mariage pour tous », etc.) et seuls 3 %

⁷³⁰ Nous avons pleinement conscience que certains des résultats exposés ici ne sont analysés que très superficiellement. L'étude de l'ironie dans les autopublications numériques politiques pourrait ainsi être le sujet d'une thèse à part entière. L'enjeu n'est pas ici d'être le plus complet possible dans l'analyse, mais simplement de démontrer que la diversité des tweets politiques ne serait se réduire à une typologie uni-critère.

des tweets emploient des termes « profanes » (c'est-à-dire des termes qui sont le plus souvent mobilisés par des agents dominés de la politique : « politicard », « pourris », etc.⁷³¹).

2.1.1.3 Classer les tweets politiques en fonction de la nature politique de leur objet

Une troisième modalité de classification possible des tweets politiques consiste à examiner la nature de l'objet sur lequel porte le message.

Tableau 5.17 Nature de l'objet évoqué dans les tweets politiques (n= 1 108tweets)

Catégories	%	Exemple
Attribut politique d'un agent politique	62	« Comment le Parti de gauche peut-il se féliciter du "NON" au référendum alsacien ? Fier d'être du côté des joyeux drilles du FN ? #beark » (homme, 35 ans, professeur, militant MoDem)
Attribut non-politique d'un agent politique	21	« RT@FO_Giesbert Nadine Morano, la Sharon Stone de l'#UMP version low cost ? pic.twitter.com/p64kkPWfqx » (homme, 59 ans, consultant en management, de gauche)
Attribut politique d'un output politique	12	« RT@Senat_Direct ce texte comporte des avancées et nous le voterons malgré le recours au vote bloqué #ANI #directsénat » (femme, 35 ans, prépare concours attaché territoriale, de centre droite)
Attribut non-politique d'un output politique	1	« J'ai rempli ma déclaration d'#impôts en ligne et à l'estimation je me rends compte que pas mal de mes Re-venus vont Re-partir » (homme, 44 ans, technicien, de gauche)
Idéal, idéologie	4	« RT@breizhgauche les média sont presque tous la propriété de grands capitalistes ne les écoutez pas sans un minimum d'esprit critique ! » (homme, 46 ans, employé informatique, militant PCF)

On peut distinguer cinq catégories d'objets politiques : un attribut politique d'un agent politique (62 % des messages), un attribut non-politique d'un agent politique (21 %), un attribut politique d'un output politique (12 %), un attribut non-politique d'un output politique (1 %) ou encore des idées politiques (4 %).

⁷³¹ Précisons néanmoins que les individus qui emploient des termes politiques « profanes » ne sont pas forcément eux-mêmes peu compétents politiquement. Certains individus, notamment des militants, expriment leur goût ou leur dégoût envers des hommes politiques avec des expressions très crues. L'expression d'un dégoût se fera alors avec des termes très proches de ceux employés par des individus peu compétents.

Le tweet peut ainsi prendre pour objet un attribut politique d'un agent politique (cet agent peut être « individuel », comme un professionnel de la représentation politique, ou « institutionnel », comme un parti politique). Par *attribut politique*, nous entendons un critère d'évaluation qui soit spécifique au champ politique : alliance partisane, positionnement programmatique, respect d'une ligne idéologique, etc.

Mais les messages publiés sur Twitter peuvent également prendre pour objet un attribut non-exclusivement politique d'un agent politique. Par *attribut non-exclusivement politique*, nous entendons un critère d'évaluation qui ne soit pas spécifique au champ politique et qui peut s'appliquer à n'importe quel agent social : l'hexis corporelle, l'honnêteté, la sincérité, etc.

Cette même dichotomie entre attributs – ou critères d'évaluation – *politiques* et *non-politiques* peut être également appliquée à des *outputs* politiques. Nous définissons les *outputs* comme des *productions* politiques d'agents politiques (texte de loi, programme électoral, etc.). Les *outputs* politiques peuvent donc être eux aussi évalués à partir de critères d'évaluation spécifiquement politiques (efficacité supposée d'un projet de loi, jugement sur une politique publique, etc.). Mais ces *outputs* peuvent être également évalués à partir de critères de jugement non-spécifiquement politiques : ce sont alors par exemple les conséquences concrètes de l'*output* dans la vie quotidienne des agents extérieurs au champ politique qui seront discutées (comme par exemple les conséquences de la réforme des rythmes scolaires sur l'emploi du temps des parents d'élèves, etc.). Ces messages sont cependant très rares et ne représentent qu'un pourcent des tweets politiques de notre corpus⁷³².

Les tweets politiques peuvent enfin porter sur une dernière catégorie d'objet : les idées et les idéologiques politiques (le communisme, le communautarisme, la République, etc.). Ces messages représentent ainsi 4 % du corpus étudié. Il ne s'agit bien sûr pas de réflexions ni d'argumentations étayées (rappelons une dernière fois la taille limitée des tweets à 140 caractères) mais simplement d'allusions à des idées et des concepts de théorie politique. Les messages portant sur ces objets sont le fait d'individus très fortement politisés.

⁷³² Soulignons que la faiblesse de ce résultat est sans doute liée pour partie à notre méthodologie. C'est par une sélection par mots-clés politiques (cf. chapitre 4) que nous avons constitué le corpus de tweets politiques sur lequel nous avons appliqué cette grille d'évaluation. Si cette liste par mots-clés contenait des termes relatifs aux *policies* (avec des occurrences de type « fiscalité », ou « impôts », etc.) ces termes étaient en majorité situés plutôt du côté du langage dominant, excluant ainsi les tweets les plus « dépolitisés » sur les *policies*, et donc les plus enclins à prendre pour objet des attributs non politiques des *outputs* politiques.

On le voit donc bien, les tweets politiques peuvent être évalués et classés en fonction de nombreux critères d'analyse. Nous allons maintenant aborder le principe classificatoire qui nous semble le plus opérant pour répondre à notre problématique, c'est-à-dire déterminer dans quelle mesure les tweets politiques peuvent être appréhendés comme des opinions ou des « biens » nécessaires à la production d'opinions.

2.1.2 Classer les tweets politiques en fonction de leur format : distinguer opinion explicite et implicite

Deux variables définissent le format d'un tweet : la nature du tweet (tweet, retweet ou retweet commenté) et la présence – ou l'absence – de liens hypertextes (absence de liens, liens non commentés, liens commentés).

Nous l'avons déjà vu, sur Twitter, un individu peut soit publier un tweet qu'il a lui-même écrit, soit publier un tweet déjà existant, écrit et publié par un autre utilisateur. Contrairement aux retweets analysés dans la première section de chapitre, précisons que les auteurs originaux de ces messages ne sont pas nécessairement des comptes de médias, de journalistes ou d'hommes politiques. Les utilisateurs sur Twitter retweetent également des messages publiés par leurs abonnés, leurs amis, leurs camarades de classe ou encore même des inconnus⁷³³. On peut distinguer maintenant plus précisément deux catégories de retweets :

- des retweets *non commentés* : il s'agit d'une republication « pure et parfaite » d'un tweet original (exemple 1 : « RT @liberation_info 53% des Français préfèrent Nicolas Sarkozy à François Hollande <http://dlvr.it/34K7QK> »)
- des retweets *commentés* : l'individu republie sur son compte un tweet publié par un autre utilisateur mais y ajoute un ou plusieurs mots de commentaire personnel (exemple 2 : « Les petits cons ! RT @liberation_info 53% des Français préfèrent Nicolas Sarkozy à François Hollande <http://dlvr.it/34K7QK> »)

Dans le premier exemple, l'auteur 1 republie à l'identique sur son compte Twitter un tweet déjà publié par le compte Twitter de *Libération*. Le tweet original du quotidien relate

⁷³³ Pour avoir une idée de la répartition des retweets selon la catégorie de leurs auteurs originaux, cf. tableau 4.15 p. 308

les résultats d'un sondage d'apparence favorable à Nicolas Sarkozy. Ce retweet non commenté ne nous renseigne donc pas sur l'opinion que porte l'auteur 1 sur les résultats de ce sondage. Est-il satisfait de ces résultats ou au contraire mécontent ? Contesté-t-il la pertinence-même du sondage ? Nous avons vu plus haut que la pratique du retweet n'était pas neutre socialement et politiquement. Retweeter une information de *Libération* peut ainsi nous laisser *suggérer* que l'auteur du retweet est en accord avec la ligne politique du quotidien, et qu'il est donc positionné à gauche. Mais ce n'est qu'une probabilité. Avec comme seule information le contenu du retweet, le chercheur ne peut donc être totalement certain de l'opinion que se fait l'auteur 1 des résultats de ce sondage.

Dans le second exemple, l'auteur 2 publie sur son compte le même tweet que l'auteur 1 : il s'agit là encore du tweet publié originellement par le compte Twitter du quotidien *Libération*. Mais à la différence de l'auteur 1, l'auteur 2 commente le tweet originel en y ajoutant l'affirmation suivante en préambule : « *Les petits cons !* », désignant ainsi de manière humoristique « *les 53 % de Français [qui] préfèrent Nicolas Sarkozy à François Hollande* ». Avec ce commentaire, l'auteur 2 dévoile son opinion sur le contenu-même du bien médiatique qu'il a consommé. On peut alors déduire à la lecture de ce retweet que l'auteur 2 n'a pas une opinion favorable de Nicolas Sarkozy, reprochant ainsi aux Français de le préférer à François Hollande. On peut cependant objecter que l'exclamation de l'auteur 2 est ironique et est à interpréter au second degré. Cette objection est bien sûr recevable. À ce stade de notre démonstration on peut simplement se contenter de répondre que, d'une part, ces quelques mots de commentaire réduisent la marge d'incertitude interprétative *par rapport* au premier exemple donné, et que, d'autre part, l'obtention d'informations sociologiques sur l'auteur 2 (et notamment son positionnement politique) permettrait de déterminer s'il s'agit ou non d'une exclamation ironique.

La deuxième variable relative au format des tweets politique concerne la présence – ou l'absence – de lien hypertexte. Un tweet peut en effet contenir un lien hypertexte renvoyant le lecteur du tweet à une page internet (dans 7 cas sur 10, il s'agit du site internet d'un média⁷³⁴). De manière analogue aux retweets, on peut alors distinguer deux types de tweet avec lien hypertexte :

- Des liens hypertextes *non commentés* : l'auteur du tweet se contente de publier un lien hypertexte vers une page internet, quelque fois en l'accompagnant du

⁷³⁴ Cf. Tableau 4.16 p. 317.

titre de l'article vers lequel renvoie le lien (exemple 3 : «*Deux morts dans une fusillade à la kalachnikov à Marseille* http://actu.orange.fr/une/deux-morts-dans-une-fusillade-a-la-kalachnikov-a-marseille-afp_1403618.html »)

- Des liens hypertextes *commentés* : l'auteur publie un lien hypertexte en l'accompagnant de quelques mots de commentaire (exemple 4 : Félicitations #Hollande, #Valls et #Taubira, la France est apaisée http://actu.orange.fr/une/deux-morts-dans-une-fusillade-a-la-kalachnikov-a-marseille-afp_1403618.html ... Dormez tranquilles braves gens, vos dieux veillent).

Dans l'exemple 3, l'auteur se contente d'indiquer dans son tweet le lien vers l'article qu'il a lu (ou du moins parcouru) sur le portail d'information *Orange.fr*, en rappelant simplement le titre de l'article. Ce tweet ne permet pas au chercheur d'identifier l'opinion de l'auteur sur l'article qu'il a consommé et relayé sur son compte.

Dans l'exemple 4, l'auteur du tweet publie un lien hypertexte vers le même article que l'auteur 3. Mais à la différence de ce dernier, il fait part dans son message de son opinion à propos de l'information qu'il vient de consommer. En indiquant ces quelques mots « *Félicitations #Hollande, #Valls et #Taubira, la France est apaisée* », l'auteur du tweet entend signifier de manière ironique que le gouvernement socialiste n'est pas compétent pour gérer la sécurité publique. Dans ce cas, comme pour l'exemple 2 (le retweet commenté), le tweet politique étudié n'est pas seulement une matérialisation empirique d'un acte de consommation informationnelle mais bien une opinion délivrée à propos du sujet de l'article « consommé ».

Tableau 5.18 Format des tweets politiques publiés (n = 1 106)

Tweet et retweet	%	Exemple
Tweet	6	-
Retweet non commenté	8	« RT @liberation_info 53% des Français préfèrent Nicolas Sarkozy à François Hollande http://dlvr.it/34K7QK »
Retweet commenté	6	« Les petits cons ! RT @liberation_info 53% des Français préfèrent Nicolas Sarkozy à François Hollande http://dlvr.it/34K7QK »
Liens	%	Exemple
Absence de lien	4	-
Lien non commenté	1	« Deux morts dans une fusillade à la kalachnikov à Marseille http://actu.orange.fr/une/deux-morts-dans-une-fusillade-a-la-kalachnikov-a-marseille-afp_1403618.html ... »
Lien commenté	5	« Félicitations #Hollande, #Valls et #Taubira, la France est apaisée http://actu.orange.fr/une/deux-morts-dans-une-fusillade-a-la-kalachnikov-a-marseille-afp_1403618.html ... Dormez tranquilles braves gens, vos dieux veillent »

Classifier les tweets politiques en fonction de leur format permet donc de distinguer deux types de message :

- certains messages sont des opinions clairement explicitées sur des objets politiques. Ils sont parfois associés aux biens informationnels qui ont (ré)activé ou « alimenté » le procès de production de l'opinion.
- d'autres messages sont simplement des matérialisations empiriques de consommation « brutes », non contextualisées, de biens médiatiques.

Nous allons maintenant voir dans quelle mesure cette seconde catégorie de tweets politiques peut nous renseigner sur les opinions – non explicitement formulées – de leurs producteurs.

2.2 Opinions politiques et biens médiatiques

Peut-on déduire de ces consommations « brutes » d'information les opinions des individus qui les relaient ? Afin de répondre à cette question, il nous faut rappeler rapidement notre cadre de réflexion théorique relatif au processus de production des opinions politiques – cadre théorique que nous avons déjà exposé dans notre deuxième chapitre – en le confrontant ici à la réalité empirique de notre terrain : Twitter.

Nous avons ainsi proposé de concevoir les modes de production des opinions politiques comme des algorithmes, c'est-à-dire comme des systèmes de sélection et décodage des informations politiques, et de dissocier ainsi les algorithmes des données traitées par ces algorithmes, ou, si l'on veut rester dans un vocabulaire marxiste, de dissocier les *facteurs de production* des *modes de production* des opinions politiques. Les facteurs de production peuvent être des biens informationnels ou médiatiques, des expériences personnelles (récits biographiques), ou encore des stocks accumulés de connaissances (capitaux culturels et cognitifs). Nous avons également spécifié que les facteurs de production pouvaient être produits « en interne » par le producteur de l'opinion lui-même ou « importés » en provenance d'autres producteurs.

Sur Twitter, les liens hypertextes ou les retweets *non commentés* peuvent être appréhendés comme des facteurs de production. On peut ainsi repérer la nature de ces facteurs (bien médiatique, récit de vie, etc.), l'origine de ces facteurs (le facteur peut être « acheté »

directement sur le marché premier de l'information – le compte de *Libération* – ou sur le marché secondaire – le compte Twitter d'un ami) et leur poids relatif par rapport aux autres facteurs mobilisés dans le procès de production (combien de tweets de médias de gauche le producteur retweete-il en moyenne par rapport aux médias de droite).

Alors, même les tweets politiques qui ne sont que des consommations « brutes » et non contextualisées d'information politique peuvent, si l'on connaît les autres variables de l'algorithme, nous renseigner sur l'opinion des individus qui les consomment, même si celle-ci n'est pas clairement explicitée.

On peut ainsi schématiquement se représenter l'opinion d'un individu comme une équation à plusieurs inconnues et postuler que : 1) la consommation ostensible d'une information sur le compte Twitter d'un individu nous permet de suggérer que cet individu a une opinion sur l'objet de l'information qu'il vient de publier, 2) cette opinion est inconnue du chercheur, 3) d'autres variables de l'équation sont en revanche (potentiellement) connues du chercheur : position sociale, orientation politique, compétence politique, habitudes de consommations médiatiques, etc., 4) ces variables permettent de résoudre l'équation et d'identifier ainsi l'opinion de l'individu sur l'objet de l'information consommée.

Notre méthode de panélisation des réseaux sociaux nous permet d'obtenir ces informations nécessaires à la résolution de l'équation : la position des producteurs dans l'espace social, leur orientation politique, leur niveau de compétence politique mais également leurs habitudes en matière de consommation médiatique (*via* l'étude statistique de leurs abonnements ou retweets médiatiques). On peut donc à partir de ces quatre types de données (position sociale, orientation politique et habitudes de consommation médiatique) déterminer le mode de production d'un individu donné. Une fois déterminé ce mode de production – ou la formule de l'algorithme – on peut alors identifier l'opinion de cet individu sur un objet si l'on connaît un ou plusieurs des facteurs de production qu'il a mobilisés dans le procès de production.

En résumé, ce n'est donc qu'en analysant les biens consommés (retweets et liens) *au prisme* des propriétés sociologiques et politiques des individus qui les consomment que l'on pourra identifier les opinions de ces individus sur les objets des informations qu'ils choisissent de diffuser.

Conclusion du chapitre 5

Nous avons voulu démontrer dans ce chapitre que la production de tweets politiques obéissait à des logiques sociologiques et politiques très similaires à celles observées dans la « sphère hors-ligne ». Les propriétés sociologiques (âge, sexe, niveau de diplôme, PCS, etc.) et politiques (orientation politique, compétence politique, pratiques militantes, etc.) des individus sur Twitter orientent très clairement leurs choix en matière de consommation d'informations et d'opinions politiques.

Étudier Twitter avec un cadre d'analyse « bourdieusien » est heuristiquement enrichissant en cela que le réseau social matérialise, donne une réalité concrète, empirique, à des concepts jusque-là purement analytiques tels que ceux de « marché de l'opinion », de « consommation », de « biens informationnels » ou encore de « transactions ». La pratique du retweet matérialise ainsi les biens médiatiques effectivement consommés par un individu, et rend également manifeste la transaction de ces biens : quelle information un individu a-t-il « achetée » à quel média ?

Cependant, nous avons pu entrevoir en filigrane de notre démonstration qu'il n'existait pas *un* mais *plusieurs* marchés de l'opinion sur Twitter. À côté de ce marché « premier » qui confronte « offre » (les agents du champ de production de l'opinion) et « demande » (les agents situés à l'extérieur de ce champ) d'opinions et d'informations politiques existent en réalité des marchés « secondaires » où s'échangent de la « main à la main » des biens achetés sur le marché premier. Or notre méthodologie ne permet pas d'appréhender ces marchés secondaires, ou ces logiques de « viralité », de « bouche-à-oreille » ou encore d'« influences interpersonnelles », pour reprendre un vocabulaire plus proche des théories du marketing viral et de l'influence interpersonnelle. Il s'agit là d'une des limites de notre dispositif que nous tenterons de mettre en lumière dans les pages qui suivent.

CONCLUSION : LES APPORTS ET LIMITES DE LA MÉTHODE DE PANÉLISATION DES RÉSEAUX SOCIAUX

Bientôt arrivés au terme de notre démonstration, nous proposons de dresser ici un bilan des apports mais aussi des limites de notre méthode de panélisation des réseaux sociaux.

Le principal défi de cette recherche consistait à construire un dispositif méthodologique opérant afin d'analyser les opinions politiques publiées par les agents sociaux sur internet. L'arrière-plan théorique qui a guidé notre démarche est clairement inspiré des thèses de Pierre Bourdieu relatives à la production des opinions politiques⁷³⁵ et à la déconstruction de la notion d'opinion publique⁷³⁶. L'enjeu était bien de concevoir une méthode permettant d'étudier rationnellement les opinions politiques publiées sur internet, en situant socialement et politiquement leurs producteurs, et ce sans reproduire le principal biais des enquêtes par questionnaire : l'imposition de problématique.

Afin de respecter ces objectifs, le dispositif méthodologique imaginé devait alors répondre à trois exigences : obtenir un nombre important d'informations comparables sur une population suffisamment nombreuse pour autoriser le traitement statistique, pouvoir qualifier sociologiquement et politiquement cette population et analyser sa production de messages sans interagir avec elle.

Échantillonner aléatoirement une population d'internautes et lui administrer un court questionnaire permet de concilier ces trois exigences. Une fois le questionnaire complété par les individus tirés au sort, plus aucune interaction n'a lieu entre le chercheur et les enquêtés. Le chercheur peut ainsi, à l'aide d'un logiciel spécialisé, archiver et collecter de façon exhaustive les messages publiés par les enquêtés et les rattacher systématiquement aux propriétés sociologiques et politiques de leurs auteurs.

Nous avons expérimenté ce dispositif méthodologique sur un terrain numérique particulier : le réseau social – ou site de *micro-blogging* – Twitter. Cette *expérimentation méthodologique* a alors généré une production importante de *données empiriques*.

⁷³⁵ Pierre Bourdieu, *La Distinction. Critique sociale du jugement*, Paris, Les Éditions de Minuit, 1979.

⁷³⁶ Pierre Bourdieu, « L'opinion publique n'existe pas », *Les Temps modernes*, n°318, janvier 1973, p. 1292-1309.

Ces données confirment d'une part des résultats déjà bien établis par la sociologie politique, concernant notamment l'inégale propension des agents sociaux à s'exprimer politiquement. Mais ces données permettent également de *valider empiriquement* certaines hypothèses qui ne l'avaient que peu été jusqu'alors. Ainsi, si l'importance des facteurs conjoncturels sur le niveau de politisation des discussions ordinaires a été soulignée dans plusieurs recherches ethnographiques⁷³⁷, rares ont été les travaux – notamment francophones – qui ont mobilisé des données quantitatives afin de le démontrer. L'analyse longitudinale de la production de nos panels a permis de valider empiriquement cette hypothèse d'une inégale distribution dans le temps des prises de parole politiques. Surtout, en comparant nos données avec celles d'un panel d'agents appartenant au champ de production idéologique, nous avons pu démontrer l'existence d'une corrélation statistique entre le niveau de l'offre et celui de la demande d'informations et d'opinions politiques sur le réseau social.

Nous reviendrons plus précisément dans la première section de cette conclusion sur ces données et ces résultats en nous interrogeant sur ce qu'ils peuvent apporter au débat théorique relatif à la notion d'opinion publique. Nous resituerons ainsi notre thèse d'une détermination de la demande par l'offre au regard de ces débats (1).

Cette discussion nous amènera alors à nous interroger sur les limites de notre dispositif. En souhaitant étudier la « demande » d'opinions – à travers notamment la mise en place d'un panel représentatif de producteurs de tweets politiques – nous avons finalement, de façon assez paradoxale, recentré la focale sur l'« offre ». Or notre dispositif ne permet pas d'appréhender pleinement les logiques spécifiques qui structurent l'« offre » d'opinions et d'informations politiques sur Twitter. Mais notre dispositif se révèle également, dans une certaine mesure, relativement inapte à restituer les logiques de consommation « horizontales » au sein des marchés « secondaires » de l'opinion (2).

Nous proposerons alors dans une dernière section de discuter la pertinence de quelques dispositifs méthodologiques alternatifs permettant de dépasser potentiellement les limites évoquées. Étudier les logiques de circulation des opinions à l'échelle de l'ensemble du réseau social – et non au sein d'un échantillon raisonné – peut ainsi permettre de mieux appréhender certaines logiques qui échappent à notre dispositif (3).

⁷³⁷ Nina Eliasoph, *L'évitement du politique. Comment les Américains produisent l'apathie dans la vie quotidienne*, Paris, Economica, 2010 ; Vincent Goulet, *Médias et classes populaires. Les usages ordinaires des informations*, Paris, INA Édition, 2010 ; Céline Braconnier et Jean-Yves Dormagen, *La démocratie de l'abstention : Aux origines de la démobilisation en milieu populaire*, Paris, Gallimard, 2007.

1. La vérification empirique d'hypothèses souvent peu objectivées

Nous n'évoquerons pas ici les avantages *méthodologiques* de notre dispositif, déjà exposés lors de notre deuxième chapitre (qualification sociologique des auteurs des messages, non-imposition de problématique, etc.) mais reviendrons davantage sur les apports *théoriques* que nous en avons tirés.

L'expérimentation de notre dispositif a d'abord permis de générer des données confirmant l'existence d'un véritable « cens caché »⁷³⁸ à la prise de parole politique sur internet mais aussi de quantifier précisément cette (non) prise de parole (*I.1*). Nous avons ensuite pu mettre en évidence que cette prise de parole politique était inégale dans le temps et pouvait s'expliquer par l'existence d'une corrélation entre le niveau de l'offre et le niveau de la demande d'opinions et d'informations politiques sur Twitter (*I.2*). Bien que présente en filigrane dans certains travaux, l'affirmation d'une détermination de la demande par l'offre a rarement été posée comme telle dans le débat sur l'opinion publique (*I.3*).

1.1 Quantifier la (non) prise de parole politique

L'expérimentation de notre dispositif méthodologique a d'abord permis de générer des données empiriques qui confirment un résultat bien connu de la sociologie politique : l'inégale propension des agents sociaux à exprimer un avis ou une opinion sur la politique.

Deux postures méthodologiques permettent généralement aux chercheurs de démontrer empiriquement ces inégalités de politisation. La première consiste à administrer un questionnaire comprenant des questions de connaissances politiques et/ou d'« opinions » à un échantillon représentatif et d'analyser les sous-catégories de la population qui acceptent de répondre, ou plus exactement celles qui ne répondent pas ou ne se « prononcent » pas. Les « non-réponses » et les « *nsp* » ont ainsi longtemps constitué le principal matériau empirique mobilisé (en première et surtout en seconde main⁷³⁹) par les chercheurs afin de démontrer les

⁷³⁸ Daniel Gaxie, *Le cens caché. Inégalités culturelles et ségrégation politique*, Édition du Seuil, 1978.

⁷³⁹ Dans les travaux français, les chercheurs qui entendent démontrer les inégalités de compétence politique ont en effet plus souvent tendance à mobiliser des données de sondages réalisés par des instituts privés plutôt que de réaliser eux-mêmes des enquêtes quantitatives.

inégalités de compétence politique des « citoyens ordinaires »⁷⁴⁰. La seconde posture, relativement similaire, consiste à analyser une population d'agents sociaux qui acceptent *volontairement* de répondre à des enquêtes politiques ou de participer à des réunions politiques – des échantillons dits « spontanés » – puis de comparer la sociologie de ces échantillons à celle de la population totale.

Notre dispositif méthodologique se situe quant à lui dans une sorte d'entre-deux. Nous avons ainsi administré un questionnaire à un échantillon représentatif d'agents sociaux. Seulement, ces agents inscrits sur Twitter, en publiant durant la campagne électorale du printemps 2012 un ou plusieurs tweets politiques, se sont déjà tous « volontairement »⁷⁴¹ exprimés politiquement. Les individus qui ont accepté de répondre à notre questionnaire constituent donc un échantillon représentatif d'un échantillon dit « spontané ».

S'étant tous exprimés politiquement sur Twitter, on a donc pu supposer qu'il existait peu de différences sociologiques entre les « répondants » et l'ensemble des individus qui composent notre échantillon, les producteurs de tweets politiques. En constituant en parallèle de ce panel de « répondants » un panel d'individus qui n'ont pas répondu au questionnaire, nous avons pu vérifier empiriquement cette hypothèse. Nous avons alors pu constater qu'il existait en effet peu de différences sociologiques entre les « répondants » et les « non répondants » ou, pour être plus précis, entre les « répondants » et les « non répondants » dont la position sociale a pu être identifiée.

Dans environ un cas sur deux, nous n'avons en effet pas pu définir les propriétés sociologiques des individus qui n'ont pas répondu au questionnaire. Or certains indices nous laissent suggérer qu'une fraction de ces « non répondants » provient d'un milieu social plus modeste, comme en atteste par exemple leurs choix en matière d'abonnement à des comptes artistiques, leurs photographies et descriptions de profil, leur niveau de langage, etc. Nous l'avons vu, ces indices sont autant de manifestations empiriques de certaines dispositions constitutives de l'habitus. Sur les réseaux sociaux, la position sociale des agents ne s'objective pas seulement en fonction de données socio-démographiques déclarées : les pratiques sociales des internautes peuvent également être analysées par le chercheur afin d'identifier leur milieu social.

⁷⁴⁰ Définissons simplement ici les « citoyens ordinaires » comme des agents sociaux n'appartenant pas au champ politique.

⁷⁴¹ Sans que nous les ayons directement incités à le faire.

L'automatisation de l'analyse de ces pratiques sociales constitue sans doute une perspective de recherche prometteuse. On pourrait en effet imaginer à terme que les logiciels de collecte soient en capacité de recenser automatiquement les pratiques sociales et culturelles des internautes (via notamment l'analyse de leurs abonnements ou de leur propension à « aimer » des pages de chanteurs, de films, de programmes télévisés, de marques, de femmes et d'hommes politiques, etc.) et puissent alors déterminer, en fonction d'une cartographie de l'espace social où figureraient, comme dans La Distinction, les pratiques culturelles, les PCS et les électorats politiques, le milieu social de ces individus. Nous y reviendrons à la fin de cette conclusion.

S'il existe donc certaines différences entre les individus qui ont répondu à notre questionnaire et ceux qui, pour diverses raisons⁷⁴², ne l'ont pas fait, nos résultats ont permis toutefois d'affirmer que les individus producteurs de tweets politiques sont, dans leur très grande majorité, bien plus politisés que la moyenne des Français. Lorsqu'un chercheur – ou un « *web analyst* » – entreprend de « sonder » les opinions politiques sur Twitter, il mesure en réalité les opinions du « pays légal », à savoir les opinions des agents sociaux qui se sentent en droit – et en devoir – de s'intéresser aux affaires politiques de leur pays et d'exprimer leur prise de position.

Pour reprendre les termes de Pierre Bourdieu, l'échantillon spontané – ici les individus inscrits sur Twitter qui publient des messages politiques – est bien la limite vers laquelle tend la population des répondants d'une enquête auprès d'un échantillon représentatif⁷⁴³. Nous avons ainsi montré que les producteurs de tweets politiques étaient, à l'instar des individus qui répondent aux enquêtes électorales, majoritairement des hommes, fortement diplômés, qui habitent dans des grandes métropoles urbaines – dont notamment Paris – et qui occupent une position sociale élevée. Et si la très nette sur-représentation des catégories les plus jeunes sur Twitter est une caractéristique sociologique qui différencie *a priori* notre population des populations classiquement « participationnistes », l'analyse des données issues de notre questionnaire nous a permis de démontrer qu'il s'agit en réalité de la fraction la plus politisée de la jeunesse qui s'exprime politiquement sur Twitter et que ses membres sont issus, comme

⁷⁴² Précisons une nouvelle fois que rien ne nous permet d'affirmer que les non répondants aient explicitement refusé de répondre à notre questionnaire.

⁷⁴³ Pierre Bourdieu, *La Distinction*, *op. cit.*

l'ensemble des producteurs de tweets politiques, des fractions supérieures des classes moyennes et des classes supérieures.

Analyser les propriétés des agents sociaux qui s'expriment politiquement sur Twitter permet donc d'affirmer que, sur internet également, la prise de parole politique obéit à des mécanismes de sélection sociale, à un « cens caché » pour reprendre l'expression de Daniel Gaxie⁷⁴⁴. En cela, notre dispositif ne fait que confirmer un résultat déjà bien connu de la sociologie politique.

Cependant, notre dispositif permet également d'analyser avec plus de précision et de finesse les différences qui existent à l'intérieur même de cette population politisée, là où les enquêtes par questionnaire « classiques » ont tendance à adopter une grille d'analyse relativement « binaire » de la prise de parole politique. Ces recherches opposent ainsi les répondants et les non répondants aux questions d'opinions, les individus qui s'expriment politiquement à ceux qui ne s'expriment pas. Notre dispositif permet de distinguer, au sein même de la catégorie des agents qui se sentent autoriser à s'exprimer politiquement, ceux qui s'expriment « beaucoup » de ceux qui s'expriment « peu ». Au sein de notre population de producteurs de tweets politiques, nous avons en effet pu constater qu'il existait de très importants écarts en termes de volume de production de tweets politiques.

Les facteurs explicatifs de ces inégalités de production sont en partie les mêmes que ceux qui permettent d'isoler les individus qui s'expriment politiquement de ceux qui ne s'expriment pas : la PCS, l'âge ou encore l'intérêt *déclaré* et *objectif* pour la politique (tel qu'il peut se manifester à travers le nombre d'abonnements à des comptes de femmes ou d'hommes politiques).

Cependant, le nombre de tweets politiques produits est également corrélé à une autre variable : le nombre *total* de tweets publiés. Les individus qui s'expriment le plus *politiquement* sur Twitter sont également ceux qui s'expriment le plus *sur tous les sujets*. En revanche, la réciproque n'est pas exacte : les individus qui publient un très grand nombre de tweets ne sont pas nécessairement ceux qui publient le plus de tweets *politiques*. Précisons cependant que même ces très faibles producteurs de tweets politiques sont bien plus politisés que la moyenne des Français.

⁷⁴⁴ Daniel Gaxie, *Le cens caché, op. cit.*

À travers ce résultat, se trouve alors posée la question de la relation entre l'expression politique *en ligne* et l'expression politique *hors ligne*. Quels sont les facteurs qui favorisent le passage de l'expression politique hors-ligne (telle qu'elle peut se manifester lors d'expériences très institutionnalisées comme le vote ou au contraire lors d'expériences très informelles comme les discussions au sein de groupes primaires) à la publication de tweets politiques ? Si nos résultats permettent d'affirmer que seuls les individus les plus politisés s'expriment politiquement sur Twitter, on ne peut en revanche affirmer que tous les individus politisés s'expriment politiquement sur le réseau social. Quels sont alors les facteurs qui motivent les individus politisés à exprimer et publiciser leurs opinions sur Twitter ? Nous n'avons pas été en mesure de répondre à cette question et d'analyser plus finement la relation existante entre le niveau de politisation et la propension à s'exprimer politiquement sur Twitter.

S'il s'agit bien là d'une limite de notre recherche, elle n'est pas imputable à notre méthode, mais à la population que nous avons choisi d'étudier. Nous avons en effet fait le choix d'échantillonner non pas la population totale des inscrits au réseau social, mais une population d'individus qui avaient déjà publié un ou plusieurs tweets politiques⁷⁴⁵. Si nous avions fait le choix d'échantillonner la population totale des inscrits ou bien des individus qui au contraire ne s'étaient pas exprimés politiquement auparavant, il aurait été alors possible d'identifier, *via* notre questionnaire, des inscrits très politisés, qui ont l'habitude de prendre la parole dans la sphère hors ligne, mais qui ne publient jamais de tweets politiques. Réaliser des entretiens ou des *focus groups* avec ces individus pourrait sans doute permettre d'identifier, au sein même de ces populations politisées, les logiques qui président à la publication – et la publicisation – des opinions politiques sur internet.

En disant cela, nous soulignons la nécessité absolue de croiser les méthodes, et notamment les méthodes numériques avec les méthodes « classiques » des sciences sociales (entretiens, observations ethnographiques, etc.). Si l'objet de cette conclusion, et plus généralement de cette recherche, consiste pour partie à valoriser scientifiquement notre dispositif méthodologique, il convient de préciser que quels que soient ses apports et ses limites, notre méthode, comme n'importe quelle autre, ne pourrait et ne saurait bien sûr se substituer aux méthodes existantes.

⁷⁴⁵ Pour les raisons qui nous ont guidé vers ce parti-pris, cf. chapitre 2.

Les données générées par notre dispositif permettent donc de démontrer l'existence d'un cens caché à l'expression politique sur internet et de quantifier avec plus de finesse ces prises de parole. Mais notre dispositif permet également de générer d'autres types de données qui permettent notamment de valider empiriquement des hypothèses qui ne l'ont que peu été jusqu'alors.

1.2 Démontrer l'inégale distribution dans le temps des expressions politiques

En archivant de façon exhaustive les tweets publiés par nos enquêtés, nous avons pu, à l'aide de notre logiciel de collecte, analyser le volume de production quotidien de tweets politiques de nos panels durant 11 mois consécutifs. Cette analyse longitudinale nous a alors permis d'apporter une preuve empirique à une hypothèse formulée dans de nombreux travaux ethnographiques mais relativement peu objectivée *quantitativement* : la différenciation du niveau de politisation des discussions ordinaires dans le temps et sa détermination par des facteurs conjoncturels.

Ainsi, c'est en observant les discussions de voisinage au sein des espaces publics d'un quartier populaire que Vincent Goulet a pu remarquer que les membres des classes populaires parlaient davantage de politique lorsqu'elle était à la « une » des journaux, notamment télévisés⁷⁴⁶. Cette hypothèse a été en revanche plus rarement objectivée quantitativement. Et ce pour une raison assez simple : les discussions politiques informelles sont insaisissables. Sauf à équiper l'ensemble des habitants d'un pays d'un microphone enregistreur, il est en effet impossible de recenser les discussions politiques échangées entre individus, et encore moins d'en déterminer leurs fréquences. Les chercheurs sont alors contraints de recourir à des données déclaratives, à savoir à des réponses à des questions du type « *au cours de ces x derniers mois, avez-vous parlé de politique avec...* ». Or, d'une part, ces questions n'autorisent souvent que deux classes de réponse (« *a parlé de politique sur la séquence* » ou « *n'a pas parlé de politique sur la séquence* »). D'autre part, ces questions ne permettent pas de déterminer si les individus ont parlé de politique, mais simplement s'ils ont *le souvenir d'en avoir parlé*. Si l'on admet assez facilement que les individus oublient régulièrement les

⁷⁴⁶ Vincent Goulet, *Médias et classes populaires*, op. cit.

élections pour lesquelles ils ont ou non voté par le passé, on peut alors logiquement supposer qu'ils aient encore plus de mal à se souvenir des discussions politiques informelles qu'ils ont pu avoir dans leur intimité ou dans leur vie sociale. De plus, les agents sociaux, notamment les moins politisés, n'ont pas toujours la même perception que les chercheurs de ce qu'est – ou de ce que devrait être – une discussion « politique ».

Notre dispositif permet de pallier ces limites inhérentes aux méthodes déclaratives. Nous n'avons pas demandé à nos enquêtés s'ils publiaient des tweets politiques, s'ils avaient le souvenir d'en avoir publié et à quelle fréquence, pas plus que nous leur avons demandé ce qu'ils entendaient par « tweets politiques ». Nous nous sommes contenté d'enregistrer et d'analyser le nombre de tweets politiques (que nous avons définis comme étant des tweets contenant au moins une occurrence relative à la *politics* parmi les 110 termes recensés⁷⁴⁷) publiés jour après jour par chacun des membres de nos panels, et ce durant une période de 11 mois.

Nous avons alors observé que le nombre de tweets publiés par l'ensemble de nos panels était très inégal dans le temps. La courbe représentant graphiquement (*cf.* graphique 4.1, p. 306) cette distribution met ainsi en évidence l'existence de quelques « pics » isolés de production de messages politiques. On a alors pu observer que, d'une part, ces pics étaient communs à toutes les catégories de producteurs, des plus politisées aux moins politisées, et que, d'autre part, ces pics se produisaient lors d'« événements politiques ».

Plutôt que d'essayer de définir de manière intrinsèque la force d'un « événement politique », nous avons simplement proposé de considérer ces « événements » comme des surproductions de biens médiatiques politiques. Nous avons ainsi pu remarquer, d'abord de manière relativement intuitive et inductive, que chaque « événement » recensé avait généré un très grand nombre d'articles et de sujets dans les journaux télévisés et dans les titres de presse écrite (papier et numérique). Constatant que chaque média, quel que soit son support d'origine (télévision, radio, presse écrite, presse en ligne) disposait aujourd'hui d'un compte Twitter, il était alors tentant de mesurer précisément la distribution dans le temps du nombre de tweets politiques publiés par ces comptes et de la comparer à celle de nos panels représentatifs (« répondants » et « non répondants »).

Notre dispositif méthodologique ne permettant pas de mesurer la production de ces comptes médiatiques, nous avons alors mobilisé le panel « Twitter audience politique »

⁷⁴⁷ *Cf.* chapitre 3.

réalisé dans le cadre d'une recherche parallèle à ce travail doctoral⁷⁴⁸. Ce panel comprenait alors non seulement les comptes des principaux médias d'information, mais également ceux d'une très grande partie des agents appartenant au champ de production idéologique (journalistes politiques, éditorialistes, « experts », hommes politiques, etc.).

Nous avons alors pu démontrer l'existence d'une corrélation statistique entre le niveau de tweets politiques publiés par ces comptes et le niveau de tweets politiques publiés par nos panels représentatifs. S'inspirant du cadre analytique mobilisé par Pierre Bourdieu, nous avons alors proposé d'appréhender Twitter comme un marché de l'opinion où se rencontrent une « offre » – les tweets publiés par les comptes des agents du champ de production idéologique – et « une demande », dont nos panels « répondants » et « non répondants » constituent un échantillon représentatif.

Enfin, nous avons pu démontrer, en analysant les logiques d'abonnement et de retweet de nos enquêtés, que leurs choix en matière de consommation d'informations et d'opinions politiques sur Twitter obéissaient à des logiques socialement et politiquement déterminées.

Nous proposons maintenant de discuter plus précisément cette hypothèse d'une détermination de la demande par l'offre et de la resituer dans le débat académique relatif à la notion d'opinion publique.

1.3 L'hypothèse d'une détermination de la demande par l'offre au prisme du débat sur l'opinion publique

L'hypothèse avancée ici, selon laquelle le niveau de l'offre détermine celui de la demande, peut rappeler d'une certaine manière les théories de l'*agenda-setting*, que l'on peut résumer en reprenant la célèbre formule de Mc Combs et Shaw :

« The press may not be successful much of the time in telling people what to think, but it is stunningly successful in telling its readers what to... think about »⁷⁴⁹.

⁷⁴⁸ Et nous tenons à remercier à ce titre une nouvelle fois l'ensemble des personnes qui ont contribué à sa réalisation : Emmanuelle Rimbault, Jean-Yves Dormagen, Marie Neihouser, Rémi Andris, Geoffrey Della Valentina, Émeline Rochefeuille et Lysne Fröydis.

⁷⁴⁹ Max Mac Combs et Donald Shaw, «The agenda-setting function of mass-media », *Public Opinion Quarterly*, n°36, 1972, p. 176-187.

Pour arriver à cette conclusion, les deux auteurs ont comparé l'ordre des enjeux *déclarés* comme étant les plus importants par un échantillon de 100 Américains interrogés par questionnaire durant une campagne électorale et la quantité d'articles consacrés à ces mêmes sujets dans les médias (presse écrite et télévision). Mc Combs et Shaw constatent alors une corrélation forte entre l'importance des enjeux accordés par les médias et celle attribuée par les électeurs. Les auteurs concluent alors que plus un enjeu est exposé dans les médias, plus il deviendra important « aux yeux de l'opinion publique ».

G. Ray Funkhouser arrivera à la même conclusion en étudiant cette fois des « enjeux » sur une plus longue période – la décennie 1960 – en comparant notamment des données de sondages avec la volumétrie des articles publiés dans la presse⁷⁵⁰. L'auteur constate alors lui aussi « *une relation nette entre le degré de préoccupation des citoyens exprimé à propos de certains problèmes et la quantité d'information consacrée par les médias à ces mêmes sujets* »⁷⁵¹.

On peut donc constater, en apparence du moins, une certaine analogie entre la démarche et les conclusions des auteurs des théories de l'*agenda-setting* et celles qui sont les nôtres. On pourrait ainsi considérer la quantité d'informations consacrées par les médias aux enjeux étudiés comme une variable mesurant *le niveau de l'offre d'opinions* et le degré des préoccupations des citoyens mesuré par sondage comme une variable mesurant *le niveau de la demande*. Constaté alors que le nombre d'articles consacrés à un sujet rend ce sujet plus « préoccupant » aux « yeux de l'opinion publique » reviendrait d'une certaine manière à affirmer que le niveau de l'offre d'opinions détermine bien celui de la demande.

Précisons néanmoins que, contrairement à ces travaux, nous n'avons pas comparé différents enjeux ou différentes idées-forces – c'est-à-dire différents produits en concurrence sur le marché de l'opinion – mais le niveau global de l'offre et de la demande. Ainsi, comparer à l'avenir plusieurs idées-forces ou plusieurs « enjeux » en concurrence sur le marché de l'opinion nous paraît être une suite logique à donner à cette recherche.

Cependant, en s'inscrivant pour la plupart dans la prolongation des thèses de Paul Lazarsfeld relatives aux effets limités des médias sur les comportements politiques des électeurs, ces travaux n'interrogent pas de façon critique la notion d'opinion publique et son

⁷⁵⁰ G. Ray Funkhouser, « The issues of the sixties: an exploratory study in the dynamics of public opinion », *Public Opinion Quarterly*, n°37, 1973, p. 62-75.

⁷⁵¹ Dorine Bregman. « La fonction d'agenda : une problématique en devenir », *Hermès, La Revue*, 1989/1, n°4, p. 195.

arrière-plan théorique et normatif. L'opinion publique est simplement appréhendée comme ce qui est mesurée par les sondages d'opinion.

Nous avons vu au cours de notre deuxième chapitre⁷⁵² que cette équivalence entre résultats de sondage et opinion publique a été déconstruite par de nombreux travaux de recherche. Nous avons, en conclusion de notre recension, identifier trois conceptions différentes de l'opinion publique : l'opinion publique sondagière, l'opinion publique « dissimulée » (le « *système de forces et de tensions* » évoqué par Pierre Bourdieu dans « L'opinion publique n'existe pas »⁷⁵³) et l'opinion publique « délibérative » développée notamment dans les travaux de James Fishkin⁷⁵⁴.

En postulant implicitement que les préoccupations des sondés sont les mêmes que celles des sondeurs, les sondages d'opinion assimilent de façon indifférenciée « offre » et « demande » d'opinions. Tous les citoyens étant considérés comme parfaitement égaux et compétents politiquement, il ne saurait alors exister de différences entre les journalistes et les ouvriers, entre les agents du champ de production idéologique et les agents situés à l'extérieur de ce champ. Comme les bulletins de vote glissés dans l'urne, les opinions des citoyens auraient donc toutes la même valeur, quelle que soit la position sociale et politique de leur auteur. L'opinion d'un ouvrier aurait donc le même poids que celle d'un éditorialiste du *Monde*.

James Fishkin et les auteurs des théories délibératives prennent acte au contraire d'une relative indifférence des citoyens aux affaires publiques et tentent d'y remédier en mettant en place des expérimentations politiques, qui sont aussi des expériences scientifiques : les sondages délibératifs. Ces sondages entendent améliorer la qualité de la délibération citoyenne relative à un enjeu donné en mettant à la portée immédiate des citoyens « sondés » des informations et des éléments de compréhension et d'analyse. Contrairement aux sondages traditionnels, les sondages délibératifs ne prétendent donc pas seulement mesurer l'opinion publique *telle qu'elle est*, mais *telle qu'elle devrait être* et *telle qu'elle pourrait être*. Le sondage post-délibération mesure ainsi une opinion publique « virtuelle », en puissance : l'opinion publique telle qu'elle pourrait être si tous les citoyens étaient « correctement » informés des enjeux.

⁷⁵² Cf. chapitre 2.

⁷⁵³ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

⁷⁵⁴ James Fishkin, *Democracy and Deliberation*, Yale, Yale University Press, 1991.

On pourrait alors voir derrière cette proposition théorique (et politique) une reformulation de notre hypothèse : augmenter l'offre d'informations, ou plutôt la consommation *effective* d'informations (en « contraignant » les citoyens à consommer ces informations) contribuerait à augmenter la demande d'opinions, à savoir le nombre d'agents ayant effectivement une opinion politiquement constituée sur le sujet traité.

Mais c'est à travers la conception bourdieusienne de l'opinion publique que notre hypothèse d'une détermination de la demande par l'offre devient réellement opérante et prend tout son sens. Dans l'« opinion publique n'existe pas », Bourdieu affirme que la fonction principale des sondages est de dissimuler que « *l'état de l'opinion à un moment donné du temps est un système de forces, de tensions* »⁷⁵⁵. Les opinions sont donc appréhendées comme des idées-forces portées, défendues et incarnées par des groupes politiques, plus ou moins mobilisés, plus ou moins nombreux numériquement et plus ou moins en capacité de mobiliser les agents sociaux.

Dans *La Distinction*, Pierre Bourdieu ne fait plus référence à « un système de forces et de tensions » mais à un « marché de l'opinion ». L'auteur met alors en exergue l'existence d'une frontière séparant d'une part « *des agents et des institutions en concurrence pour le monopole de la production et de l'imposition de la représentation légitime du monde social* » – l'offre d'opinions – et d'autre part des agents « *qui, faute de pouvoir participer activement au jeu proprement politique, qui est de facto réservé aux professionnels à plein temps (hommes politiques, permanents des appareils – mot adéquat en ce cas – de partis, journalistes politiques, idéologues professionnels)* » – la « demande » – n'ont d'autre choix que de consommer des « *opinions toutes faites* » [...], des « *produits offerts* » caractérisés par une « *marque de fabrique* »⁷⁵⁶. Pour Bourdieu, la frontière entre offre et demande d'opinion est donc très difficilement franchissable. Les « profanes », aussi politisés soient-ils, n'ont pas pour la plupart les moyens de produire et surtout de diffuser leurs « propres » opinions, produites à partir de leur petit moyen de production.

En analysant par la suite les logiques qui orientent les choix des différentes classes et fractions de classe vers différents titres de presse, Pierre Bourdieu s'est alors concentré sur les logiques *particulières* de consommation des opinions. Dit autrement, l'auteur ne s'est pas intéressé aux structures, aux mécanismes et aux lois d'ensemble de ce marché de l'opinion, en

⁷⁵⁵ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

⁷⁵⁶ Pierre Bourdieu, *La Distinction*, op. cit., p. 504.

étudiant par exemple l'impact du niveau *global* de l'offre à un moment du temps sur le niveau de la demande.

On retrouve cependant cette idée d'une détermination du niveau de la demande d'opinions par le niveau de l'offre dans l'« opinion publique n'existe pas », lorsque l'auteur affirme que dans « *les situations de crise* [soit des périodes de surproduction de l'offre d'opinions], *les gens* [la demande] *sont devant des opinions constituées* [qui sont alors plus nombreuses qu'en situation de non-crise] »⁷⁵⁷. La situation de crise politique correspond donc bien à un niveau élevé de l'offre, et ce niveau élevé de l'offre entraîne une élévation du niveau de la demande.

Notre dispositif méthodologique, appliqué à ce terrain particulier qu'est Twitter, a donc permis de matérialiser empiriquement ce marché de l'opinion, en identifiant d'une part l'offre et la demande d'opinion, et en analysant d'autre part leurs relations.

En revanche, notre dispositif n'a pas permis de rendre compte des tensions et des concurrences qui traversent l'offre, ni des logiques de consommation « horizontales » qui parcourent la demande.

2. Les limites de la logique de panélisation

Alors que notre volonté initiale consistait à étudier un panel représentatif de producteurs de tweets politiques, nous avons finalement été amené à recentrer notre focale sur des comptes d'agents et d'institutions appartenant au champ de production idéologique. En étudiant plus finement les comptes des individus qui composent notre panel représentatif, et notamment leurs logiques d'abonnement, nous nous sommes en effet rendu compte de l'existence d'une frontière relativement étanche entre offre et demande d'opinions sur Twitter. Les individus s'abonnent et retweetent en réalité très peu de comptes différents et ces comptes appartiennent, dans leur très grande majorité, au champ de production idéologique. Or notre dispositif d'échantillonnage aléatoire de la population des inscrits au réseau social ne

⁷⁵⁷ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

permet pas d'étudier cette seconde catégorie de comptes (l'offre d'opinions). Pour partie calquée sur la technologie du sondage, notre méthode donne en effet une chance égale à tous les producteurs de tweets politiques – qu'ils soient employés ou journalistes politiques – de figurer dans notre panel. Or du côté de l'offre, les comptes sont numériquement bien moins nombreux que du côté de la demande (environ 200 comptes sur l'ensemble du réseau social), la chance pour un compte de se retrouver dans notre panel représentatif est donc très faible (4 chances sur 1000).

Afin d'étudier cette offre, nous avons alors mobilisé le panel « Twitter audience politique », réalisé dans le cadre d'un projet de recherche parallèle à ce travail doctoral. Ce dispositif repose sur la même démarche méthodologique que la nôtre : il s'agit en effet d'étudier un échantillon de comptes – et non l'ensemble du réseau social – afin de pouvoir qualifier plus finement les individus étudiés. Seules changent les modalités de sélection de l'échantillon panélisté. Les comptes qui composent ce panel « Twitter audience politique » ne sont pas un échantillon représentatif de producteurs de tweets politiques, mais correspondent à un classement hiérarchisé des comptes politiques bénéficiant de la plus forte audience sur le réseau social. L'étude parallèle de ces deux panels – représentatif et influence – nous a alors permis de mesurer une corrélation statistique entre le niveau de production de l'offre et celui de la demande.

Cependant, cette logique d'échantillonnage, qu'elle soit appliquée à la demande (panel représentatif) ou à l'offre (panel « Twitter audience politique »), comporte dans tous les cas une même limite : elle tend à figer l'analyse, en étudiant les comptes panélistés de façon isolée et statique, en ne tenant pas compte de leur intégration dans des réseaux d'interconnaissances.

Au niveau de l'offre, les logiques d'échantillonnage et de panélisation ne sont ainsi pas à même de rendre compte des tensions qui parcourent le champ de production idéologique, qui, comme les autres univers sociaux, est à la fois un champ de force mais aussi un champ de lutte (2.1).

Au niveau de la demande, la logique de panélisation ne permet pas d'étudier les logiques de consommation horizontales, les phénomènes de bouche-à-oreille ou encore les logiques de délégation politique de pair à pair au sein de groupes primaires. Dit autrement, la logique d'échantillonnage isole les individus de leur contexte relationnel d'interconnaissances, contexte réticulaire qui peut être pourtant, sur Twitter, matérialisé et donc analysé (2.2).

2.1 Logique d'échantillonnage et offre d'opinions

Lors de nos précédents chapitres, nous avons étudié l'offre d'opinions politiques de deux manières différentes. Dans notre cinquième chapitre, consacré aux logiques sociales et politiques de consommation des opinions politiques, nous avons analysé l'offre au prisme de la demande, en étudiant les abonnements des individus de nos panels représentatifs aux comptes d'agents et d'institutions appartenant au champ de production idéologique. Dans notre quatrième chapitre, nous avons étudié l'offre *via* le panel « Twitter audience politique ». Nous avons mobilisé ce panel afin d'étudier d'une part la composition et la structure de l'offre sur Twitter, et afin d'analyser d'autre part sa production et la comparer à celle de notre panel représentatif. C'est à cette seconde manière d'étudier l'offre – le panel « Twitter audience politique » – que nous allons ici nous intéresser.

Avant d'aborder les limites de cette méthode de panélisation de l'offre, il convient au préalable de préciser que toutes ses potentialités n'ont pas été forcément exploitées au cours cette recherche. Comme nous l'avons déjà évoqué plus haut, seul le niveau global de production du panel a été analysé. Il pourrait être par la suite intéressant de comparer la volumétrie des tweets concernant cette fois différents enjeux, objets ou idées-forces politiques. Le nombre de tweets publiés sur un enjeu pourrait être de plus ventilé en fonction des différentes variables qui nous ont permis de catégoriser les comptes qui composent le panel « Twitter audience politique » : la nature du compte (média, journaliste, homme politique, etc.), son positionnement politique, son positionnement éditorial (journal de sensation ou de réflexion), la place relative accordée à la politique (journaux omnibus ou journaux d'opinion), etc. Cette ventilation permettrait par exemple de déterminer dans quelle mesure certains arguments en faveur d'un projet de loi sont plus diffusés par certains agents plutôt que par d'autres, etc. Enfin, une des autres potentialités de ce panel serait d'établir une liste des opinions politiques les plus diffusées sur l'ensemble du réseau social. Ce dernier type d'informations est d'ailleurs déjà mobilisé dans certaines études commerciales de *web opinion*, telles que celles produites par l'institut Netscope, sous l'appellation de « top des retweets » (un top 5 ou top 10 des tweets les plus retweetés sur l'ensemble du réseau social, sur une période de temps donnée). Ces données permettent ainsi de connaître les opinions ou les informations les plus diffusées et les plus partagées sur le réseau social.

Si la logique de panélisation de l'offre permet donc bien de générer, du moins potentiellement, des données et des résultats empiriques intéressants, elle se confronte à une

principale limite : une tendance à figer l'analyse et à faire abstraction du contexte dans lequel les agents produisent et publient leurs opinions. Or le champ de production idéologique, auquel appartiennent les membres de ce panel est un champ de forces et un champ de luttes. Il est traversé par des tensions et des logiques de concurrence. Les productions de ces agents sur Twitter (les tweets politiques) doivent donc être analysées au prisme de ces logiques. Or notre dispositif se trouve relativement inadapté à les saisir. Centré sur des unités individuelles, le panel « Twitter audience politique », comme les sondages d'opinions ou les questionnaires à choix fermés, tend à réduire les opinions à des pourcentages. Or, pour reprendre les mots de Bourdieu, « *l'état de l'opinion à un moment donné du temps est un système de forces, de tensions et [...] il n'est rien de plus inadéquat pour représenter l'état de l'opinion qu'un pourcentage* »⁷⁵⁸. Ces logiques de concurrences et de tensions sont pourtant bien présentes sur Twitter.

Ainsi, nous l'avons vu, une grande partie des journalistes français sont inscrits sur le réseau social⁷⁵⁹. Ces journalistes, comme n'importe quels inscrits, ont la possibilité de s'abonner à des comptes d'autres journalistes, qui peuvent eux-mêmes s'abonner en retour aux comptes qui les ont « *followés* ». De manière relativement inédite, Twitter permet donc d'établir une cartographie des relations entre agents à l'intérieur du champ journalistique. Pour être plus précis, Twitter *rend compte* et en même temps *reconfigure* ces relations. Twitter *rend compte* de ces relations lorsque les relations qui se matérialisent à l'écran (*via* les logiques croisées d'abonnement ou de retweet) traduisent des affinités électives qui (pré)existent *réellement* dans le champ, indépendamment de Twitter. *Twitter reconfigure* ces relations lorsque les relations qui se matérialisent à l'écran ont été créées, facilitées ou modifiées par les nouveaux usages, personnels ou professionnels, qui se sont développés avec l'apparition du réseau social.

Ainsi d'après T., un journaliste d'un titre de presse quotidienne régionale, Twitter a facilité la capacité des journalistes, et notamment des nouveaux entrants, à « se faire du réseau », c'est-à-dire à rentrer en contact et à échanger avec d'autres agents du champ avec qui ils n'avaient pas de relation auparavant, à augmenter leur capital social :

[Pourquoi avoir un profil Twitter personnel ? C'est important ?] « *Oui tout à fait c'est important pour tout, d'abord pour se faire du réseau, c'est toujours bien de connaître des*

⁷⁵⁸ Pierre Bourdieu, « L'opinion publique n'existe pas », art. cit.

⁷⁵⁹ Pour une analyse des usages de Twitter par les journalistes, cf. Arnaud Mercier, « Twitter l'actualité : usages et réseautage chez les journalistes français », *Recherches en communication*, 39, 2013.

gens, quand on a besoin d'une info, ou voilà, ça permet de faire son propre réseau et vaut mieux avoir un compte personnel parce que au moins on est vraiment libre d'interpeler qui on veut et dire ce qu'on veut ou à peu près, puis le compte du journal doit vraiment rester la propriété du journal, au moins avoir mon propre compte me permet d'avoir plus de liberté même si moi-même je me fixe des lignes » [Est-ce qu'il vous arrive d'échanger avec d'autres journalistes sur Twitter ?] « *Ah oui bien sûr, tout à fait, on peut discuter entre nous, on peut relayer les infos de l'autre, l'autre peut relayer nos propres infos, on peut s'envoyer des petites blagues, faire de l'humour sur Twitter ensemble, au su et au vu de tout le monde, ça m'arrive fréquemment* » [Cet échange entre journalistes s'est plus développé avec Twitter ou ça existait déjà avant sous d'autres formes ?] « *Non je crois que ça s'est vraiment plus développé avec Twitter, clairement, on arrive plus facilement à rentrer en contact avec des gens qu'on connaît pas, à leur parler comme si on s'était déjà rencontrés alors qu'on s'était jamais rencontrés, sans ça effectivement ça serait beaucoup moins pratique* ». (T., homme, journaliste salarié en CDI, 30 ans, de gauche).

« *Se faire du réseau* », c'est-à-dire accroître son capital social professionnel, peut ainsi répondre à la fois à une fonction utilitariste (prendre contact avec d'autres journalistes pour obtenir des informations, des contacts ou des « off », *etc.*) mais aussi à une fonction stratégique de mobilité dans le champ (se familiariser avec des journalistes et des chefs de services de titres concurrents en vue d'une éventuelle promotion professionnelle, *etc.*).

Augmenter son nombre d'abonnés est un autre exemple de pratique dont la finalité consiste à favoriser une ascension dans le champ. Obtenir un nombre élevé d'abonnés sur Twitter constitue ainsi un nouvel « hochet symbolique ». Avant l'apparition du réseau social à la fin des années 2000, seuls les médias pouvaient être évalués (et hiérarchisés) en fonction de leur audience. Désormais, sur Twitter, les journalistes eux-mêmes bénéficient d'une audience personnelle – leur nombre d'abonnés – distincte de l'audience du média pour lequel ils travaillent. Le nombre d'abonnés est alors perçu par certains de ces agents comme un indice de mesure de leur capital social et, partant, comme une donnée potentiellement valorisable dans le champ médiatique.

Un journaliste qui bénéficie d'un nombre élevé d'abonnés peut être en effet perçu comme en capacité de générer une importante audience – « *de générer du trafic* » pour reprendre les mots du journaliste interrogé – vers les sites internet des médias auxquels il est – ou pourrait être – affilié : « *les sites internet cherchent à élargir leur audience, ils regardent*

qui a beaucoup de followers, parce qu'ils savent que s'ils embauchent cette personne-là cette personne-là va tweeter des informations du site internet donc ça sera visible par autant de gens qui seront abonnés à son compte, donc quand on est dans une période où tous les sites internet cherchent à faire de la page vue, à augmenter leur audience, c'est sûr qu'il vaut mieux avoir des journalistes qui ont beaucoup de followers ça permet d'élargir l'audience ».

Certains agents entreprennent donc, sur Twitter, d'augmenter, par diverses manières, leur nombre d'abonnés. Cette pratique, particulièrement dévalorisée et stigmatisée dans le champ, est ainsi qualifiée de « *personal branding* ». Elle consiste à se démarquer des autres agents, « à se mettre en avant », selon une logique distinctive, en publiant notamment des informations « rares », c'est-à-dire peu relayées par les médias *mainstream* (en réalisant par exemple une veille d'information sur des sites étrangers), ou encore à « *live-tweeter* » certaines émissions télévisées afin de livrer un bon mot ou encore de pratiquer le *fact-checking*⁷⁶⁰. Autant de pratiques visant à augmenter son nombre d'abonnés, dans l'optique de valoriser sa position dans le champ.

[Vous cherchez à augmenter votre nombre de followers, à élargir votre réseau ?] « *C'est quelque chose que je faisais au début évidemment quand on est à 3 followers, on essaye de monter un peu, mais là maintenant non, je fais pas de trucs pour trouver spécialement d'autres followers, je tweete ce que je veux et ce qui m'intéresse, euh donc non c'est quelque chose que je faisais au début, j'essayais d'être très réactif, tout ça, de commenter les émissions en direct, maintenant je suis arrivé à un nombre de followers, 900 et quelques, c'est pas mal, je vais pas non plus faire la compétition pour devenir le premier donc là maintenant non, je cherche plus à essayer de gagner des followers* » [D'autres journalistes sont dans cette compétition du nombre de followers ?] « *Ah oui clairement, tous ceux qui sont sur cet outil qui s'appelle le Klout, qui est censé améliorer votre influence sur internet, y a des gens qui vont passer leur journée à faire ça, qui vont décortiquer tous les sites d'information étrangers pour faire vraiment de l'info fraîche et donc attirer plus de gens, y en a ça se repère et c'est un peu agaçant* »

Les opinions produites par ce type d'agents doivent donc être analysées au prisme des tensions qui structurent le champ médiatique.

⁷⁶⁰ Cf. chapitre 3.

Cependant, la logique de panélisation, en centrant l'analyse sur des individus, des taux et des pourcentages, ne permet pas d'appréhender analytiquement ces mouvements d'opinion, qui sont également des mouvements de position au sein du champ médiatique.

De manière relativement analogue, au niveau cette fois de la « demande », la logique de panel ne permet pas de rendre compte des logiques de consommation horizontale et de l'existence de « marchés secondaires » de l'opinion.

2.2 Logique d'échantillonnage et demande d'opinions

En s'abonnant ou en devenant « amis » avec d'autres utilisateurs, les individus inscrits sur les réseaux sociaux numériques (Twitter, Facebook, *etc.*) se trouvent de fait insérés dans des réseaux interconnectés d'utilisateurs. Des réseaux étudiés, nous l'avons vu, par de nouvelles disciplines universitaires dont les approches se situent à la croisée des théories du marketing viral, de l'influence interpersonnelle, des mathématiques et de l'informatique⁷⁶¹.

Notre dispositif ne prend pas pour objet ces réseaux d'interconnaissances, mais des individus « isolés », tirés au sort, et, pourrait-on dire, « interchangeables ». La logique de panel tend ainsi à dissimuler le contexte réticulaire dans lequel se trouvent insérés les individus inscrits sur Twitter.

Nous l'avons vu, c'est afin de pouvoir qualifier socialement et politiquement les individus inscrits sur les réseaux sociaux que nous avons choisi d'étudier un échantillon raisonné, limité à environ 1200 individus. Qualifier de la même manière, c'est-à-dire « manuellement » (*via* l'administration d'un questionnaire ou *via* l'observation directe des descriptions de profil) les « abonnés » ou les « abonnements » des 1228 membres de nos panels représentatifs serait techniquement et humainement irréalisable. À eux seuls, les 608 individus qui ont répondu à notre questionnaire sont ainsi abonnés à plus de 100 000 comptes différents.

Contrairement à notre recherche, les travaux issus de la nouvelle science des réseaux⁷⁶² ne cherchent pas à dépasser cette contrainte de qualification manuelle. Ces

⁷⁶¹ Cf. chapitre 1.

⁷⁶² Voir notamment ; Kévin Mellet, « Aux sources du marketing viral », *Réseaux*, 2009/5-6, n° 157-158 ainsi que d'autres références recensées dans notre premier chapitre.

recherches étudient en effet les structures des réseaux (et les logiques de circulation de contenus numériques en leur sein) sans tenir compte des caractéristiques sociologiques et politiques de leurs membres. Pourtant, nombre de ces travaux invoquent les théories de l'influence interpersonnelle et des leaders d'opinion. Pourtant, les travaux de l'école de Columbia accordent une attention primordiale aux caractéristiques sociales et politiques des individus et des groupes primaires au sein desquels s'exercent ces logiques d'influence⁷⁶³. Il ne suffit pas de constater qu'un individu *x* est retweeté par un individu *y* pour en déduire une relation d'influence, ou de délégation politique, pour revenir à un vocabulaire plus proche de Bourdieu⁷⁶⁴. Ce n'est qu'en étudiant précisément la position sociale et le niveau de compétence – technique et statutaire – de ces deux individus que l'objet étudié – le retweet – peut éventuellement éclairer le chercheur sur la relation de délégation qui peut exister entre ces deux individus.

Notre méthode de panélisation – pas plus d'ailleurs que les méthodes prenant pour objet des réseaux macroscopiques de comptes non-qualifiés socialement – ne permet pas d'analyser ces logiques de délégation politique au sein de groupes primaires, ou, si l'on veut rester dans l'univers sémantique du marché, ces logiques de consommation horizontale d'opinions.

Dans notre chapitre cinq, nous proposons en effet d'appréhender la pratique du retweet comme un acte de consommation. Nous nous focalisons alors sur les messages publiés originellement par des comptes médiatiques et politiques et retweetés par des agents situés à l'extérieur de ces champs. Cependant, seul un retweet politique sur deux provient d'un compte appartenant au champ de production idéologique. Pour le dire autrement, 50 % des retweets politiques publiés par les membres de notre panel « répondants » ne sont pas des messages originellement publiés par des journalistes ou des médias, mais des messages publiés par des individus situés à l'extérieur de ces champs ; des messages publiés par des individus appartenant aux réseaux d'interrelations – réelles ou virtuelles – des membres de notre panel représentatif.

Parmi ces retweets « de pair à pair », deux catégories de messages peuvent être identifiées : des messages publiés originellement par un pair et des messages publiés

⁷⁶³ Elihu Katz et Paul Lazarsfeld, *Influence personnelle*, Paris, Arman Colin, 2008.

⁷⁶⁴ Daniel Gaxie nous rappelle une certaine analogie entre les concepts d'influence et de délégation politique. Cf. Daniel Gaxie, « Retour sur les modes de production des opinions politiques », in Philippe Coulangeon et Julien Duval (dir.) *Trente ans après La Distinction de Pierre Bourdieu*. Paris, La Découverte, 2013.

originellement par des médias puis retweetés par un pair. Dans le premier cas de figure, l'individu x retweete un message écrit et publié en première personne par l'individu y . Il s'agit donc d'une consommation localisée de pair à pair. Dans le second cas de figure, l'individu x retweete un tweet publié par un individu y qui est un retweet d'un message publié originellement par un compte de média. L'individu x consomme sur un marché secondaire (logique horizontale) un produit déjà consommé par l'individu y sur le marché premier (logique verticale).

Un enjeu majeur de la recherche sur les réseaux sociaux sera alors d'analyser ces logiques de consommation horizontale de pair à pair au prisme des caractéristiques sociologiques et politiques des consommateurs (et ne pas se contenter comme aujourd'hui de cartographier et de décrire ces réseaux en faisant abstraction de leur ancrage dans le monde social « réel »). Les individus sur Twitter retweetent-ils davantage des messages publiés par des individus qu'ils connaissent réellement ou au contraire par des individus qu'ils ne connaissent que *via* Twitter ? Discutent-ils de politique avec des individus qui partagent les mêmes opinions politiques que les leurs ? Ou l'homophilie politique et sociale est-elle au contraire moins marquée sur internet ? Autant de questions auxquelles notre dispositif méthodologique ne peut répondre, faute de pouvoir sociologiser un nombre suffisamment important d'individus.

On le voit donc bien, notre méthode de panélisation des réseaux sociaux, qu'elle soit appliquée à l'offre ou à la demande, comporte certaines limites. Elle tend ainsi à figer l'analyse en étudiant des individus isolés de leur contexte et de leurs réseaux d'interconnaissances (virtuelles ou réelles).

En questionnant les limites de notre méthode, nous nous trouvons donc de nouveau face au « dilemme » évoqué lors des premiers chapitres de cette recherche. Face aux données numériques, le chercheur a deux possibilités : analyser de façon exhaustive la population étudiée ou l'échantillonner. S'il choisit d'analyser la population dans sa totalité, il se trouve dans l'impossibilité de situer socialement et politiquement les individus qu'il étudie, contraint alors d'« abandonner la référence sociodémographique »⁷⁶⁵. S'il choisit de l'échantillonner,

⁷⁶⁵ Dominique Boullier et Audrey Lohard, *Opinion Mining et Sentiment analysis, op. cit.*, p. 189.

de nombreux phénomènes peuvent alors échapper à l'analyse, comme par exemple les logiques de consommation horizontale des opinions que nous venons d'évoquer.

Nous proposons maintenant d'esquisser quelques pistes de réflexion visant à résoudre cette tension entre exhaustivité et qualification manuelle.

3. Des dispositifs méthodologiques alternatifs de mesure des opinions numériques

Nous évoquerons dans cette dernière section quelques dispositifs méthodologiques susceptibles de pallier certaines limites de notre méthode. Nous présenterons d'abord des outils de cartographie des réseaux socionumériques et questionnerons la manière dont ils pourraient nous permettre de matérialiser et d'analyser les logiques de consommation « horizontale » des opinions politiques sur Twitter (3.1). Nous esquisserons ensuite quelques pistes de réflexion visant à qualifier sociologiquement un plus grand nombre d'individus, ce afin de pouvoir resituer ces réseaux socionumériques dans l'espace social « réel » (3.2).

3.1 Matérialiser et analyser les logiques de consommation horizontale des opinions

Si l'arrière-plan théorique des recherches qui les mobilisent d'ordinaire est clairement différent du nôtre⁷⁶⁶, les outils cartographiques des réseaux socionumériques peuvent néanmoins nous permettre d'analyser certains phénomènes qui échappent à notre dispositif méthodologique et qui sont pourtant centraux dans l'analyse de la production et de la consommation des opinions politiques. Ces outils permettent en effet de cartographier le champ de production idéologique, de rendre compte, de façon dynamique – avec des graphiques animés – des tensions qui le parcourent, mais aussi de matérialiser les logiques de consommation horizontale des opinions politiques. C'est à ce second point que nous allons ici nous intéresser.

⁷⁶⁶ Cf. chapitre 1.

Sur Twitter, la cartographie des réseaux peut être réalisée à partir de deux principales variables : les abonnements des inscrits (leurs liens affinitaires) ou leurs retweets (leur consommation d'opinions et d'informations).

Dans le premier cas de figure, les comptes – d'individus ou d'institutions – sont représentés graphiquement par des nœuds et les abonnements par des arêtes. Chaque compte est ainsi défini par un certain nombre de liens entrants (les abonnés) et de liens sortants (les abonnements). On peut alors représenter graphiquement les liens affinitaires qui unissent les différents individus qui composent une population donnée à un moment t .

Dans le second cas de figure, les comptes sont représentés par des nœuds et les retweets par des arêtes. Dans cette configuration, chaque compte (x) est également défini par des liens entrants (les comptes $n...$ à l'origine des retweets publiés par le compte x) et des liens sortants (les tweets publiés par le compte x puis retweetés par les comptes $n...$). Le diamètre des nœuds peut soit correspondre au nombre de tweets publiés par le compte, soit au nombre de retweets. Ce sont alors les logiques de circulation des opinions au sein des marchés secondaires et les logiques de consommation de pair à pair qui peuvent être représentées graphiquement.

Dans cette seconde configuration, le temps devient alors une variable intégrée au modèle. La cartographie prend ainsi la forme d'un graphique animé où se reconstituent, à la manière d'un film, les dynamiques de circulation et de consommation des messages. La séquence temporelle étudiée (de x jours ou de x mois) peut être condensée en une plage de temps réduite (de x secondes ou de x minutes). Selon l'échelle temporelle adoptée, le chercheur peut alors représenter graphiquement des dynamiques de courte, moyenne ou longue durée.

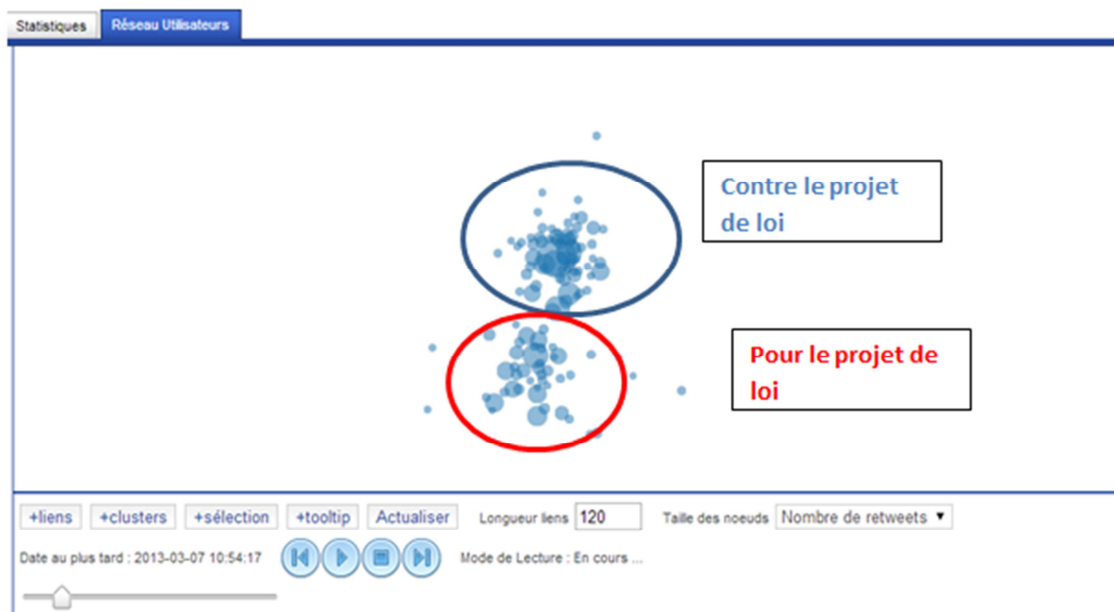
À titre d'exemple, nous avons représenté graphiquement, à l'aide de notre logiciel *Ami Opinion Tracker*⁷⁶⁷, les dynamiques de retweet concernant les messages abordant le projet de loi relatif au mariage homosexuel en France, sur une période allant du 7 janvier 2013 (soit 6 jours avant la première grande manifestation d'opposition) au 22 avril 2013 (veille de l'adoption définitive du texte par l'Assemblée Nationale). Précisons que cette analyse cartographique n'est pas basée sur la production d'un panel particulier, mais sur l'ensemble des tweets publiés sur le réseau social contenant des occurrences relatives au projet de loi

⁷⁶⁷ *Ami Opinion Tracker* intègre depuis 2012 une application de cartographie de Twitter, développée par sa division recherche et développement, cf. Annexe 6, p. 487.

(« #mariagepourtous », « mariage gay », « mariage homo », etc.). Le corpus étudié est donc composé de 896 903 tweets provenant de 291 891 auteurs uniques.

La cartographie 6.1 représente l'ensemble des comptes Twitter ayant retweeté – ou ayant publié un tweet ayant été retweeté – sur le mariage homosexuel, à un moment t_n , le 7 mars 2013, soit deux mois après le premier jour de notre période d'étude.

Illustration 6.1 Cartographie des retweets concernant le mariage homosexuel le 7 mars 2013



Source : *Ami Opinion Tracker*, 2014

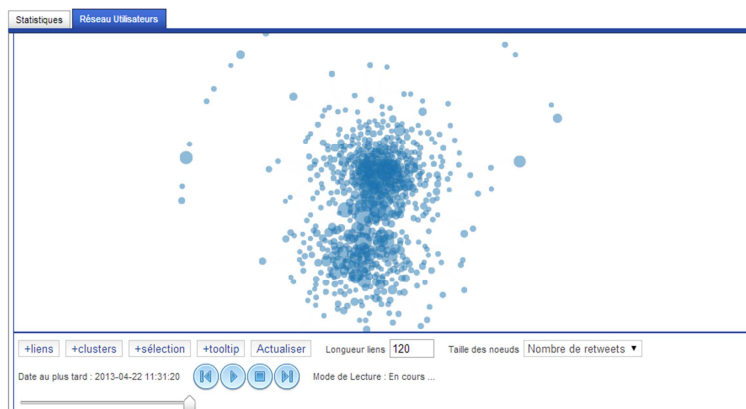
Trois résultats peuvent être rapidement tirés de l'analyse de cette cartographie. Le premier concerne le nombre de nœuds, relativement faible. Peu d'événements relatifs au projet de loi ont marqué la journée étudiée, le 7 mars 2013. Peu de comptes ont alors produit et/ou consommé des opinions et des informations concernant ce sujet. Le deuxième résultat concerne la taille des nœuds, quant à elle relativement élevée. Les comptes qui ont consommé et/ou produit des messages ce jour-là sont de « gros » consommateurs/producteurs : ils ont publié un nombre important de messages à ce sujet. Il s'agit donc des comptes les plus mobilisés (qu'ils soient contre ou en faveur du projet de loi) : souvent des militants, des associations, des médias « engagés ». Enfin, le troisième résultat concerne la séparation

relativement nette entre deux groupes de comptes. Au « nord » de la carte, on voit se former un premier groupe de nœuds agglomérés, relativement homogène, éloigné d'un second groupe, situé « au sud » de la carte. En plaçant le curseur de la souris sur les nœuds de cette carte, l'utilisateur du logiciel peut d'une part connaître le sens de la relation entre deux nœuds (« qui a retweeté qui ? »), et, dans certains cas⁷⁶⁸, des informations sur le compte lui-même (son nom, sa description de profil, son nombre de tweets, d'abonnés et d'abonnements). En « glissant » la souris sur les nœuds de chacun des deux groupes, on peut alors rapidement se rendre compte que le premier groupe, situé au nord, est composé de comptes hostiles au projet de loi (avec principalement des comptes issus du collectif de la « manif pour tous »), tandis que celui situé au sud de la carte est au contraire composé de comptes favorables au projet de loi (principalement des militants socialistes et des associations et militants LGBT). Cette cartographie rend donc manifeste les logiques de consommation politiquement sélectives des individus inscrits sur Twitter. Les comptes hostiles au projet de loi ont ainsi très majoritairement retweeté des messages publiés par des comptes eux-mêmes hostiles au projet de loi et inversement, les comptes favorables ont retweeté des messages publiés par des comptes favorables. Il s'agit donc bien là d'une autre façon de démontrer empiriquement les logiques de consommation socialement et politiquement déterminées des opinions sur le réseau social.

La cartographie 6.2 représente l'état de l'opinion sur ce même sujet à un moment t_{n+1} , à la fin de notre période d'étude, le 22 avril 2013, soit la veille de l'adoption définitive du projet de loi par l'Assemblée Nationale. À la différence de la période étudiée dans la cartographie 6.1, cette journée est donc marquée par une forte actualité concernant le projet de loi. On peut ainsi voir que le nombre de nœuds est considérablement plus important que le 7 mars 2013. Les consommateurs et producteurs d'informations et d'opinions relatives à ce débat ne sont plus seulement des agents ou des institutions politiquement mobilisés, mais également des individus non-militants. Cependant, la séparation entre les deux groupes de nœuds est toujours nette. On remarque alors que même les individus moins engagés dans ce débat consomment des opinions produites par des comptes qui partagent la même opinion que la leur.

⁷⁶⁸ Les créateurs du logiciel ont entrepris de capitaliser « au fil de l'eau », par strates successives, une base de données recensant les informations des comptes Twitter répertoriés par le logiciel.

Illustration 6.2 Cartographie des retweets concernant le mariage homosexuel le 22 avril 2013



Source : *Ami Opinion Tracker*, 2014

Ces outils de cartographie permettent donc de représenter graphiquement ces dynamiques d'opinion. En revanche, nous ne pouvons rendre compte « sur le papier » de l'ensemble des séquences qui s'enchaînent à l'écran de manière dynamique et continue de t_0 à t_n . L'animation graphique de ces dynamiques de retweet rend alors manifestes les logiques temporelles de diffusion, de structuration et de mobilisation des idées-forces politiques par et autour de groupes mobilisés.

Si le logiciel fournit des informations sur les comptes cartographiés (les données « utilisateurs » – du type nom et description de profil, nombre de tweets, d'abonnements et d'abonnés – accessibles à n'importe quel utilisateur sur Twitter), il ne permet pas en revanche de catégoriser automatiquement les caractéristiques sociologiques (PCS, âge, niveau d'études, etc.) et politiques (niveau de compétence politique, orientation politique, engagement militant, etc.) de ces comptes, même lorsque celles-ci sont « auto-déclarées » par les utilisateurs eux-mêmes dans leur description de profil. Le logiciel scanne l'intégralité de la description de profil, sans toutefois pouvoir y repérer automatiquement les termes témoignant par exemple d'une orientation politique.

On peut cependant raisonnablement penser qu'il s'agit là d'une limite technique qui sera bientôt résolue. Nous avons d'ailleurs déjà eu l'occasion de tester un logiciel concurrent, *Visibrain*, utilisé par le *Service d'information du Gouvernement* (SIG). Ce logiciel permet déjà de catégoriser automatiquement des données sociologiques et politiques à partir des

descriptions de profil des utilisateurs. En revanche, nous l'avons vu, une majorité de comptes Twitter n'indique pas dans sa description de profil ces informations socio-démographiques et politiques. Pourtant, qualifier finement la position sociale et politique de l'ensemble des individus qui forment ces réseaux socionumériques est la condition *sine qua non* d'une importation de ces outils de cartographie en sociologie politique.

Afin de qualifier en plus grand nombre les propriétés sociologiques des individus inscrits sur Twitter, et resituer ainsi ces *réseaux numériques* dans *l'espace social réel*, quelques pistes de réflexion peuvent être ici esquissées.

3.2 Des modalités alternatives de sociologisation des comptes

Si les outils de représentation graphique des dynamiques de production et de consommation des opinions sur Twitter peuvent apporter une réelle plus-value heuristique à l'analyse sociologique des opinions, encore faut-il, rappelons-le encore une fois, être en mesure de situer socialement et politiquement ces producteurs et ces consommateurs. Or, à ce problème, les logiciels de veille et de *tracking* ne proposent pas (encore ?) de solution toute faite.

L'administration d'un questionnaire à une population composée de centaines de milliers d'individus est très difficilement envisageable. C'est pour cette raison que nous avons finalement choisi d'étudier des échantillons raisonnés. Cependant, cette logique d'échantillonnage est insuffisante dès lors que l'on entreprend de cartographier les dynamiques de circulation et de consommation des opinions à l'échelle du réseau social dans son ensemble.

Nous allons alors évoquer ici trois pistes de réflexion visant à « sociologiser » un plus grand nombre de comptes, sans avoir recours à des protocoles trop coûteux. L'administration d'un questionnaire en boule-de-neige peut être ainsi une première piste. Déléguer la tâche de passation du questionnaire aux enquêtés permettrait de diminuer sensiblement le coût de l'opération (3.2.1). Une seconde piste consisterait à recourir à une analyse automatisée des caractéristiques sociologiques des individus à partir d'une analyse de leurs abonnements. Cette option relève plus cependant de la prospective méthodologique, et nécessiterait une collaboration étroite entre informaticiens et sociologues (3.2.2). Enfin, une dernière manière

d'augmenter le nombre d'internautes « sociologisés » pourrait être de mutualiser et de capitaliser les bases de données de la communauté scientifique travaillant sur ces questions et ces terrains. Si toutes les équipes de recherche décidaient de mettre en commun leurs bases de données, on pourrait alors accroître considérablement la connaissance de ces populations internautes (3.2.3).

3.2.1 Déléguer aux enquêtés la tâche de passation du questionnaire

La technique de l'échantillonnage en boule de neige – ou *snowball sampling* – a été popularisée par Paul Lazarsfeld et Elihu Katz dans *Personal Influence*⁷⁶⁹. Ce dispositif consistait à demander à un échantillon réduit d'enquêtés de désigner les personnes de leur entourage qu'ils jugeaient exercer sur eux une influence dans différents domaines. Les personnes désignées comme influentes étaient ensuite interrogées à leur tour par les enquêteurs. Ce protocole devait ainsi permettre aux chercheurs d'identifier des leaders d'opinion et des chaînes d'influence interpersonnelle.

Julien Audemard, dans sa recherche doctorale⁷⁷⁰, a mobilisé cette même méthode en lui apportant deux modifications majeures. D'une part, les enquêtés-souches sont dans son enquête des membres « intimes » de son entourage immédiat : sa famille, ses amis. D'autre part, c'est à ses enquêtés qu'il confie la tâche de passation du questionnaire. Les enquêtés-souches sont ainsi chargés d'administrer eux-mêmes à leur entourage immédiat le questionnaire auquel ils ont préalablement répondu.

Appliqué aux réseaux socionumériques, ce dispositif permettrait de générer un nombre très important de données, tout en rendant identifiables des réseaux sociaux « réels », en retraçant les chaînes de passation du questionnaire. On pourrait ainsi envisager de demander à un échantillon initial d'inscrits sur Twitter (tirés au sort, comme lors de notre propre enquête) de 1) répondre à notre questionnaire, 2) diffuser le questionnaire à leurs abonnés, 3) adresser à leurs abonnés les mêmes consignes que celles que nous leur avons livrées. Afin d'être rempli très rapidement, le questionnaire, identique au nôtre, comporterait seulement une vingtaine de questions socio-démographiques et de préférences politiques. Le questionnaire étant hébergé

⁷⁶⁹ Elihu Katz et Paul Lazarsfeld, *Influence personnelle*, *op. cit.*

⁷⁷⁰ Julien Audemard, *Influences interpersonnelles. Comment le contexte structure les opinions et les votes*, Thèse de doctorat soutenue à Montpellier le 4 décembre 2013.

et auto-administré en ligne, ce dispositif générerait de plus très peu de problèmes « logistiques ». Il suffirait aux enquêtés de demander à leurs abonnés de retweeter un message les invitant à répondre à notre questionnaire, suivi du lien hypertexte vers notre questionnaire (il peut être intéressant de signaler que lors de l'administration de notre propre questionnaire, des enquêtés avaient spontanément entrepris de diffuser le questionnaire à leurs abonnés⁷⁷¹). Cette absence de difficultés logistiques autorise alors un nombre très important de « cascades » de passation (les abonnés des abonnés des abonnés, *etc.*).

Ce dispositif, en plus de générer un nombre très élevé de questionnaires remplis, permettrait également d'éclairer sous un jour nouveau les réseaux socionumériques : identifier les logiques de passation du questionnaire, les propriétés des individus à qui les enquêtés choisissent de l'administrer, celles des individus qui acceptent – ou au contraire refusent – de répondre, le nombre de répondants par cascade, le nombre de dyades identiques d'une cascade à l'autre, *etc.* Ces données permettraient de plus de vérifier empiriquement certaines hypothèses sur les liens faibles ou l'existence des petits mondes⁷⁷², *etc.*

Il est donc inutile de préciser que les résultats issus d'un tel dispositif suffiraient à alimenter une recherche doctorale à part entière. Nous nous bornerons simplement ici à envisager rapidement deux principales limites. La première est que ce dispositif ne garantit pas au chercheur un taux de réponse élevé. Avec ce dispositif, comme avec notre propre méthode, la sociologisation des internautes n'est pas « systématique » : tous ne sont pas susceptibles de répondre au questionnaire. De plus, en déléguant aux enquêtés le rôle d'enquêteur, le chercheur perd en même temps la maîtrise du processus de passation et de diffusion du questionnaire. À partir de la troisième cascade, il devient très vite difficile pour l'enquêteur de vérifier le bon déroulement de la passation, de s'assurer que les différentes strates d'enquêtés-enquêteurs administrent correctement le questionnaire, *etc.* On peut alors craindre que plus le questionnaire s'« éloigne » du chercheur et plus le public à qui il est diffusé se segmente et se réduise aux individus les plus enclins à répondre à une enquête politique, c'est-à-dire aux plus gros producteurs de tweets politiques.

⁷⁷¹ Nous n'avons pas tenu compte des réponses de ces abonnés, car ils ne figuraient pas dans notre échantillon aléatoire.

⁷⁷² Stanley Milgram et Jeffrey Travers, « An Experimental Study of the Small World Problem », *Sociometry*, 1969, Vol.32, No. 4. (1), p. 425-443 ; Mark S. Granovetter, « The Strength of Weak Ties », *American Journal of Sociology*, Vol. 78, n°6, 1973, p. 1360-1380.

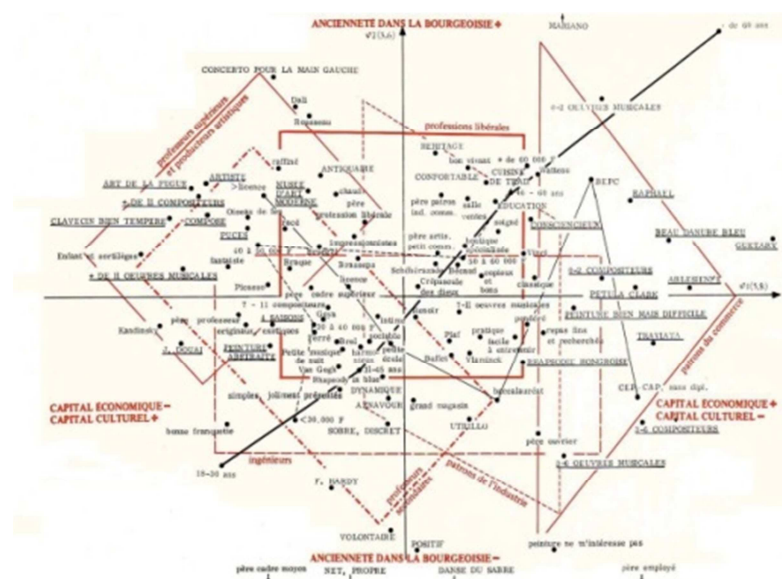
3.2.2 Automatiser l'analyse des goûts

Un autre dispositif de caractérisation sociologique des membres des réseaux socionumériques, qui relève plus, en 2014, de la prospective méthodologique, consisterait à identifier automatiquement la position sociale et politique des internautes à partir de leurs goûts tels qu'ils se manifestent dans leurs choix d'abonnements.

Nous avons démontré lors de notre cinquième chapitre que les choix de nos enquêtés en matière d'abonnements ne se distribuaient pas au hasard, mais étaient bien déterminés socialement. Si nous avons focalisé l'analyse sur les abonnements aux comptes de médias, de journalistes et d'hommes politiques, ce même constat peut être fait concernant les abonnements à des comptes artistiques (chanteurs, acteurs, humoristes, *etc.*). Selon qu'ils se situent à un point ou à un autre de l'espace social, les agents sociaux n'« aiment » pas les mêmes musiques, les mêmes films, les mêmes sports ou les mêmes photographies.

À l'aide d'une analyse de correspondances multiples (ACM), Pierre Bourdieu, dans *La Distinction*, avait alors représenté graphiquement, sur un même espace, les propriétés sociologiques des agents sociaux et leurs préférences culturelles.

Cartographie 6.3 « Variantes du goût petit-bourgeois » (graphiques 15 et 16 de La Distinction, p.392)



Une telle cartographie pourrait être réalisée à partir des propriétés sociologiques et politiques des individus inscrits sur Twitter et de leurs préférences culturelles, telles qu'elles se manifestent dans leurs choix d'abonnements à des comptes artistiques, médiatiques ou politiques.

Une fois cette cartographie réalisée, il suffirait alors de connaître les abonnements d'un individu donné pour le situer sur cette carte, c'est-à-dire pour le situer dans l'espace social, et ainsi déduire ses caractéristiques sociales (âge, sexe, PCS, niveau de diplôme, etc.) et ses orientations politiques. Dans cette optique, la représentation graphique de l'espace social ne serait plus le produit d'une méthode (l'ACM), mais la méthode elle-même. La position sociale d'un individu serait alors appréhendée comme étant le résultat d'une équation à plusieurs inconnues, et ses préférences culturelles, telles qu'elles se manifestent dans ses choix en matière d'abonnement, comme les différentes inconnues de l'équation. Connaître ces inconnues permettrait de résoudre l'équation. Connaître les goûts culturels permettrait de déduire la position sociale d'un individu.

Nous l'avons dit, un tel dispositif n'existe pas encore à notre connaissance et relève plus à ce jour de la prospective méthodologique. Sa création nécessiterait une collaboration très étroite entre sociologues et informaticiens. Une collaboration qui est, à l'heure actuelle, très embryonnaire dans le champ académique⁷⁷³. On peut de plus supposer qu'un tel dispositif comporterait une inévitable marge d'erreur. Il n'existe pas en effet de déterminisme pur et parfait dans le monde social. Dès lors que certains individus, aussi minoritaires soient-ils, continuent d'échapper aux déterminismes sociaux, le risque d'erreur serait nécessairement inhérent à une telle méthode.

3.2.3 Mutualiser et capitaliser les bases de données sur les réseaux socionumériques

Enfin, une dernière solution, dans l'immédiat sans doute plus envisageable, serait de mutualiser et de capitaliser les bases de données des recherches en sciences humaines et sociales portant sur ces terrains et ces populations « numériques ».

L'idée de mutualiser et de capitaliser les bases de données en sciences humaines et sociales n'est pas nouvelle. En 2003, Françoise Cribier, directrice de recherche au Centre

⁷⁷³ Dominique Boullier Audrey Lohard, *Opinion mining et sentiment analysis, op. cit.*

Maurice Halbwachs, remettait au Ministère délégué à la Recherche et aux nouvelles technologies un « rapport sur la sauvegarde des données qualitatives des sciences sociales »⁷⁷⁴. Le but du projet consistait à « *sauvegarder, inventorier et rendre accessibles à la communauté scientifique l'essentiel des données qualitatives de qualité collectées depuis 30 ou 40 ans auprès des gens ordinaires par les SHS, et qui n'ont pas encore disparu* ». Et l'auteur de préciser que « *la sauvegarde que nous proposons ne serait pas seulement une sauvegarde à "visée conservatoire", même si son intérêt patrimonial est évident. Elle aurait pour finalité, comme le Centre Quételet, et avec l'accord de leurs auteurs, la mise à disposition et la réutilisation des données* »⁷⁷⁵.

Ce même projet de mutualisation pourrait être mis en œuvre concernant les recherches quantitatives sur les populations et les terrains numériques. Dans la démarche qui est la nôtre, s'en tenir uniquement aux questions socio-démographiques permettrait sans doute de ne pas tomber dans « *l'irréalité abstraite de l' "analyse secondaire" des données disparates que thésaurisent mécaniquement les "banques de données" sans capital théorique* »⁷⁷⁶. En effet dans notre volonté de situer socialement les membres des réseaux socionumériques, seules les questions permettant de définir les propriétés sociologiques des agents sociaux nous intéressent⁷⁷⁷. Cette mutualisation permettrait alors d'augmenter considérablement la connaissance sociologique des réseaux socionumériques, en permettant de situer ces réseaux *virtuels* dans l'espace social *réel*.

Ces quelques pistes de réflexion ont un objectif commun : permettre aux chercheurs de situer socialement et politiquement la population des inscrits aux réseaux socionumériques sans avoir recours à une logique d'échantillonnage. Cette logique d'échantillonnage, au cœur de notre dispositif méthodologique, ne permet pas, nous l'avons démontré, d'appréhender les logiques de concurrence et de tension au sein du champ de production idéologique d'une part, et les logiques horizontales de circulation et de consommation des opinions d'autre part.

⁷⁷⁴ Françoise Cribier, « Rapport sur la sauvegarde des données qualitatives des sciences sociales », Avril 2003.

⁷⁷⁵ *Ibidem*.

⁷⁷⁶ Pierre Bourdieu, *La Distinction, op. cit.*, p. 592.

⁷⁷⁷ Une telle démarche devrait scrupuleusement respecter les conditions de confidentialité et de respect de la vie privée des internautes.

Conclusion

À travers cette conclusion, nous avons cherché à expliciter les apports théoriques mais aussi les limites heuristiques de notre dispositif méthodologique de panélisation des réseaux sociaux.

Nous avons d'abord démontré que la logique d'échantillonnage du réseau social, en rendant possible une caractérisation fine des propriétés sociologiques et politiques des inscrits sur Twitter, a permis de collecter des données et des résultats empiriques qui n'auraient pas pu être obtenus en analysant simplement la production de l'ensemble du réseau social.

Ces données confirment des résultats bien établis de la sociologie politique, comme l'inégale propension des agents sociaux à prendre la parole politique selon leur position sociale ou leur niveau de compétence politique. Mais ces données ont également permis de vérifier empiriquement des hypothèses qui ne l'avaient que peu été jusqu'alors, notamment l'inégale distribution dans le temps des discussions politiques ou encore une détermination du niveau de la demande d'opinions par le niveau de l'offre.

Nous avons ensuite mis en avant quelques limites de notre dispositif. Nous avons ainsi démontré qu'en échantillonnant aléatoirement la population des inscrits au réseau social, notre dispositif tendait à isoler les individus du contexte réticulaire dans lequel ils étaient insérés. Au niveau de l'offre, la logique d'échantillonnage ne permet pas de rendre compte des tensions qui parcourent le champ de production idéologique et qui déterminent la production du champ. Au niveau de la demande, l'échantillonnage ne permet pas d'analyser les logiques de consommation horizontale des opinions et des informations politiques.

Nous avons enfin esquissé quelques pistes de réflexion visant à pallier ces limites. Avoir recours aux outils de cartographie des réseaux sociaux pourrait ainsi apporter une certaine plus-value heuristique à l'analyse sociologique des réseaux siconomériques, en permettant précisément de matérialiser les dynamiques de production, de circulation et de consommation des opinions et des informations politiques sur Twitter. Mais ces outils ne pourraient être opérants qu'à la condition de pouvoir caractériser socialement en plus grand nombre les consommateurs de ces opinions.

En résumé, notre dispositif méthodologique n'est parvenu à résoudre que très partiellement cette tension constitutive des recherches numériques entre d'un côté la plus-value d'une *analyse exhaustive* des autopublications numériques et de l'autre la nécessité de

qualifier sociologiquement les producteurs de ces messages. La résolution de cette tension constitue sans doute l'un des nombreux leviers à l'importation de ces terrains, matériaux, outils et méthodes numériques en sociologie et sociologie politique.

Table des matières

Remerciements	5
Sigles et abréviations	8
INTRODUCTION GÉNÉRALE : LES CONDITIONS MÉTHODOLOGIQUES D'UNE ANALYSE SOCIOLOGIQUE DES OPINIONS NUMÉRIQUES.....	13
Une tentative d'importation d'outils de <i>web-tracking</i> en sociologie politique.....	14
Des perspectives de recherche stimulantes pour la sociologie politique	15
Des données empiriques générées en dehors de toute interaction de recherche.....	15
Les avantages méthodologiques des logiciels de web-tracking	16
Les limites des logiciels de web-tracking.....	21
Une agrégation de données hétéroclites	22
Une analyse centrée sur les flux de messages plutôt que sur les individus	23
Une impossible caractérisation socio-démographique des producteurs d'autopublications numériques	23
À partir de quel cadre théorique et selon quelle méthodologie analyser les données collectées par les logiciels de <i>web-tracking</i> ?.....	25
Le cadre délibératif habermassien et les logiques d'influence interpersonnelle de Lazarsfeld comme principaux horizons théoriques des recherches sur les autopublications numériques ..	26
Le cadre théorique habermassien de la délibération.....	26
Les phénomènes de « viralité » et les théories de l'influence interpersonnelle.....	29
Les théories de Pierre Bourdieu : un cadre opérant d'analyse des opinions numériques.....	32
Désarticuler la question des opinions de celle des sondages.....	33
Concilier dimension conjoncturelle et dispositionnelle de production des opinions politiques	35
Notre méthodologie : panéliser un échantillon représentatif d'internautes inscrits sur le réseau social Twitter	38
Notre méthode de panélisation du réseau Twitter permet-elle de rendre compte de l'existence d'un marché de l'opinion ?	42
Twitter, un observatoire de l'« offre » d'opinions.....	43
Twitter, un observatoire de la « demande » d'opinions	44
Plan de thèse.....	45
PREMIÈRE PARTIE : LES AUTOPUBLICATIONS NUMÉRIQUES COMME SOURCE DE RENOUVELLEMENT THÉORIQUE ET MÉTHODOLOGIQUE DE L'ANALYSE DES OPINIONS POLITIQUES	48

CHAPITRE 1 - LES USAGES COMMERCIAUX ET SCIENTIFIQUES DES AUTOPUBLICATIONS NUMÉRIQUES.....	50
1. L'autopublication numérique, une nouvelle pratique sociale	52
1.1 Les facteurs de démocratisation	52
2. La diversification de la pratique sociale d'autopublication.....	56
2. Les spécificités heuristiques d'un nouveau matériau d'étude du monde social.....	58
2.1. Un matériau produit hors laboratoire.....	59
2.1.1. Des verbatims non altérés par la situation d'enquête	59
2.1.2 Des données de seconde main décontextualisées et désincarnées.....	62
2.1.3 Les stratégies de (non) qualification sociodémographique des auteurs des autopublications numériques	64
2.2. Un matériau adapté à des méthodes combinant approches quantitatives et qualitatives....	67
2.2.1 Les autopublications au prisme du dualisme quantitatif/qualitatif	68
2.2.2 Les instruments de mesure des autopublications numériques	71
2.3 L'objectivation et la modélisation de la circulation temporelle et spatiale des autopublications.....	76
2.3.1 Un nouveau rapport à la temporalité	77
2.3.2. Les représentations graphiques des dynamiques de messages	81
3. Un nouveau champ de l'opinion ?.....	85
3.1. L'intérêt différencié des acteurs scientifiques et commerciaux	86
3.1.1. Un matériau commun aux études scientifiques et commerciales	86
3.1.2. La création d'un nouveau secteur d'activité économique marqué par l'arrivée de nouveaux acteurs, concurrents des instituts de sondage.....	88
3.1.3. Désintérêt de la sociologie politique et informatisation des sciences sociales	96
3.2. Les trois grandes traditions d'étude des autopublications numériques	100
3.2.1. Les autopublications numériques comme indices de participation politique	101
3.2.2 Autopublications, « viralité », réseaux sociaux et influence	107
3.2.2.1. Les origines du « marketing viral »	107
3.2.2.2 Les trois approches de la contagion.....	113
3.2.3 Opinion mining, sentiment analysis et prédictivité de Twitter.....	119
3.2.3.1. Les enjeux scientifiques de l' <i>opinion mining</i> et du <i>sentiment analysis</i>	119
3.2.3.2 Les analyses prédictives de <i>sentiment analysis</i>	123
CHAPITRE 2 - PANÉLISER TWITTER, UNE MÉTHODE HYBRIDE DE MESURE DES OPINIONS.....	130
1. Représenter, définir et mesurer les opinions en sciences sociales	133
1.1 Les définitions et dispositifs d'étude de l'opinion antérieurs aux sondages	134

1.1.1 L'opinion publique comme objet historiographique	135
1.1.2 L'opinion comme catégorie de la philosophie politique avant les sondages.....	137
1.1.3 Les dispositifs expérimentaux de mesure de l'opinion avant les sondages	139
1.2 L'avènement des sondages : quantification, représentativité et dépolitisation de l'opinion	142
1.2.1 La transformation de l'opinion en « construit mesurable »	143
1.2.2 Échantillonnage, représentativité et prédictivité de l'opinion	144
1.2.3 Une définition scientifique et politique de l'opinion publique.....	146
1.3 La critique universitaire des sondages et de l'opinion publique.....	147
1.3.1 L'opinion publique (mesurée par les sondages) n'existe pas	147
1.3.2 Une représentation alternative de l'opinion publique ?.....	150
1.4 Les études alternatives de l'opinion	152
1.4.1. Les approches qualitatives de formation des opinions	153
1.4.2. Les mesures délibératives de l'opinion	156
1.4.3 Mesures virtuelles et mesures empiriques de l'opinion publique.....	157
2. Les autopublications numériques, une matérialisation empirique des théories critiques de l'opinion publique.....	159
2.1 Mesurer l'inégale capacité à produire une opinion	160
2.1.1 Un enregistrement « neutre » des opinions numériques.....	161
2.1.2 Objectiver le mode de production des opinions numériques.....	162
2.2 Mesurer l'inégale valeur des opinions numériques	165
2.3 Mesurer les préoccupations des enquêtés plutôt que celles des enquêteurs	167
2.4 Contextualiser la production des opinions numériques.....	171
2.4.1 Intégrer des éléments contextuels à l'analyse dispositionnelle des modes de production des opinions politiques	172
2.4.1.1 Modes de production et facteurs de production.....	173
2.4.1.2 Les facteurs contextuels d'activation de la production des opinions politiques..	176
2.4.2 Objectiver les éléments de contexte sur internet	178
3. Les limites de l'usage exclusif d'une posture passive d'enregistrement des opinions numériques.....	180
3.1 Une unité de mesure problématique	181
3.2 L'impossible caractérisation sociologique des producteurs d'opinion numérique.....	184
3.3 « Accepter d'abandonner la référence sociodémographique » ou abandonner l'analyse statistique ?	186
4. Panéliser Twitter : concilier une posture passive d'observation et une logique active de qualification sociologique des producteurs de tweets politiques	188

4.1 Panéliser un réseau social particulier : le choix de Twitter	189
4.2 Le principe et les avantages méthodologiques de la panélisation de Twitter.....	192
4.2.1 Le principe de la panélisation.....	192
4.2.2 Les avantages méthodologiques de la panélisation	193
3. Les étapes de constitution du panel	195
4.3.1 Définition et délimitation de la population-mère	195
4.3.2 Échantillonnage et analyse d'éligibilité.....	198
4.3.3 Administration du questionnaire et taux de réponse	202
4.3.4 Constitution du panel de contrôle des individus qui n'ont pas répondu au questionnaire	203

SECONDE PARTIE : TWITTER, UN OBSERVATOIRE DU MARCHÉ DES OPINIONS POLITIQUES 207

CHAPITRE 3 – TWITTER, UN NOUVEL OBSERVATOIRE 209

DU « PAYS LÉGAL » 209

1. Les propriétés sociologiques et politiques singulières des producteurs de tweets politiques 212

1.1 Les propriétés sociologiques des producteurs de tweets politiques.....	212
1.1.1 Une jeunesse surreprésentée mais non représentative.....	213
1.1.2 Une position sociale élevée	222
1.1.3 Des producteurs masculins, Parisiens, et fortement diplômés.....	229
1.1.3.1 Sexe et production de tweets politiques	230
1.1.3.1 Lieu d'habitation et production de tweets politiques.....	233
1.1.3.1 Niveau d'étude et production de tweets politiques.....	236
1.2 Une population fortement politisée et politiquement ancrée à gauche.....	239
1.2.1 Une population fortement politisée	240
1.2.2 Une population fortement ancrée à gauche	246

2. Les usages socialement distinctifs de Twitter 250

2.1 L'« identité numérique » comme manifestation empirique de l'habitus	251
1.1.1 Du concept d'« identité numérique » à celui d'habitus	252
2.1.2 Se « décrire » en moins de 160 caractères, une pratique socialement distinctive	255
2.2 Une tentative de typologie des producteurs de tweets politiques	264
2.2.1 Le profil « post-it ».....	266
2.2.2 Le « profil militant » et les « auteurs militants ».....	268
2.2.3 Le consommateur « ostentatoire » d'informations politiques	274

CHAPITRE 4 – LES INÉGALITÉS DE PRODUCTION DE TWEETS POLITIQUES..... 279

1. Les facteurs individuels de production de tweets politiques.....	281
1.1 Comparaison volumétrique de la production des panels « répondants » et « non répondants »	281
1.2 Les facteurs individuels de production de tweets politiques	288
2. L'inégale distribution du nombre de tweets politiques publiés dans le temps.....	304
2.1 Une inégale distribution dans le temps.....	304
2.2 Une inégale distribution temporelle qui concerne toutes les catégories de producteurs ...	308
3. « Offre » et « demande » d'informations et d'opinions politiques sur Twitter.....	313
3.1 Le marché de l'opinion.....	314
3.2 Mesurer l'« offre » d'opinions sur Twitter : le panel « Twitter audience politique »	319
3.2.1 La réalisation du panel « Twitter audience politique ».....	319
3.2.1 La composition du panel « Twitter audience politique »	323
3.3 Une corrélation entre « offre » et « demande » d'opinions et d'informations politiques..	326

CHAPITRE 5 – LES LOGIQUES SOCIALES ET POLITIQUES DE CONSOMMATION DES BIENS POLITIQUES SUR TWITTER..... 331

1. Les logiques sociales et politiques de consommation d'informations et d'opinions politiques	333
1.1 Les logiques sociales et politiques d'abonnement.....	333
1.1.1 Les abonnements à des comptes de médias.....	342
1.1.2 Les logiques d'abonnement aux comptes de femmes et d'hommes politiques	352
1.2 La consommation effective de biens politiques	357
1.2.1 Des logiques de consommation très similaires à la pratique d'abonnement	358
1.2.2 Les logiques spécifiques des pratiques sociales de consommation effectives de biens politiques	363
2. Formes de consommation <i>indirecte</i> et marqueurs d'opinion.....	366
2.1 Une grille de lecture multicritères des tweets politiques	366
2.1.1 Les principes classificatoires des tweets politiques non retenus pour l'analyse.....	367
2.1.1.1 Classer les tweets politiques en fonction de leur contexte d'émission	368
2.1.1.2 Classer les tweets politiques en fonction de leurs qualités formelles	370
2.1.1.3 Classer les tweets politiques en fonction de la nature politique de leur objet	373
2.1.2 Classer les tweets politiques en fonction de leur format : distinguer opinion explicite et implicite.....	375
2.2 Opinions politiques et biens médiatiques	378

CONCLUSION : LES APPORTS ET LIMITES DE LA MÉTHODE DE PANÉLISATION DES RÉSEAUX SOCIAUX	381
1. La vérification empirique d’hypothèses souvent peu objectivées	383
1.1 Quantifier la (non) prise de parole politique	383
1.2 Démontrer l’inégale distribution dans le temps des expressions politiques	388
1.3 L’hypothèse d’une détermination de la demande par l’offre au prisme du débat sur l’opinion publique	390
2. Les limites de la logique de panélisation.....	394
2.1 Logique d’échantillonnage et offre d’opinions	396
2.2 Logique d’échantillonnage et demande d’opinions.....	400
3. Des dispositifs méthodologiques alternatifs de mesure des opinions numériques	403
3.1 Matérialiser et analyser les logiques de consommation horizontale des opinions	403
3.2 Des modalités alternatives de sociologisation des comptes	408
3.2.1 Déléguer aux enquêtés la tâche de passation du questionnaire	409
3.2.2 Automatiser l’analyse des goûts	411
3.2.3 Mutualiser et capitaliser les bases de données sur les réseaux socionumériques	412
Table des matières	416
Bibliographie.....	423
Table des illustrations	458
Annexes	463
Annexe 1. Glossaire	464
Annexe 2. Le questionnaire	467
Annexe 3. Caractéristiques générales de la population « répondants »	475
Annexe 4. Caractéristiques générales de la population « non répondants ».....	482
Annexe 5. Caractéristiques générales du panel « Twitter audience politique »	485
Annexe 6. Présentation du logiciel AMI Opinion Tracker	487
Annexe 7. Requête de sélection des tweets politiques.....	494
Annexe 8. Grille de codage des tweets politiques.....	496

Bibliographie

ADAMIC Lada et GLANCE Natalie, « The political Blogosphere and the 2004 U.S. Election: Divided They Blog », *Proceedings of the 3rd International Conference on Knowledge Discovery Data*, Chicago, 2005, p. 36-43.

AGARWAL Nitin, LIU Huan, TANG Lei et YU Philip, « Identifying the Influential Bloggers in a Community », *Proceedings of the 2008 International Conference on Web Search and Data Mining*, New York, 2008, p. 207-218

ALBOUY Valérie et WANECQ Thomas, « Les inégalités sociales d'accès aux grandes écoles », *Économie et statistique*, n° 361, 2003, p. 27-52.

ALDRIN Philippe, *Sociologie politique des rumeurs*, Paris, PUF, 2005.

ALLARD Laurence et VANDENBERGHE Frédéric, « Express yourself ! Les pages perso. Entre légitimation technopolitique de l'individualisme expressif et authenticité réflexive peer to peer », *Réseaux*, vol. 1, n°117, 2003, p. 191-219.

ANDERSON Christopher et PASKEVICUTE Aida, « Macro-politics and micro-behavior: mainstream politics and the frequency of political discussion in contemporary democracies », in Alan Zuckerman (dir.), *The social logic of politics. Personal networks as contexts for political behaviour*, Philadelphie, Temple University Press, 2005, p. 228-250.

ANSTEAD Nick, « Social media metrics: measuring manifestations of the public in online spaces », *Colloque international Regards critiques sur la participation politique en ligne*, Paris, 2013.

ASUR Sitaram et HUBERMAN Bernardo, « Predicting the Future with Social Media », *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Washington, 2010, vol. 1, p. 492-499.

AUDEMARD Julien, *Influences interpersonnelles. Comment le contexte structure les opinions et les votes*, Thèse de doctorat soutenue à Montpellier le 4 décembre 2013.

BAKER Keith et CHARTIER Roger, « Dialogue sur l'espace public », *Politix*, vol. 7, n°26, 1994, p. 5-22.

BAKSHY Eytan, HOFMAN Jake, MASON Winter et WATTS Duncan, « Identifying “Influencers” on Twitter », *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM 2011)*, Hong Kong, 2011.

BAKSHY Eytan, HOFMAN Jake, MASON Winter et WATTS Duncan, « Everyone's an Influencer: Quantifying Influence on Twitter », *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM 2011)*, Hong Kong, 2011.

BARABÁSI Albert-Lászlo et ALBERT Réka, « Emergence of Scaling in Random Networks », *Science*, n° 286, 1999, p. 509-512.

BARABÁSI Albert-Lászlo, *Linked. How Everything is Connected to Everything Else and What it Means for Business, Science and Everyday Life*, New York, Penguin, 2003.

BARATS Christine (dir.), *Manuel d'analyse du web*, Paris, Armand Colin, 2013.

BARATS Christine, LEBLANC Jean-Marc et FIALA Pierre, « Approches textométriques du web : corpus et outils », in Christine Barats (dir.), *Manuel d'analyse du web*, Paris, Armand Colin, 2013.

BARBONI Thierry et TREILLE Éric, « L'engagement 2.0. Les nouveaux liens militants au sein de l'e-parti socialiste », *Revue française de science politique*, 2010/6 (vol. 60), p. 1137-1157.

BASTIEN Frédérick et GREFFET Fabienne, « Les campagnes électorales sur Internet : Une comparaison entre France et Québec », *Hermès*, n° 54, 2009, p. 209-217.

BASTIEN Frédéric et BLANCHARD Gersende, « Les internautes face à la communication électorale à l'ère des campagnes postmodernes », in Philippe J. Maarek (dir.) *Présidentielle 2012, une communication politique bien singulière*, Paris, L'Harmattan, 2013, p. 137-148.

BAUMARD Philippe, « Quelles approches avec quelles données ? » in Marie-Laure Gavard-Perret, David Gotteland, Christophe Haon et Alain Jolibert (dir.), *Méthodologie de la recherche en sciences de gestion, Réussir son mémoire ou sa thèse*, Paris, Pearson, 2012.

BEAUDOUIN Valérie et VELKOVSKA Julia, « Constitution d'un espace de communication sur Internet. (Forums, pages personnelles, courrier électronique...) », *Réseaux*, vol. 17, n° 97, 1999, p. 121-177.

BEAUVALLET Godefroy et RONAI Maurice, « Vivre à temps réels. Le renouvellement des pratiques militantes autour des TIC est-il possible au sein des partis de gouvernement ? », *Réseaux*, vol. 1-2, n° 129-130, 2005, p. 275-309.

BEAUVISAGE Thomas, BEUSCART Jean-Samuel, COURONNE Thomas et MELLET Kevin, « Le succès sur Internet repose-t-il sur la contagion ? Une analyse des recherches sur la viralité », *Tracés*, 21/2011, p. 151-166.

BEAUVISAGE Thomas, « Compter, mesurer et observer les usages du web : outils et méthodes », in Christine Barats (dir.), *Manuel d'analyse du Web*, Paris, Arman Colin, 2013.

BELISLE Claire (dir.), *Lire dans un monde numérique*, Villeurbanne, Presses de l'enssib, 2011.

BENKLER Yochai, *La Richesse des réseaux. Marchés et libertés à l'heure du partage social*, Lyon, Presses Universitaires de Lyon, 2009.

BENVEGNI Nicolas, « Le débat public en ligne. Comment s'équipe la démocratie dialogique ? », *Politix*, vol.3, n° 75, 2006, p. 103-124.

BENVEGNI Nicolas et BRUGIDOU Mathieu, « Prendre la parole sur Internet. Des dispositifs sociotechniques aux grammaires de la discussion », *Réseaux*, vol. 4, n° 150, 2008, p. 51-82.

BERELSON Bernard, LAZARSELD Paul et MCPHEE William, *Voting. A study of Opinion Formation in a Presidential Campaign*, Chicago, The University of Chicago Press, 1954.

BERNARD Lise, « Réflexions sur la “petite bourgeoisie nouvelle” dans les années 2000 », in Philippe Coulangeon et Julien Duval (dir.) *Trente ans après "La Distinction" de Pierre Bourdieu*, La Découverte, Paris, p. 266-277.

BEUSCART Jean-Samuel, « Sociologie des activités en ligne (introduction) », *Terrains & Travaux*, vol. 1, n°15, 2009, p. 3-28.

BEUSCART Jean-Samuel, CARDON Dominique, PISSARD Nicolas et PRIEUR Christophe, « Pourquoi partager mes photos de vacances avec des inconnus ? Les usages de Flickr », *Réseaux*, vol. 2, n°154, 2009, p. 91-129.

BIGEY Magali, « Littérature et sentiments. Assumer ses choix de lecture sur le web », *Les cahiers du numérique*, vol. 7, 2011, p. 85-102.

BLANCHARD Gersende, GADRAS Simon et WOJCIK Stéphanie, « Analyser la participation politique en ligne : des traces numériques aux pratiques sociales » in Christine Barats (dir.), *Manuel d'analyse du Web*, Paris, Arman Colin, 2013.

« LES BLOGS », *Réseaux*, vol. 24, n° 138, 2006.

BLONDIAUX Loïc, « Paul F. Lazarsfeld (1901-1976) et Jean Stœtzl (1910-1987) et les sondages d'opinion : genèse d'un discours scientifique », *Mots*, vol. 23, n° 23, 1990, p. 5-23.

BLONDIAUX Loïc, « Le chiffre et la croyance. L'importation des sondages d'opinion en France ou les infortunes d'une opinion sans publics », *Politix*, vol. 7, n°25, 1994, p. 117-152.

BLONDIAUX Loïc, « Mort et résurrection de l'électeur rationnel. Les métamorphoses d'une problématique incertaine », *Revue française de science politique*, 1996/5 (vol. 46), p. 753-791.

BLONDIAUX Loïc, *La Fabrique de l'opinion. Une histoire sociale des sondages*, Paris, Le Seuil, 1998.

BLONDIAUX Loïc, « Les théories contemporaines de l'opinion publique : un retour aux “classiques” », *Hermès*, n°31, 2001.

BLONDIAUX Loïc et ZALLER John, « Repenser l'opinion. Extrait de *The Nature and Origins of Mass Opinion*, New York, Cambridge University Press, 1992, p. 91-96 », *Hermès*, vol. 3, n°31, 2001, p. 29-38.

BLONDIAUX Loïc, « Sondages et délibération. Une épistémologie alternative de l'opinion publique », *Politix*, vol. 15, n°57, 2002, p. 167-180.

BLONDIAUX Loïc, « Faut-il se débarrasser de la notion de compétence politique ? Retour critique sur un concept classique de la science politique », *Revue française de science politique*, 2007/6 (vol. 57), p. 759-774.

BLONDIAUX Loïc, « La démocratie par le bas. Prise de parole et délibération dans les conseils de quartier du vingtième arrondissement de Paris », *Hermès*, n°26-27, 2010, p. 323-338.

BOLLEN Johan, MAO Huina et ZENG Xiao-Jun, « Twitter mood predicts the stock market », *Journal of Computational Science*, vol. 2, n°1, 2010, p. 1-8.

BOULLIER Dominique, « La nouvelle matérialité de l'espace public : les dispositifs de la net-politique », *Hermès*, n° 26-27, 2000, p. 201-211.

BOULLIER Dominique et LOHARD Audrey, *Opinion mining et Sentiment analysis : Méthodes et outils*, Marseille, OpenEdition Press, 2012.

BOURDIEU Pierre et PASSERON Jean-Claude, « Sociologues des mythologies et mythologies des sociologues », *Les temps modernes*, n°211, 1963, p. 998-1021.

BOURDIEU Pierre, CHAMBOREDON Jean-Claude et PASSERON Jean-Claude, *Le métier de sociologue. Préalables épistémologiques*, Paris, Éditions de l'EHESS, 1968.

BOURDIEU Pierre, « L'opinion publique n'existe pas », *Les temps modernes*, n°318, 1972, p. 1292-1309.

BOURDIEU Pierre, « Les doxosophes », *Minuit*, 1972.

BOURDIEU Pierre, « Un jeu chinois », *Actes de la recherche en sciences sociales*, vol. 2, n°2-4, 1976, p. 91-101.

BOURDIEU Pierre, « Question de politique », *Actes de la recherche en science sociales*, vol. 16, n°1, septembre 1977, p. 55-89.

BOURDIEU Pierre, *La Distinction, Critique sociale du jugement*, Les Éditions de Minuit, 1979.

BOURDIEU Pierre, « La jeunesse n'est qu'un mot », *Questions de sociologie*, Paris, Éditions de Minuit, 1984, p. 143-154.

BOURDIEU Pierre, *La noblesse d'État. Grandes écoles et esprit de corps*, Paris, Éditions de Minuit, 1989.

BOURDIEU Pierre, *Ce que parler veut dire. L'économie des échanges linguistiques*, Paris, Éditions Fayard, 1991.

BOURDIEU Pierre, « Esprit d'État », *Actes de la recherche en sciences sociales*, vol. 96, n° 96-97, 1993. p. 49-62.

BOURDIEU Pierre, *La misère du monde*. Paris, Seuil, 1993.

BOURDIEU Pierre, « Le mystère du ministère. Des volontés particulières à la "volonté générale" », *Actes de la recherche en sciences sociales*, vol. 140, n°140, 2001, p. 7-11.

BOY Daniel et MAYER Nona, *L'électeur a ses raisons*, Paris, Les Presses de Sciences Po, 1997.

BOYD Danah, GOLDER Scott et LOTAN Gilad, « Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter », *International Conference on System Sciences (HICSS)*, 43rd, Hawaii, 2010.

BRACONNIER Céline et DORMAGEN Jean-Yves, *La démocratie de l'abstention. Aux origines de la démobilisation en milieu populaire*, Paris, Gallimard, 2007.

BRAUD Philippe, « Données psycho-biographiques et formation des opinions politiques », *Revue française de science politique*, vol.24, n°3, 1974, p. 596-620.

BREGMAN Dorine, « La fonction d'agenda : une problématique en devenir », *Hermès*, vol. 1, n°4, 1989, p.191-202.

BROOKS Rod, « Word-of-Mouth in Selling New Products », *Journal of Marketing*, 22 (2), 1957, p. 154-161.

BRUGIDOU Mathieu, *L'opinion et ses publics. Une approche pragmatique de l'opinion publique*, Paris, Les Presses de Science Po, 2008.

CAMPBELL Donald et FISKE Donald, « Convergent and discriminant validation by the multitrait-multimethod matrix », *Psychological Bulletin*, vol. 56(2), 1959.

CANTRELL Paul, « Opinion polling and American democratic culture », *International Journal of Politics, Culture and Society*, vol.5, n°3, 1992, p. 415-416.

CARDON Dominique, « Présentation », *Réseaux*, vol. 3, n°197, 2006, p. 9-25.

CARDON Dominique, « La production de soi comme technique relationnelle. Un essai de typologie des blogs par leurs publics », *Réseaux*, vol. 4, n°138, 2006, p. 15-71.

CARDON Dominique, « Le design de la visibilité. Un essai de cartographie du web 2.0 », *Réseaux*, vol. 6, n°152, 2008, p. 93-137.

CARDON Dominique, *La démocratie internet. Promesses et limites*, Paris, Éditions du Seuil, 2010.

CARDON Dominique et GRANJON Fabien, *Médiactivistes*, Paris, Les presses de Science Po, 2010.

CARDON Dominique, FOUETILLOU Guilhem et LERONDEAU Clémence, « Esquisse de géographie de la blogosphère politique (2007-2009), in Fabienne Greffet (dir.) *Continuerlalutte.com, Les partis politiques sur le web*, Paris, Les Presses de Science Po, 2011.

CARDON Dominique, « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », *Réseaux*, vol. 1, n°177, 2013, p. 69-95.

CEVIPOF, « Le baromètre de la confiance politique », 2011.

CEVOPOF, « Sondage post-électoral 2012 », 2012.

CHA Meeyoung, HADDADI Hamed, BENEVENUTO Fabricio et GUMMADI Krishna, « Measuring User Influence on Twitter: the Million Follower Fallacy », *ICWSM'10: Proceedings of international AAAI Conference on Weblogs and Social*, Washington, Mai 2010.

CHADWICK Andrew et HOWARD Philip (dir.), *Routledge Handbook of Internet Politics*, London, Routledge, 2008.

CHADWICK Andrew, « Recent Shifts in the Relationship Between the Internet and Democratic Engagement in Britain and the United States: Granularity, Informational Exuberance, and Political Learning », in Michael Jensen et Laia Jobra (dir.), *Digital Media and Political Engagement Worldwide: A Comparative Study*, Cambridge, Cambridge University Press, 2012.

CHAMPAGNE Patrick, « Le cercle politique. Usages sociaux des sondages et nouvel espace politique », *Actes de la recherche en sciences sociales*, vol. 71, n°71-72, 1988, p. 71-97.

CHAMPAGNE Patrick, « Qui a gagné ? Analyse interne et analyse externe des débats politiques à la télévision », *Mots*, n°20, 1989, p. 5-22.

CHAMPAGNE Patrick, MANIN Bernard et PARODI Jean-Luc, « Quand les sondages se débattent – Débat entre Patrick Champagne, Bernard Manin et Jean-Luc Parodi », *Politix*, vol. 2, n°5, 1989, p. 25-46.

CHAMPAGNE Patrick, *Faire l'opinion. Le nouveau jeu politique*, Paris, Les Éditions de Minuit, 1990.

CHAMPAGNE Patrick, « La loi des grands nombres », *Actes de la recherche en science sociale*, vol. 101, n°101-102, 1994, p. 10-22.

CHAPUT Mathieu, « Analyser la discussion politique en ligne. De l'idéal délibératif à la reconstruction des pratiques argumentatives », *Réseaux*, n°150, 2008, p. 83-106.

HAZEL François, « La méthode du panel et ses possibilités d'application à la structure politique française », *Revue française de sociologie*, vol. 7, n° HS, 1966, p. 664-699.

CLAMENS Gilles, « En finir avec l'opinion », *Autres temps*, n°58, 1998, p. 40-48.

COAVOUX Samuel, « Introduction. Des classes sociales 2.0 ? », *Reset*, n°1, 2012, p. 4-24.

COCHOY Frank, *Une Histoire du marketing. Discipliner l'économie de marché*, Paris, La Découverte, 1999.

COINTET Jean-Philippe et ROTH Camille, « Social-semantic dynamics in a blog network », *IEEE International Conference on Social Computing (SocialCom-09)*, Vancouver, Canada, 2009.

COINTET Jean-Philippe et ROTH Camille, « Social and semantic coevolution in knowledge networks », *Social Networks*, n°32, 2010, p. 16-29.

COLEMAN James, KATZ Elihu et MENZEL Herbert, *Medical Innovation. A Diffusion Study*, The Bobbs-Merrill Cy, 1966.

COLEMAN Stephen, « Connecting Parliament to the Public via the Internet: Two Case Studies of Online Consultations », *Information, Communication & Society*, vol. 7, n°1, 2004, p. 3-22.

COLEMAN Stephen et MOSS Giles, « Governing at a distance – politicians in the blogosphere », *Information Polity*, vol. 13, 2008, p. 7-20.

COLEMAN Stephen et BLUMLER Jay, *The Internet and Democratic Citizenship, Theory, Practice and Policy*, Cambridge, Cambridge University Press, 2009.

COLOVALD Annie et LEFEBVRE Rémi, « Le militantisme socialiste n'est plus ce qu'il n'a jamais été. Modèle de l'engagement distancié et transformations du militantisme au Parti Socialiste », *Politix*, vol. 26, n°102, 2013, p. 9-33.

COMBY Jean-Baptiste, DEVILLARD Valérie, DOLEZ Charlotte et RIEFFEL Rémy, « Les appropriations différenciées de l'information en ligne au sein des catégories sociales supérieures », *Réseaux*, vol. 6, n°170, 2011, p. 75 -102.

COMBY Jean-Baptiste, « L'orientation sociale des goûts en matière d'actualité », in Josiane Jouët et Rémy Rieffel (dir.), *S'informer à l'ère numérique*, Rennes, Presses Universitaires de Rennes, 2013.

CONOVER Michael, RATKIEWICZ Jacob, FRANCISCO Matthew, GONCALVES Bruno, MENCZER Filippo et FLAMMINI Alessandro, « Political polarization on Twitter », *Fifth International AAAI Conference on Weblogs and Social Media*, Barcelone, 2011.

CONOVER Michael, RATKIEWICZ Jacob, FRANCISCO Matthew, GONCALVES Bruno, MENCZER Filippo et FLAMMINI Alessandro, « Predicting the Political Alignment of Twitter Users », *Privacy, security, risk and trust (passat), 2011 ieee third international conference on on social computing (socialcom)*, Boston, 2011, p. 192-1999.

CONVERSE Philip, « The Nature of Mass Beliefs in Mass Publics », in David E. Apter (dir.), *Ideology and Discontent*, New York, The Free Press, 1964, p. 206-261.

CORI Marcel et LEON Jacqueline, « La constitution du TAL. Étude historique des dénominations et des concepts », *TAL*, vol. 43, n° 3, 2002, p. 21-55

COUTANT Alexandre et STENGER Thomas, « Les activités quotidiennes des jeunes sur les réseaux socionumériques », *Réel-virtuel*, n°2, 2011.

CRAMER WALSH Katherine, *Talking About Politics. Informal Groups and Social Identity in American Life*, Chicago, The University of Chicago Press, 2004

CRIBIER Françoise, *Rapport sur la sauvegarde des données qualitatives des sciences sociales*, Avril 2003.

DAGIRAL Éric et PARASIE Sylvain, « Presse en ligne : où en est la recherche ? », *Réseaux*, vol. 2-3, n°160-161, 2010, p. 13-42.

DAHLBERG Lincoln, « Computer-Mediated Communication and the Public Sphere: a critical analysis », *Journal of Computer-Mediated Communication*, vol. 7, n°1, 2001.

DAHLBERG Lincoln, « Rethinking the fragmentation of the cyberpublic: from consensus to contestation », *New Media & Society*, vol. 9, n° 5, 2007, p. 827-847.

DARGERÉ Christophe, *L'observation incognito en sociologie, Notions théoriques, démarche réflexive, approche pratique et exemples concrets*, Paris, L'Harmattan, 2012.

DARES (enquêtes de la), « Emploi et chômage des 15-29 ans en 2012 », novembre 2013.

DENOUEËL Julie et GRANJON Fabien (dir.), *Communiquer à l'ère numérique. Regards croisés sur la sociologie des usages*, Paris, Presse des mines, 2011.

DENVER David, HANDS Gordon et JONES Bill, « Fishkin and the Deliberative Opinion Poll: Lessons from the Study of the Granada 500 TV Program », *Political Communication*, 12, 1995, p. 147-156.

DESQUINABO Nicolas, « Interactions et argumentations dans les webforums partisans », *Communication au colloque Les usages partisans de l'Internet organisé par l'équipe IRÉNÉE*, Université de Nancy 2, Nancy, 21-22 juin 2007.

DESQUINABO Nicolas, « Dynamiques et impacts des propositions politiques dans les webforums partisans », *Réseaux*, vol. 4, n°150, 2008, p. 109-132.

DESROSIÈRES Alain, *La politique des grands nombres. Histoire de la raison statistique*, Paris, Éditions La Découverte, 1993.

DESROSIERES Alain, « Entre réalisme métrologique et conventions d'équivalence : les ambiguïtés de la sociologie quantitative », *Genèses*, vol. 2, n°43, 2001, p. 112-127.

DEVILLARD Valérie, DOLEZ Charlotte et RIEFFEL Rémy, « La consommation de l'information entre engagement professionnel et implication civique », in Josiane Jouët et Rémy Rieffel (dir.), *S'informer à l'ère numérique*, Rennes, Presses Universitaires de Rennes, 2013.

DIAKOPOULOS Nicholas et SHAMMA David, « Characterizing debate performance via aggregated twitter sentiment », *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 1195-1198.

DI MAGGIO Paul, HARGITTAI Eszter, NEUMAN Russel et ROBINSON John, « Social Implication of the Internet », *Annual Review of Sociology*, 27, 2001, p. 307-336.

DONNA Olivier, *Les pratiques culturelles des Français à l'ère numérique. Enquête 2008*, Paris, Éditions La Découverte, 2009.

DORMAGEN Jean-Yves et MOUCHARD Daniel, *Introduction à la sociologie politique*, de boeck, 2010, 3ème édition.

DOURY Marianne et MARCOCCIA Michel, « Forum internet et courrier des lecteurs : l'expression publique des opinions », *Hermès*, n°47, 2007, p. 41-50.

DREZNER Daniel et FARRELL Henry, « The power and politics of blogs », *Public Choice*, n°134, 2008, p. 15-30.

DUTHIL Benjamin, TROUSSET François, DRAY Gérard, MONTMAIN Jacky et PONCELET Pascal, « Vers une caractérisation automatique de critères pour l'opinion mining », *Les cahiers du numériques*, vol. 7, 2011, p. 41-62.

DUSCHENE Sophie et HAEGEL Florence, « La politisation des conversations au croisement des logiques de spécialisation et de conflictualisation », *Revue Française de Science Politique*, vol. 54, n°6, décembre 2004, p. 877-909.

DUSCHENE Sophie et HAEGEL Florence, *L'entretien collectif. L'enquête et ses méthodes*, Paris, Armand Colin, 2008.

DUMOULIN Michaël, « Les forums électroniques : délibératifs et démocratiques ? », in Denis Monière (dir.), *Internet et la Démocratie*, Québec, Monière et Wollank éditeurs, 2002, p. 141-157.

DUSSUEIL Jeanne, « Comment le gouvernement écoute ce qui se dit sur Twitter grâce la technologie d'une start up », *Challenges*, 26 novembre 2012.

DUTTON William, « Network rules of order: regulating speech in public electronic fora », *Media, Culture & Society*, vol. 18, n° 2, 1996, p. 269-290.

EENSOO-RAMDANI Egle, BOURION Evelyne, SLODZIAN Monique et VALETTE Mathieu, « De la fouille de donnée à la fabrique de l'opinion », *Les Cahiers du numérique*, vol. 7, 2011, p. 15-39.

ELIASOPH Nina, *L'évitement du politique. Comment les Américains produisent l'apathie dans la vie quotidienne*, Paris, Economica, 2010.

ELSTER Jon, « Argumenter et négocier dans deux assemblées constituantes », *Revue française de science politique*, vol. 44, n°2, 1994, p. 187-256.

ERTZSCHEID Olivier, GALLEZOT Gabriel et SIMONNOT Brigitte, « À la recherche de la "mémoire" du web : sédiments, traces et temporalités des documents en ligne », in Christine Barats (dir.), *Manuel d'analyse du Web*, Paris, Arman Colin, 2013.

EUROSTAT, *Statistiques sur la société de l'information*, DEPS, 2013.

EVELAND William et HUTCHENS Myiah, « Political Discussion Frequency, Network Size, and "Heterogeneity" of Discussion as Predictors of Political Knowledge and Participation », *Journal of Communication*, n°59, 2009, p. 205-224.

EVELAND William, MARTON Krisztina et SEO Mihye, « Moving Beyond “Just the Facts”. The Influence of Online News on the Content and Structure of Public Affairs Knowledge », *Communication Research*, vol. 31, n°1, 2004, p. 81-108.

FAVRE Pierre, « Politix, 1988-2012 : changement de génération, basculement de paradigmes », *Politix*, vol. 4, n° 100, 2012, p. 41-62.

FISHKIN James, *Democracy and Deliberation*, Yale, Yale University Press, 1991.

FLICHY Patrice, *L'imaginaire d'internet*, Paris, La Découverte, 2001.

FLICHY Patrice, « Internet et le débat démocratique », *Réseaux*, n°150, 2008, p. 159-185.

FLICHY Patrice, « La démocratie 2.0 », *Études*, n°412, mai 2010, p.41-50.

FLICHY Patrice, *Le sacre de l'amateur. Sociologie des passions ordinaires à l'ère numérique*, Paris, Seuil, 2010.

FLORES Laurent, « Web 2.0 et études de marché : vers une nouvelle génération d'études de marché ? », *Revue Française du Marketing*, n°220, 2008, p. 7-16.

FOUETILLOU Guilhem, « Le web et le traité constitutionnel européen. Écologie d'une localité thématique compétitive », *Réseaux*, vol. 5, n°144, 2007, p. 279-304.

« FRACTURE NUMERIQUE ET JUSTICE SOCIALE », *Les Cahiers du numérique*, vol. 5, n° 1, 2009.

« LA FRACTURE NUMERIQUE », *Réseaux*, vol. 5-6, n° 127-128, 2004.

FRIPPIAT Didier et MARQUIS Nicolas, « Les enquêtes par Internet en sciences sociales : un état des lieux », *Population-F*, n°65, 2010, p. 309-338.

FUNKHOUSER G. Ray, «The issues of the sixties: an exploratory study in the dynamics of public opinion», *Public Opinion Quarterly*, n°37, 1973, p. 62-75.

GADRAS Simon et GREFFET Fabienne, « Espaces officiels et espaces non officiels de campagne », in Jacques Gerstlé et Raul Magni Berton (dir.), 2012, *la campagne présidentielle*, Paris, L'Harmattan, 2013, p. 73-86

GAMSON William, *Talking politics*, Cambridge University Press, 1992.

GARCIA-BARDIDIA Renaud, « Des communications interpersonnelles au marketing viral: état des lieux », Document de travail, 2006.

GAXIE Daniel, *Le cens caché. Inégalités culturelles et ségrégation politique*, Paris, Édition du Seuil, 1978.

GAXIE Daniel, « Au-delà des apparences... », *Actes de la recherche en science sociale*, vol. 81, n°81-82, 1990, p. 97-112.

GAXIE Daniel, COLLOVALD Annie, GAÏTI Brigitte, LEHINGUE Patrick et POIRMEUR Yves (dir.), « Le “social” transfiguré - sur la représentation politique des préoccupations “sociales” », Paris, PUF-CURAPP, 1990.

GAXIE Daniel, « Appréhender le politique à l'aune des expériences sociales », *Revue Française de Science Politique*, vol. 52, n°2-3, 2002, p. 145-178.

GAXIE Daniel, « Cognitions, auto-habilitation et pouvoir des “citoyens” », *Revue française de science politique*, 2007/6 (vol. 57), p. 737-757.

GAXIE Daniel, HUBÉ Nicolas, DE LASSALLE Marine et ROWELL Jay (dir.), *L'Europe des Européens. Enquête comparative sur les perceptions de l'Europe*, Paris, Economica, 2010.

GAXIE Daniel, « Retour sur les modes de production des opinions politiques », in Philippe Coulangeon et Julien Duval (dir.), *Trente ans après La Distinction de Pierre Bourdieu*, Paris, La Découverte, 2013.

GENSOLLEN Michel, « Économie non-rivale et communautés d'information », *Réseaux*, vol. 2, n°124, 2004, p. 141-206.

GEORGES Fanny, « Représentation de soi et identité numérique. Une approche sémiotique et quantitative de l'emprise culturelle du web 2.0 », *Réseaux*, vol. 2, n°154, 2009, p. 165-193.

GJOKA Minas, KURANT Maciej, BUTTS Carter et MARKOPOULOU Athina, « Walking in Facebook: A Case Study of Unbiased Sampling of OSNs », *Proceedings of the 29th conference on Information communications*, 2010, p. 2498-2506.

GLADWELL Malcolm, *The Tipping Point: How little things can make a big difference*, New York, Little, Brown and Company, 2002.

GLASER Barney et STRAUSS Anselm, *The Discovery of Grounded Theory*, New York, Aldine Publishing Company, 1967, cité dans David Silverman, *Interpreting Qualitative Data*, London, Sage, 1993.

GOLLAC Michel et KRAMARZ Francis, « L'informatique comme pratique et comme croyance », *Actes de la recherche en sciences sociales*, n° 134, 2000, p. 4-21.

GOMBAULT Vincent, « L'internet de plus en plus prisé, l'internaute de plus en plus mobile », *INSEE Première*, n°1452, juin 2013.

GOULET Vincent, *Médias et classes populaires. Les usages ordinaires des informations*, Paris, INA Édition, 2010.

GRANJON Fabien, *L'internet militant. Mouvement social et usages des réseaux télématiques*, Rennes, éditions Apogée, 2001.

GRANJON Fabien, « Les répertoires d'action télématiques du néo-militantisme », *Le Mouvement Social*, vol. 3, n° 200, 2002, p. 11-32.

GRANJON Fabien, « L'Internet militant. Entretien avec Fabien Granjon », *Matériaux pour l'histoire de notre temps*, 2005, vol. 79, n°79, p. 24-29.

GRANJON Fabien, « Une approche critique de la fracture numérique. Champ de l'Internet, pratiques télématiques et classes populaires », *Cahiers de recherche M@rsouin*, n° 1, janvier 2005.

GRANJON Fabien, BLANCO Catherine, LE SAULNIER Guillaume et MERCIER Grégory, « Sociabilités et familles populaires », *Réseaux*, vol. 6-7, n°145-146, 2007, p. 117-157.

GRANJON Fabien, LELONG Benoît et METZGER Jean-Luc, *Inégalités numériques – Clivages sociaux et mode d’appropriation des TIC*, Paris, *Hermès*, 2009.

GRANJON Fabien, « Inégalités numériques et reconnaissance sociale. Des usages populaires de l’informatique connectée », *Les Cahiers du numériques*, 2009/1, vol. 5, p. 19-44.

GRANJON Fabien et DENOÛËL Julie, « Exposition de soi et reconnaissance des singularités subjectives sur les sites de réseaux sociaux », *Sociologie*, vol. 1, n°1, 2010, p. 25-43.

GRANJON Fabien et LE FOULGOC Aurélien, « Les usages sociaux de l’actualité. L’Expérience médiatique des publics internautes », *Réseaux*, vol. 2, n°160-161, 2010, p. 225-253.

GRANJON Fabien et LE FOULGOC Aurélien, « Penser les usages sociaux de l’actualité », *Réseaux*, vol. 6, n°170, 2011, p. 17-43.

GRANJON Fabien, *Reconnaissance et usages d’Internet. Une sociologie critique des pratiques de l’informatique connectée*, collection Sciences Sociales, Presses des Mines, Paris, 2012.

GRANOVETTER Mark, « The Strength of Weak Ties », *American Journal of Sociology*, vol. 78, n°6, 1973.

GRAWITZ Madeleine, « Les sondages d’opinion publique », *Revue des travaux de l’Académie des sciences morales et politiques*, 125ème année, 1er semestre 1972, p. 99-108.

GRAWITZ Madeleine, *Méthodes des sciences sociales*, Paris, Dalloz, 2000.

GREFFET Fabienne, « Les blogs politiques : enjeux et difficultés de recherche à partir de l’exemple français », *Communication*, vol. 25, n° 2, 2007, p. 200-211.

GREFFET Fabienne et WOJCIK Stéphanie, « Parler politique en ligne. Une revue des travaux français et anglo-saxons », *Réseaux*, vol. 4, n°150, 2008, p. 19-50.

GROSSETTI Michel, « Plus largement, les sciences sociales... », *SociologieS* [En ligne]. En ligne depuis le 27 janvier 2012.

GRUMBACH Michel et HERPIN Nicolas, « À propos de quelques travaux de Lazarsfeld et de son école », *Enquête*, n°4, 1988, mis en ligne le 16 novembre 2005, consulté le 05 mai 2014. URL : <http://enquete.revues.org/65>

HABERMAS Jürgen, *L'espace public. Archéologie de la publicité comme dimension constitutive de la société bourgeoise*, Paris, Payot, 1998 (version française).

HABERMAS Jürgen, « "L'espace public" 30 ans après », *Quaderni*, vol. 18, n°18, 1992. p. 161-191.

HABERMAS Jürgen, *Droit et démocratie. Entre faits et normes*, Paris, Gallimard, 1997.

HACKING Ian, *Rewriting the Soul: Multiple Personality and Sciences of Memory*, Princeton University Press, 1995.

HAEGEL Florence, « La mobilisation partisane de droite. Les logiques organisationnelles et sociales d'adhésion à l'UMP », *Revue française de science politique*, 59 (1), février 2009, p. 7-27.

HALL Catherine, « Private Persons versus Public Someones: Class, Gender and Politics in England, 1780-1850 », in Carolyn Steedman, Cathy Urwin, Valerie Walkerdine (dir.), *Language, Gender and Childhood*, Londres, 1985.

HAMMAN Philippe, « Quand le souvenir fait le lien... De la délimitation des domaines de validité des énoncés recueillis par le sociologue en situation d'entretien », *Sociologie du travail*, 44 (2), 2002, p. 175-191.

HARGITTAI Eszter, « Second-Level Digital Divide: Differences in People's Online Skills », *First Monday*, volume 7, n° 4, 2002.

HARGITTAI Eszter, « Digital Na(t)ives? Variation in Internet Skills and Uses among Members of the “Net Generation” », *Sociological Inquiry*, n°80(1), 2009, p. 92-113.

HARGITTAI Eszter et LITT Eden, « The tweet smell of celebrity success: Explaining variation in Twitter adoption among a diverse group of young adults », *New Media & Society*, n°13(5), 2011, p. 824-842.

HARGITTAI Eszter et HINNANT Amanda, « Digital Inequality: Differences in Young Adults' Use of the Internet », *Communication Research*, n°35(5), p. 602-621.

HARGITTAI Eszter, « The Digital Reproduction of Inequality », in David Grusky, *Social Stratification: Class, Race, and Gender in Sociological Perspective*, Boulder, Westview Press, p. 936-944.

HARRIS-LACEWELL Melissa, *Barbershops, Bibles and BET: Every-Day Talk and Black Political Thought*, Princeton, Princeton University Press, 2004.

HATZIVASSILOGLOU Vasileios et WIEBE Janyce, « Effects of Adjective Orientation and Gradability on Sentence Subjectivity », *Proceedings of the 18th International Conference on Computational Linguistics*, vol. 1, no 11-12, 2000, p. 299-305.

HENRY Nathalie et FEKETE Jean-Daniel, « Représentations visuelles alternatives pour les réseaux sociaux », *Réseaux*, n°152, 2008, p. 61-92.

HERAULT Adeline et MOLINIER Pierre, « Les caractéristiques de la communication sociale via internet », *Empan*, n°76, p. 13-21.

HONG Liangjie, DAN Ovidiu et DAVISON Brian, « Predicting popular messages in Twitter », *Proceedings of the 20th international conference companion on World wide web*, ACM New York, p. 57-58.

HOPKINS Daniel et KING Gary, « A Method of Automated Nonparametric Content Analysis for Social Science », *American Journal of Political Science*, vol. 54, n°1, 2010, p. 229-247.

HINDMAN Matthew, *The Myth of Digital Democracy*, Princeton, Princeton University Press, 2009.

HIRSCHMAN Albert, *Exit, voice, loyalty. Défection et prise de parole*, Bruxelles, Éditions de l'Université de Bruxelles, 2011.

HUBERMAN Bernardo et ADAMIC Lada, « Growth Dynamics of the World Wide Web », *Nature*, vol. 401, n° 6749, 1999.

HUBERMAN Bernardo et ADAMIC Lada, « Power Law Distribution of the World Wide Web », *Science*, vol. 294, n°5641, 2000.

HUBERMAN Bernardo, ROMERO Daniel et WU Fang, « Social networks that matter: Twitter under the microscope », *First Monday*, vol. 14, n°1, janvier 2009, p. 1-9.

HUCKFELDT Robert, *Politics in Context: Assimilation and Conflict in Urban Neighborhoods*, New York, Agathon Press, 1986.

HUMPHREYS Laud, *Le commerce des pissotières. Pratiques homosexuelles anonymes dans l'Amérique des années 1960*, Paris, La découverte, 2007.

IFOP, « Observatoire des réseaux sociaux », 2011.

IFOP, « Observatoire des réseaux sociaux », 2012.

IFOP, « Observatoire des réseaux sociaux », 2013.

IFOP, « Twitter vu par ses utilisateurs », 2010.

INSEE (enquête de), « Technologies de l'information et de la communication », avril 2010.

ION Jacques, *La fin des militants ?*, Paris, Éditions de l'atelier, 1997.

JANSEN Harold et KOOP Royce, « Pundits, Ideologues, and Ranters: The British Columbia Election Online », *Canadian Journal of Communication*, vol.4, n°30, 2005, p. 613-632.

JANSEN Bernard, ZHANG Mimi, SOBEL Kate et CHOWDURY Abdur, « Twitter power: Tweets as electronic word of mouth », *Journal of the Association for Information Science and Technology*, vol. 60, n°11, 2009, p. 2169–2188.

JAVA Akshay, SONG Xiaodan, FININ Tim et TSENG Belle, « Why We Twitter: Understanding Microblogging Usage and Communities », *Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007*, 2007, p. 56-65.

JEANNE-PERRIER Valérie, « Parler de la télévision sur Twitter : une réception oblique à partir d'une conversation médiatique », *Communication & langages*, n°166, 2010, p. 127-147.

JENKINS Henry, « If it doesn't spread, it's dead: Media Viruses and Memes », *Confessions of an Aca fan – The blog of Henry Jenkins*, février 2009.

JERECZEK-LIPINSKA Joana, « Le blog en politique. Outil de démocratie électronique participative ? », *Revue de sociolinguistique en ligne*, n° 10, 2007, p. 159-172.

JOIGNANT Alfredo, « au politique », *Revue française de science politique*, 2007/6 (vol. 57), p. 799-817.

JOUËT Josiane, « Pratiques de communication et figures de la médiation », *Réseaux*, vol. 11, n° 60, 1993, p. 99-120.

JOUËT Josiane, « Retour critique sur la sociologie des usages », *Réseaux*, vol. 18, n°100, 2000, p. 487-521.

JOUËT Josiane, VEDEL Thierry et COMBY Jean-Baptiste, « Political information and interpersonal conversations in a multimedia environment: a quantitative and qualitative examination of information practices in France », *European Journal of Communication*, vol. 26, n°4, décembre 2011, p. 361-375.

JOUËT Josiane et LE CAROFF Coralie, « L'actualité politique et la participation en ligne », in Josiane Jouët et Rémy Rieffel (dir.), *S'informer à l'ère numérique*, Rennes, Presses Universitaires de Rennes, 2013, p.69.

JUNGHERR Andreas, JÜRGENS Pascal et SCHOEN Harald, « Why the Pirate Party Won the German Election of 2009 or the Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. “Predicting Elections With 140 Characters Reveal About Political Sentiment” », *Social Science Computer Review*, p. 1-6.

KATZ Elihu et LAZARSELD Paul, *Influence personnelle : ce que les gens font des médias*, Paris, Arman Colin, 2008.

KATZ Elihu, « The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis Author », *The Public Opinion Quarterly*, vol. 21, n°1, 1957, p. 61-78.

KELLER Ed et BERRY Jon, *The Influentials: one American in ten tells the other nine how to vote, where to eat and what to buy*, New York, Free Press, 2003.

KINDER Donald et SEARS David, « Public opinion and political action » in Gardner Lindzey et Eliot Aronson (dir.), *The Handbook of social psychology*, 4th. Ed. New York, Random House, 1985.

KOC-MICHALSKA Karolina, LILLEKER Darren et BAGOT Perle, « Interagir avec les candidats : les stratégies interactives lors de l'élection présidentielle française de 2012 », in Philippe J. Maarek (dir.) *Présidentielle 2012, une communication politique bien singulière*, Paris, L'Harmattan, 2013, p. 149-164.

KRAIF Olivier et PONTON Claude, « Du bruit, du silence et des ambiguïtés : que faire du TAL pour l'apprentissage des langues ? », *Actes de TALN*, Toulouse, 2007.

KWAK Haewoon, LEE Changhyun, PARK Hosung et MOON Sue, « What is Twitter, a social network or a news media ? », *Proceedings of the 19th international conference on World wide web*, NewYork, 2010, p. 591-600

LAHIRE Bernard, *La culture des individus. Dissonances culturelles et distinction de soi*, Paris, La Découverte, 2004.

LANDS Joan, *Women and the Public Sphere in the Age of the French Revolution*, Ithaca, 1988.

LARDELLIER Pascal et BRYON-PORTET Céline, « Ego 2.0. Quelques considérations théoriques sur l'identité et les relations à l'ère des réseaux », *Les cahiers du numérique*, vol.6, Lavoisier, 201, p. 13-34.

LATTE Stéphane « La “force de l'événement” est-elle un artefact ? », *Revue française de science politique*, 3/ 2012 (vol. 62), p. 409-432.

LATZKO-TOTH Guillaume et PROULX Serge, « Enjeux éthiques de la recherche sur le Web », in Christine Barats (dir.), *Manuel d'analyse du Web*, Paris, Arman Colin, 2013.

LAZARSELD Paul et FISKE Marjorie, « The “panel” as a New Tool for Measuring Opinion », *The Public Opinion Quarterly*, vol. 2, n°4, 1938, p. 596-612.

LAZARSELD Paul, BERELSON Bernard et GAUDET Hazel, *The people's choice: How the voter makes up his mind in a presidential campaign*, New York, Duell, Sloan, and Pearce, 1944.

LEGAVRE Jean-Baptiste, « Les “règles” du médiateur du *Monde* ou la mise en scène d'une écriture de presse », *Questions de communication*, 2007/2, n°12, p. 311-334.

LEGAVRE Jean-Baptiste, « Des effets paradoxaux de stratégies de communication : quelques réceptions inattendues d'un magazine territorial », *Questions de communication*, 2013/2, n°24, p. 199-217.

LE GRIGNOU Brigitte, « Les études de réception: un mauvais objet ? », in Didier Georgakakis et Jean-Michel Utard (dir.), *1. Sciences des médias. Jalons pour une histoire politique*, Paris. 2001, p. 179-194

LE GRIGNOU Brigitte, *Du côté du public : Usages et réceptions de la télévision*, Paris, Economica, 2003.

LE HAY Viviane, VEDEL Thierry et CHANVRIL Flora, « Usages des médias et politique : une écologie des pratiques informationnelles », *Réseaux*, vol. 6, n°170, 2011, p. 45-73.

LE HAY Viviane, VEDEL Thierry, CHANVRIL Flora, « Usages des médias et politique : une écologie des pratiques informationnelles », *Réseaux*, vol. 6, n° 170, 2011, p. 45-74.

LEHINGUE Patrick, « Le Non français au traité constitutionnel européen (mai 2005). Sur deux lectures “polaires” du scrutin », *Actes de la Recherche en Sciences Sociales*, n°166-167, mars 2007, p. 123-139.

LEHINGUE Patrick, « L'emprise des sondages sur le jeu électoral : vrais et faux débats », *Savoir/agir*, 1, septembre 2007, p. 37-48.

LEHINGUE Patrick, *Subunda. Coups de sonde dans l'océan des sondages*, Broissieux, Éditions du croquant, 2007.

LEHINGUE Patrick, « La détermination par les instituts de sondage de l'indétermination des électeurs », GUILBERT Philippe., HAZIZA David., RUIZ-GAZEN Anne, TILLÉ Yves, (dir), *Méthodes de sondage*, Dunod, 2008, p. 158-164.

LEHINGUE Patrick, *Le vote. Approches sociologiques de l'institution et des comportements électoraux*, Paris, La Découverte, 2011.

LE SAULNIER Guillaume, « La lecture de la presse en ligne. L'appropriation des contenus d'actualité au défi de la technique », in Josiane Jouët et Rémy Rieffel (dir.), *S'informer à l'ère numérique*, Rennes, Presses Universitaires de Rennes, 2013, p. 57-84.

LERMAN Kristina et GHOSH Rumi, « Information Contagion: an Empirical Study of the Spread of News on Digg and Twitter Social Networks », *Proceedings of 4th International Conference on Weblogs and Social Media (ICWSM-10)*, 2010.

LEV-ON Azi et MANIN Bernard, « Internet : la main invisible de la délibération », *Esprit*, vol. 5, 2006, p.195-212.

LIPPMAN Walter, *Le public fantôme*, Paris, Demopolis, 2008.

LIU Bing, « Sentiment Analysis and Subjectivity », Nitin Indurkha et Fred Damerau (dir.), *Handbook of Natural Language Processing* (2ème édition), Boca Raton, Chapman & Hall/CRC Press, 2010.

LIU Bing, « Sentiment Analysis and Subjectivity », in Nitin Indurkha et Fred Damerau (dir.), *Handbook of Natural Language Processing* (2ème édition), Boca Raton, Chapman & Hall/CRC Press, 2010.

LIVINGSTONE Peter et LUNT Sonia, *Talk on Television: Audience Participation and Public Debate*, Londres, Routledge, 1994.

LUNDBERG George « The newspaper and public opinion », *Social Forces*, vol. IV(4), 1926, p. 709-715.

MAC COMBS Max et SHAW Donald, «The agenda-setting function of mass-media », *Public Opinion Quarterly*, n°36, 1972, p. 176-187.

MALLEIN Philippe et TOUSSAINT Yves, « L'intégration sociale des technologies de l'information et de communication : une sociologie des usages », *Technologies de l'information et société*, vol. 6, n°4, 1994, p. 315-335.

MARCHEVA Marta, « Facebook, un outil de communication électorale supplémentaire à maîtriser. La page Facebook de Marine Le Pen, entre présence et absence », in Philippe J. Maarek (dir.) *Présidentielle 2012, une communication politique bien singulière*, Paris, L'Harmattan, 2013, p. 201-215.

MARCOCCIA Michel, « Parler de politique dans un forum de discussion », *Langage et société*, n°104, 2003, p. 9-55.

MARCOCCIA Michel, « Les webforums des partis politiques français : quels modèles de discussion politique ? », *Mots*, n° 80, 2006, p. 49-60.

MARCOCCIA Michel, « Les forums de discussion d'adolescents : pratiques d'écritures et compétences communicatives », *Revue française de linguistique appliquée*, vol. 15, 2010.

MARTIN Philippe, « L'état de la blogosphère 2008 selon Technorati indique 133 millions de blogs », article en ligne : <http://www.nayezpaspeur.ca/blog/2008/09/letat-de-la-blogosphere-2008-selon-technorati-indique-133-millions-de-blogs.html> (consulté le 6/10/2013).

MARTY Emmanuel, REBILLARD Franck, SMYRNAIOS Nikos et TOUBOUL Annelise, « Variété et distribution des sujets d'actualité sur Internet. Une analyse quantitative de l'information en ligne », *Mots*, n°93, 2010, p. 107-126.

MATHIAS Paul, *Qu'est-ce que l'internet ?*, Paris, Vrin, 2009.

MATHIEU Lilian, *Comment lutter ? Sociologie et mouvements sociaux*, Paris, Textuel, 2004.

MAYAFFRE Damon, « Vocabulaire et discours électoral de Sarkozy. Entre modernité et pétainisme », *La Pensée*, n° 352, 2007, p. 65-80.

MAYAFFRE Damon, « Débat de l'entre-deux-tours 2007. La conquête de l'espace lexical », *Mots*, vol. 1, n°89, 2009, p. 105-117.

MAYAFFRE Damon, « De la lexicométrie à la logométrie », Encyclopédie en ligne de l'Astrolabe http://hal.archivesouvertes.fr/docs/00/55/19/21/PDF/18_Mayaffre_Astrolabe_2005.pdf

MAYER Nona, *Sociologie des comportements politiques*, Paris, Armand Colin, 2010.

MELLET Kévin, « Aux sources du marketing viral », *Réseaux*, vol. 5-6, n° 157-158, 2009, p. 267-292.

MENEZES Telmo, ROTH Camille et COINTET Jean-Philippe, « Precursors and Laggards: An Analysis of Semantic Temporal Relationships on a blog Network », *IEEE SocialCom Intl Conf on Social Computing*, Minneapolis, Minnesota, 2010.

MERCIER Arnaud, « Médias et recherche d'information politique », in Pascal Perrineau et Luc Roban (dir.), *La politique en France et en Europe*, Paris, Presses de Science Po, 2007, p. 251-286.

MERCIER Arnaud, « Les médias vecteurs de mondialisation et “ d’événementialisation ” », in Paul Rasse (dir.), *La mondialisation de la communication*, Paris, CNRS éditions, 2010, p. 93-111.

MERCIER Arnaud, « L’utile fiction de l’opinion publique », in Arnaud Mercier (dir.), *Médias et opinion publique*, Paris, CNRS éditions, 2012, p. 15-40.

MERCIER Arnaud, « Étudier les internautes par des questionnaires en ligne », in Amandine Degand et Benoît Grevisse (dir.), *Journalisme en ligne*, Bruxelles, de Boeck, 2012, p. 292-297.

MERCIER Arnaud, « Avènement du Twiléspectateur et hashtags contestataires, faits marquants de la campagne sur les réseaux sociaux », in Philippe J. Maarek (dir.) *Présidentielle 2012, une communication politique bien singulière*, Paris, L’Harmattan, 2013, p. 165-201.

MERCIER Arnaud, « Twitter l’actualité : usages et réseautage chez les journalistes français », *Recherches en communication*, 39, 2013.

MERCKLE Pierre, *Sociologie des réseaux sociaux*, Paris, La Découverte, 2011.

MEUSNIER Norbert, « Argumentation et démonstration de la loi des grands nombres » in Evelyne Barbin (dir.), *La démonstration mathématique dans l’histoire*, Besançon, IREM, 1989, p. 89-97.

MICHELAT Guy et SIMON Michel, « Les “sans réponse” aux questions politiques », *Pouvoir*, n°33, 1985, p. 41-56.

MIHALCEA Rada, BANEA Carmen et WIEBE Janyce, « Learning multilingual subjective language via crosslingual projections », *Proceedings of the Association for Computational Linguistics (ACL)*, Prague, Czech Republic, 2007.

MILES Matthew et HUBERMAN Michael, *Analyse des données qualitatives : Recueil de nouvelles méthodes*, Bruxelles, De Boeck, 2003.

MILGRAM Stanley et TRAVERS Jeffrey, « An Experimental Study of the Small World Problem », *Sociometry*, 1969, vol.32, No. 4. (1), p. 425-443

MONNOYER-SMITH Laurence, « La participation en ligne, révélateur d'une évolution des pratiques politiques ? », *Participations*, 2011/1 (n° 1), p. 156-185.

MUXEL Anne, *L'expérience politique des jeunes*, Paris, Presses de Science Po, 2001.

NEIJENS Peter, DE RIDDER Jan et SARIS Willem, « An Instrument for Collecting Informed », *Opinions Quality and Quantity*, n°26, 1992, p. 245-258.

NORRIS Pippa, « The Gender Generation Gap in British Election », *British Elections and Parties Yearbook*, vol. 3, 1993, p. 129-142.

NORRIS Pippa, *Digital Divide: Civic engagement, information poverty, and the internet worldwide*, Cambridge, Cambridge University Press, 2001.

NORRIS Pippa et CURTICE John, « Getting the Message Out: A two-step model of the role of the Internet in campaign communication flows during the 2005 British General Election », *Paper presented at the annual meeting of the American Political Science Association*, Marriott, Loews Philadelphia, and the Pennsylvania Convention Center, Philadelphia, PA, Aug 31, 2006.

O'CONNOR Brendan, BALASUBRAMANYAN Ramnath, ROUTLEDGE Bryan et SMITH Noah, « From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series », *Proceedings of International AAAI Conference on Weblogs and Social Media*, Washington, 2010, p. 1-8.

OGER Claire, *Le façonnage des élites de la République. Culture générale et haute fonction publique*, Paris, Presses de Sciences Po, 2008.

OUAKRAT Alan, « La consommation de l'information en ligne : un cadrage quantitatif », in Josiane Jouët et Rémy Rieffel (dir.), *S'informer à l'ère numérique*, Rennes, Presses Universitaires de Rennes, 2013, p.69.

PAK Alexander et PAROUBEK Patrick, « Twitter as a Corpus for Sentiment Analysis and Opinion Mining », *Proceedings of the Seventh conference on International Language Resources and Evaluation LREC'10*, Valletta, Malta, European Language Resources Association ELRA, 2010.

PANG Bo et LEE Lillian, « Opinion Mining and Sentiment Analysis », *Foundations and Trends in Information Retrieval*, vol. 2, n°1-2, p. 1-135.

PAPA Françoise et FRANCONY Jean-Marc, « Twitter en campagne », in Jacques Gerstlé et Raul Magni Berton (dir.), *2012, la campagne présidentielle*, Paris, L'Harmattan, 2013, p. 87-102.

PAPACHARISSI Zizi, « The Virtual Sphere. The Internet as a Public Sphere », *New Media & Society*, vol. 4(1), 2002, p. 9-27.

PARAISIE Sylvain et COINTET Jean-Philippe, « La presse en ligne au service de la démocratie locale. Une analyse morphologique de forums politiques », *Revue Française de Science Politique*, 2012/1 (vol. 62).

PARAISIE Sylvain et DAGIRAL Éric, « Le site et son audience. Enquêter sur les “métriques du web” », in Christine Barats (dir.), *Manuel d'analyse du web*, Paris, Armand Colin, 2013.

PASSERON Jean-Claude, « Symbolisme dominant et symbolisme dominé », *Enquête*, 1985, mis en ligne le 16 novembre 2005, consulté le 05 mai 2014. URL : <http://enquete.revues.org/38>

PERRINEAU Pascale, *Le vote normal. Les élections présidentielle et législatives d'avril-mai-juin 2012*, Paris, Presses de Sciences Po, 2013.

PEW RESEARCH CENTER, « Twitter News Consumers: Young, Mobile and Educated », novembre 2013.

PHELAN Owen, MC CARTHY Kevin et SMYTH Barry, « Using twitter to recommend real-time topical news », *Proceedings of the third ACM conference on Recommender systems*, 2009, p. 385-388.

PINÇON-CHARLOT Monique et PINÇON Michel, *Voyage en grande bourgeoisie*, Paris, PUF, 1997.

PINSON Gilles, « Peut-on vraiment se passer de l'entretien en sociologie de l'action publique ? », *Revue Française de Science politique*, 2007/5 (vol. 57), p. 555-597.

PLANTIN Jean-Christophe, « D'une carte à l'autre, Le potentiel heuristique de la comparaison entre graphe du web et carte géographique », in Christine Barats (dir.), *Manuel d'analyse du web*, Paris, Armand Colin, 2013.

REBILLARD Franck et SMYRNAIOS Nikos, « Les infomédiaires, au cœur de la filière de l'information en ligne. Les cas de Google, Wikio et Paperblog », *Réseaux*, vol. 2-3, n°160-161, 2010, p. 163-194.

REBILLARD Franck, *Le Web 2.0 en perspective : Une analyse socio-économique de l'Internet*, L'Harmattan, 2007.

REPERES ET REFERENCES STATISTIQUES SUR LES ENSEIGNEMENTS, LA FORMATION ET LA RECHERCHE. STATISTIQUES - publications annuelles - Édition 2012 http://media.education.gouv.fr/file/2012/36/9/DEPP-RERS-2012_223369.pdf

REY Henri, « Les adhérents socialistes : permanences et changements », *Revue socialiste*, n°42, 2011, p. 125-135.

RICHARD-ZAPPELLA Jeannine, « La mise en mots du référent dans la construction de l'opinion publique : les questions en question », *Langage et société*, n°55, p. 41-59.

RIEDER Bernhard, « Pratiques informationnelles et analyse des traces numériques : de la représentation à l'intervention », *Études de communication*, n°35, 2010, p. 91-104.

RIEDER Bernhard et SMYRNAIOS Nikos, « Pluralisme et infomédiation sociale de l'actualité : le cas de Twitter », *Réseaux*, vol. 6, n°176, 2012, p. 105-139.

RITTERMAN Joshua, OSBORNE Miles et KLEIN Ewan, « Using Prediction Markets and Twitter to Predict a Swine Flu Pandemic », *Proceedings Title 1st International Workshop on Mining Social Media*, 2009.

ROZENBERG Olivier et BACKMAN François, « Mortels transferts ? La pratique des entretiens collectifs par les Grands Instituts de Sondages Internationaux et par la recherche en Science Sociale », *Focus groupes, entretiens collectifs : état des lieux - Journée d'étude organisée dans le cadre du groupe MOD de l'AFSP* - 8 avril 2009.

SAUGER Niolas, *The French Electoral Study 2012*, données disponibles sur <http://cee.sciences-po.fr/fr/elections-2012/enquete-electorale-francaise-2012.html>.

SAWICKI Frédéric et SIMEANT Johanna, « Décloisonner la sociologie de l'engagement militant. Note critique sur quelques tendances récentes des travaux français », *Sociologie du travail*, n°51, 2009, p. 97-125.

SCHLOZMAN Kay, VERBA Sydney et BRADY Henry, « Weapon of the Strong ? Participatory Inequality and the Internet », *Perspectives on Politics*, 8(2), 2010. p. 487-459.

SCHMITZ Andreas, « Elective Affinities 2.0? A Bourdieusian Approach to Couple Formation and the Methodology of E-Dating », *Reset*, n°1, 2012, p. 176-202.

SCHNEIDER Steven, « Creating a Democratic Public Sphere Through Political Discussion: A Case Study of Abortion Conversation on the Internet », *Social Science Computer Review*, vol. 14, n° 4, 1996, p. 373-393.

SILVERMAN David, *Interpreting Qualitative Data*, London, Sage, 1993.

SMITH Ted, COYLE James, LIGHTFOOT Elizabeth et SCOTT Amy, « Reconsidering models of influence: the relationship between consumer social networks and word-of-mouth effectiveness », *Journal of Advertising Research*, décembre 2007, p. 387-397.

SNIDERMAN Paul, « The new look in public opinion research », in Ira Katznelson et Helen V. Milner (dir.), *Political science: the state of the discipline*, vol. 2, Washington, American Political Science Association, 1993, p. 193-218.

SNIDERMAN Paul, « Les nouvelles perspectives de la recherche sur l'opinion publique », *Politix*, vol. 11, n°141, 1998, p. 123-175.

STROMER-GALLEY Jenifer, « New voices in the public sphere: a comparative analysis of interpersonal and online political talk », *The public*, vol. 9, n°2, 2002, p. 23-42.

STROMER-GALLEY Jenifer, « Diversity of Political Conversation on the Internet: Users Perspectives », *Journal of Computed-mediated communication*, 8(3), 2003.

STROMER-GALLEY Jenifer, « Decoding deliberation online », *Paper presented at the Second Conference on Online Deliberation*, Palo Alto, Californie, mai 2005.

SUH Bongwon, HONG Lichan, PIROLI Peter et CHI Ed, « Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network », *IEEE Second International Conference on Social Computing (SocialCom)*, Palo Alto, 2010, p. 177-184.

SUNSTEIN Cass, *Republic.com*, Princeton, Princeton University Press, 2001.

SUNSTEIN Cass, *Republic.com 2.0*, Princeton, Princeton University Press, 2007.

TALPIN Julien, « Qualité de la délibération », in Ilaria Casillo avec Rémi Barbier, Loïc Blondiaux, Francis Chateauraynaud, Jean-Michel Fourniau, Rémi Lefebvre, Catherine Neveu et Denis Salles (dir.), *Dictionnaire critique et interdisciplinaire de la participation*, Paris, GIS Démocratie et Participation, 2013.

TCHUENTE Dieudonné, CANUT Marie-Françoise, BAPTISTE-JESSEL Nadine, COUTANT Alexandre, STENGER Thomas et RAMPNOUX Olivier, « Pour une approche interdisciplinaire des TIC. Le cas des réseaux socionumériques », *Document numérique*, 2011/1 (vol. 14).

THELWALL Mike, « Blog searching. The first general-purpose source of retrospective public opinion in the social sciences? », *Online Information Review*, vol. 31, n°3, 2007, p. 227-289.

THEVIOT Anaïs, « Qui milite sur Internet ? Esquisse du profil sociologique du « cybermilitant » au PS et à l'UMP », *Revue Française de Science Politique*, 2013/3-4 (vol.63), p. 663-678.

THEVIOT Anaïs, « Twitter en regardant la télévision : une campagne transmédiatique interactive ? Analyse comparée des stratégies numériques au Parti Socialiste et à l'Union pour un Mouvement Populaire lors des ripostes-party », *Télévision*, 5 (2014), p. 95-112.

TSAGARROUSIANOU Roza, « Electronic democracy in practice... One, two, three... countless variants », *Hermès*, 26-27, 2000, p. 233-246.

TUMASJAN Andranik, SPRENGER Timm, SANDNER Philipp et WELPE Isabell, « Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment », *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 2010, Menlo Park, CA: The AAAI Press, p. 178-185.

TURKLE Sherry, *Life on the Screen. Identity in the Age of Internet*, New York, Simon & Shuster, 1995.

VEDEL Thierry, « Internet et les pratiques politiques », in Anne-Marie Gingras (dir.), *La communication politique. Etat des savoirs, enjeux et perspectives*, Québec, Presses de l'Université du Québec, 2003, p. 189-214.

VEDEL Thierry, « L'idée de démocratie électronique : origines, visions, questions », in Pascal Perrineau (dir.), *Le désenchantement démocratique*, Paris, Editions de l'Aube, 2003, p. 243-266.

VEDEL Thierry, « La Révolution en sera plus télévisée. Internet, information et démocratie », *Pouvoirs*, n°119, 2006, p. 41-54.

VEDEL Thierry, « Les blogs et la politique : la démocratie en kit ? », in Xavier Greffe et Nathalie Sonnac (dir.), *Culture Web : Création, contenus, économie numérique*, Paris, Éditions Dalloz, 2008, p. 61-77.

VITIELLO Thomas, « Les systèmes d'information et d'aide au vote : le cas de La Boussole présidentielle », in Jacques Gerstlé et Raul Magni Berton (dir.), 2012, *la campagne présidentielle*, Paris, L'Harmattan, 2013, p. 137-151.

WATTS Duncan et STROGATZ Steven, « Collective Dynamics of “Small-World” Networks », *Nature*, n° 393, 1998, p. 440-442.

WATTS Duncan, *Small Worlds: The Dynamics of Networks Between Order and Randomness*, Princeton, Princeton University Press, 1999.

WATTS Duncan, « The “new” science of networks », *Annual Review of Sociology*, n° 30, 2004, p. 243-70.

WATTS Duncan et DODDS Peter, « Influentials, Networks and Public Opinion Formation », *Journal of Consumer Research*, 34(4), 2007, p. 441-458.

« WEB 2.0 », *Réseaux*, vol. 27, n° 154, 2009.

WIEBE Janyce, WILSON Theresa et BELL Matthew, « Identifying collocations for recognizing opinions », *Proc. of the ACL/EACL Workshop on Collocation*, 2001.

WIEBE Janyce, BRUCE Rebecca et O'HARA Thomas, « Development and Use of a Gold-Standard Data Set for Subjectivity Classifications », *Proceedings of the Association for Computational Linguistics (ACL)*, vol. 37, 1999, p. 246-253.

WILSON Theresa, WIEBE Janyce et HOFFMAN Paul, « Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis », *Proceedings of Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*.

WITSCHGE Tamara, « Online Deliberation: Possibilities of the Internet for Deliberative Democracy », Peter M. Shane (dir.), *Democracy Online. The Prospects for Political Renewal through the Internet*, New York, Routledge, 2004, p. 109-122.

WOJCIK Stéphanie, « Prendre au sérieux la démocratie électronique. De quelques enjeux et controverses sur la participation politique en ligne », in Elsa Forey et Christophe Geslot (dir.), *La démocratie électronique*, L'Harmattan, coll. « Questions contemporaines », 2011.

WOLTON Dominique, *Internet et après ? Une théorie critique des nouveaux médias*, Flammarion, coll. « Champs », 2000.

WRIGHT Scott, « From “third place” to “third space”: everyday political in non-political online spaces », *javnost-the public*, vol. 19, n°3, 2012, p. 5-50.

WU Shaomei, HOFMAN Jake, MASON Winter et WATTS Duncan, « Who says what to whom on Twitter », *Proceedings of the 20th international conference on World wide web*, Hyderabad, India, 2011, p. 705-714.

YOUNG Marion, « Communication and the Other: Beyond Deliberative Democracy », in Seyla Benhabib (dir.), *Democracy and Difference. Contesting the Boundaries of the Political*, Princeton, Princeton University Press, 1996, p. 120-135.

ZALLER John, *The Nature and Origins of Mass Opinion*, Cambridge, Cambridge University Press, 1992.

ZHAO Dejin et ROSSON Mary Beth, « How and why people Twitter: the role that micro-blogging plays in informal communication at work », *Proceedings of the ACM 2009 international conference on Supporting group work*, 2009, p. 243-252.

Table des illustrations

Introduction

Illustration 0.1 Représentation sous forme textuelle de données quantitatives	p. 19
Illustration 0.2 Cartographie d'un ensemble d'inscrits au réseau social Twitter	p. 21
Tableau 0.1 Les avantages et limites des analyses de <i>web-tracking</i> et leur contournement	p. 39

Chapitre 1

Graphique 1.1 Évolution du nombre d'internautes en France de 1995 à 2011	p. 53
Tableau 1.1 Taux de pénétration d'internet dans les foyers français par PCS (en %)	p. 54
Tableau 1.2 Taux d'inscription à Facebook et Twitter (en %) par PCS	p. 56
Illustration 1.1 Cartographie de la blogosphère politique en 2007	p. 82

Chapitre 2

Tableau 2.1 Activité des comptes soumis à l'analyse d'éligibilité au panel	p. 199
Tableau 2.2 Langue des comptes soumis à l'analyse d'éligibilité au panel	p. 200
Tableau 2.3 Nature des comptes soumis à l'analyse d'éligibilité au panel	p. 201
Tableau 2.4 Date du dernier tweet publié par les individus éligibles au panel	p. 201
Tableau 2.5 Taux de remplissage des informations socio-démographiques des non-répondants	p. 205

Chapitre 3

Tableau 3.1 Répartition par classe d'âge (en %) des populations « répondants » et « non répondants »	p. 213
Tableau 3.2 Répartition par classe d'âge (en %) des « répondants » et des inscrits sur Twitter	p. 215
Tableau 3.3 Taux de participation au premier tour des élections présidentielles 2012 par	p. 216

tranche d'âge (en %)

Tableau 3.4 Situation de l'emploi des jeunes de 15 à 24 ans (en %)	p. 218
Tableau 3.5 Répartition (en %) des étudiants en fonction du type d'établissement en 2011-2012	p. 219
Tableau 3.6 Répartition (en %) des étudiants selon leur origine sociale	p. 220
Tableau 3.7 Répartition (en %) des individus panélisés selon leur origine sociale	p. 221
Tableau 3.8 Répartition (en %) des panélisés selon la PCS	p. 223
Tableau 3.9 Distribution de la population selon le sexe (en %)	p. 230
Tableau 3.10 Distribution de la population « répondants » selon le sexe et l'âge (en %)	p. 231
Tableau 3.11 Distribution de la population « répondants » selon le sexe et la PCS (en %)	p. 232
Tableau 3.12 Distribution des populations répondants et non répondants selon le lieu de résidence (en %)	p. 233
Tableau 3.13 Distribution des répondants selon la PCS et le lieu de résidence (en %)	p. 235
Tableau 3.14 Distribution de la population répondants selon le niveau de diplôme (en %)	p. 237
Tableau 3.15 Distribution des populations âgées de 25 à 49 ans selon le niveau de diplôme (en %)	p. 238
Tableau 3.16 Distribution de la population répondants selon l'intérêt déclaré pour la politique (en %)	p. 240
Tableau 3.17 Pratiques politiques des répondants	p. 243
Tableau 3.18 Vote au premier tour des élections présidentielles du 22 avril 2012	p. 246
Illustration 3.1 Dernière étape de l'inscription à Twitter	p. 255
Tableau 3.19 Distribution des répondants selon l'âge, le niveau scolaire et le fait d'avoir rédigé ou non sa biographie	p. 257
Tableau 3.20 Répartition (en %, n = 460) des « biographies » selon leur forme et leur contenu	p. 259
Tableau 3.21 Type de « biographie » en fonction de l'âge (en %, n = 460)	p. 263
Tableau 3.22 Exemples de biographies stylisées	p. 264

Chapitre 4

Tableau 4.1 Comparaison statistique de la production des deux panels	p. 284
Tableau 4.2 Quartiles de la variable « nombre de tweets politiques » selon la population étudiée	p. 287
Tableau 4.3 Test de corrélation de Pearson entre l'âge et le nombre de tweets politiques publiés par les « répondants »	p. 292
Tableau 4.4 Indices de dispersion et de tendance centrale du nombre de tweets politiques par classes d'âge	p. 293
Tableau 4.5 ANOVA nombre de tweets publiés et PCS	p. 294
Tableau 4.6 Test de corrélation de Spearman entre l'intérêt déclaré pour la politique et le nombre de tweets politiques publiés	p. 296
Tableau 4.7 Comparaison du nombre de tweets politiques publiés entre militants et non militants	p. 296
Tableau 4.8 Test de corrélation de Pearson entre le nombre de tweets et le nombre de tweets politiques	p. 298
Tableau 4.9 Test de corrélation de Pearson entre le nombre de tweets et le nombre d'abonnés	p. 298
Tableau 4.10 Test de corrélation de Pearson entre le nombre de tweets et le nombre d'abonnés	p. 299
Tableau 4.11 Intérêt déclaré pour la politique et nombre d'abonnements à des médias et hommes politiques	p. 300
Tableau 4.12 Analyse de régression linéaire du nombre de tweets politiques publiés par les « répondants »	p. 301
Tableau 4.13 Production quotidienne des panels « répondants » et « non répondants »	p. 305
Graphique 4.1 Distribution journalière des tweets politiques publiés par les panels « répondants » et « non répondants »	p. 306
Tableau 4.14 Volume et taux de production quotidiens les plus élevés des panels	p. 307

Tableau 4.15 Production de tweets à l’occasion d’« événements » politiques	p. 308
Graphique 4.2 Évolution quotidienne du taux de production des « répondants » selon leur intérêt déclaré pour la politique	p. 312
Tableau 4.16 Origine du compte retweeté (n = 576 tweets)	p. 317
Tableau 4.17 Nature des sites vers lesquels renvoient les liens hypertextes	p. 318
Tableau 4.18 Objet des tweets politiques publiés par le panel « répondants » (n = 1 108 tweets)	p. 318
Graphique 4.3 Distribution du nombre d’abonnés aux comptes producteurs de tweets politiques	p. 321
Tableau 4.19 Catégories des 1041 et des 200 comptes du panel « Twitter audience politique » connaissant la plus forte audience sur le réseau	p. 323
Tableau 4.20 Les vingt comptes du panel « Twitter audience politique » bénéficiant de la plus forte audience sur le réseau	p. 325
Tableau 4.21 Corrélation entre le nombre de tweets politiques publiés par les panels « Répondants » et « Audience politique » (n = 92 jours)	p. 326
Graphique 4.4 Évolution quotidienne des tweets politiques publiés par les panels « Répondants » et « Audience politique » du 01/03/2014 au 31/05/2014	p. 327
Tableau 4.22 Occurrences politiques les plus courantes contenues dans les tweets politiques publiés par les panels	p. 328

Chapitre 5

Tableau 5.1 Comptes auxquels les membres des panels sont le plus souvent abonnés	p. 337
Tableau 5.2 Comptes bénéficiant du plus important nombre d’abonnés sur Twitter	p. 338
Tableau 5.3 Comptes « exclusifs » auxquels sont abonnés les panels	p. 340
Tableau 5.4 Comparaison des comptes du panel « Twitter audience politique » et des abonnements des « répondants »	p. 341
Tableau 5.5 Taux d’abonnement (en %) des « répondants » à des comptes de médias en	p. 343

fonction de leur orientation politique déclarée (n = 558)

Tableau 5.6 Nombre moyen d'abonnements à des médias politiques selon l'orientation politique de l'enquêté p. 345

Tableau 5.7 Support originel des comptes de médias auxquels les « répondants » sont le plus souvent abonnés p. 347

Tableau 5.8 Origine des comptes médiatiques majoritairement privilégiés selon la PCS (en %) p. 349

Tableau 5.9 Taux d'abonnement (en %) à des médias par PCS p. 351

Tableau 5.10 Nombre d'abonnement moyen à des comptes d'hommes politiques p. 356

Tableau 5.11 Taux et nombre de retweets moyens provenant de médias, de journalistes et d'hommes politiques p. 358

Tableau 5.12 Nombre de tweets publiés contenant des liens vers des sites d'information en ligne par PCS p. 361

Tableau 5.13 Taux (en %) et nombre de retweets moyens provenant de comptes de médias, de journalistes et d'hommes politiques p. 363

Tableau 5.14 Destinataires des tweets politiques publiés (n = 1 108tweets) p. 368

Tableau 5.15 Contexte d'émission du tweet (n = 1 108tweets) p. 369

Tableau 5.16 Registre de discours des tweets politiques (n = 1 108tweets) p. 372

Tableau 5.17 Nature des objets évoqués dans les tweets politiques (n = 1 108tweets) p. 373

Tableau 5.18 Format des tweets politiques publiés (n = 1 106) p. 377

Conclusion

Illustration 6.1 Cartographie des retweets concernant le mariage homosexuel le 7 mars p. 405

Illustration 6.2 Cartographie des retweets concernant le mariage homosexuel le 22 avril p. 407

Illustration 6.3 « Variantes du goût petit-bourgeois » (graphiques 15 et 16 de La Distinction, p.392) p. 411

Annexes

Annexe 1. Glossaire	p. 464
Annexe 2. Le questionnaire	p. 467
Annexe 3. Caractéristiques générales de la population « répondants »	p. 475
Annexe 4. Caractéristiques générales de la population « non répondants »	p. 482
Annexe 5. Caractéristiques générales du panel « Twitter audience politique »	p. 485
Annexe 6. Présentation du logiciel <i>AMI Opinion Tracker</i>	p. 487
Annexe 7. Requête de sélection des tweets politiques	p. 494
Annexe 8. Grille de codage des tweets politiques	p. 496

Annexe 1. Glossaire

Abonné / (s') abonner ou « follower » : Être abonné ou « suivre » (to *follow* en anglais) un compte Twitter permet à un utilisateur de recevoir dans sa *timeline* les tweets publiés par ce compte.

Abonnement : Compte Twitter auquel un utilisateur est abonné.

Application Programming Interface (API) : Interface de programmation d'application en français. Une API est un service informatique développé par une société (ou un développeur) à destination d'autres développeurs. Twitter propose ainsi des API permettant aux logiciels de *web-tracking* de collecter les données issues de son site internet.

Avatar : Image ou photographie personnelle associée à un profil Twitter.

Biographie : Description de profil d'une taille maximum de 160 caractères présentant en quelques mots le titulaire du compte.

Blog : « Abréviation de l'anglais *weblog*, *carnet de bord sur le Web*. Site Web sur lequel un internaute tient une chronique personnelle ou consacrée à un sujet particulier », (source : Larousse, 2014).

Blogueur : Auteur d'un blog.

Lien hypertexte : (ou hyperlien) : « Référence dans un système hypertexte permettant de passer automatiquement d'un document consulté à un document lié. Les hyperliens sont notamment utilisés dans le World Wide Web pour permettre le passage d'une page Web à une autre à l'aide d'un clic », (source : <http://fr.wikipedia.org/wiki/Hyperlien>).

Like : Sur Facebook, action d'« aimer » la page d'une institution, d'une entreprise, d'une organisation, d'une marque ou d'une personnalité publique (artiste, journaliste, etc.). « *Liker* » permet à un utilisateur d'avoir accès au contenu de la page et d'être tenu informé de son actualisation.

Live-tweeter : Émettre des tweets afin de commenter en direct la tenue d'un événement (compétition sportive, *meeting* électoral, etc.) ou d'une émission télévisée.

Micro-blogging : Service en ligne de publication de très courts messages textuels. Twitter est considéré comme un service (ou site) de *micro-blogging*.

PageRank : « *Algorithme d'analyse des liens concourant au système de classement des pages Web utilisé par le moteur de recherche Google. Il mesure quantitativement la popularité d'une page web* », (source : <http://fr.wikipedia.org/wiki/PageRank>).

Profil Twitter : « *Page Twitter contenant les informations d'un utilisateur, ainsi que les Tweets envoyés depuis [son] compte* », (source : <https://support.twitter.com/articles/515487-glossaire-twitter>).

Pureplayer : (« tout en ligne » en français). Entreprise ou titre de presse exclusivement disponible sur internet. *Mediapart* et *Rue89* sont des *pureplayer*. *Lemonde.fr* n'est pas un *pureplayer*.

Hashtag ou mot-dièse : « *Correspond à un terme précédé du symbole #. Sert à recenser les Tweets autour d'un thème commun et à créer une conversation. Un hashtag transforme aussi un terme en lien de recherche* », (source : <https://support.twitter.com/articles/515487-glossaire-twitter>).

Retweet : Publication par un utilisateur sur son profil d'un tweet publié par un autre utilisateur.

Retweeter : Action d'émettre un retweet, c'est-à-dire de republier à l'identique un message publié originellement par un autre utilisateur.

Source de type count : Source logicielle collectant l'ensemble des tweets publiés par un compte donné.

Source de type search : Source logicielle collectant tous les tweets publiés sur le réseau social contenant une occurrence définie par le créateur de la source.

Timeline : Espace d'édition où défilent en temps réel et selon un ordre antéchronologique les tweets publiés par l'ensemble des comptes auxquels un utilisateur est abonné.

Tweet : Message d'une taille maximum de 140 caractères publié sur le réseau social Twitter.

Tweeter : Action de publier un tweet.

Twitto : Individu inscrit sur Twitter.

Uniform Resource Locator (URL) : adresse permettant d'identifier et d'accéder à une page internet.

Widget : Contraction des termes *window* (fenêtre en anglais) et *gadget*. Interface graphique permettant de figurer des informations et de les actualiser en temps réel.

Web-crawler : (Robot indexation du Web en français). Logiciel qui explore automatiquement le Web afin de collecter différents types de ressources (dont des autopublications numériques), afin de permettre à un moteur de recherche de les indexer.

Web-tracker : Logiciel permettant de détecter et d'archiver des échanges ou des messages publiés sur les réseaux sociaux.

@réponse : « *Un Tweet commençant par le nom d'un autre utilisateur et qui est en réponse à l'un de ses Tweets* », (source : <https://support.twitter.com/articles/515487-glossaire-twitter>).

@mention : « *Un Tweet contenant le nom d'utilisateur Twitter d'un autre utilisateur, précédé du symbole "@"* », (source : <https://support.twitter.com/articles/515487-glossaire-twitter>).

Annexe 2. Le questionnaire

Bonjour, nous menons un projet de recherche universitaire afin de mieux connaître la population des twittos. Merci de nous aider dans notre démarche en répondant aux questions suivantes :

Q1. Vous êtes :

1. Un homme
2. Une femme

Q2. Quel âge avez-vous ? _____ ans

Q3. Vous habitez dans une agglomération de :

1. Moins de 2 000 habitants
2. De 2 000 à 20 000 habitants
3. De 20 000 à 100 000 habitants
4. De 100 000 à 500 000 habitants
5. De plus de 500 000 habitants

Q4. Merci d'indiquer le code postal à 5 chiffres de votre commune de résidence (exemple : si vous habitez le 13^{ème} arrondissement de Paris : 75013 ; si vous habitez hors de France, merci d'indiquer votre code postal sans tenir compte de la restriction à 5 chiffres, ou, à défaut, merci de rentrer le code : 00000) _____

Q5. Quel est le dernier diplôme que vous ayez obtenu ?

1. Je ne possède aucun diplôme
2. Certificat d'études
3. Brevet des collèges ou BEPC
4. CAP ou BEP
5. Baccalauréat d'enseignement professionnel
6. Baccalauréat technologique (STI, STG, ST2S, etc.)
7. Baccalauréat général (ES, S, L)
8. Deug, BTS ou équivalent Bac+2

9. Licence ou équivalent Bac+3
10. Maîtrise ou équivalent Bac+4
11. Master (ex. DEA ou DESS) ou équivalent Bac+5
12. Doctorat

Q6. Quel est, ou a été, le dernier établissement d'enseignement supérieur que vous ayez fréquenté ? (si vous n'avez pas suivi d'études supérieures, merci de cocher la réponse 1)

1. Je n'ai pas suivi d'études supérieures
2. Universités (hors IUT, IUFM et formations d'ingénieurs)
3. Instituts universitaires de formation des maîtres (IUFM)
4. Instituts universitaires de technologie (IUT)
5. Classes préparatoires aux grandes écoles (CPGE)
6. Sections de techniciens supérieurs (STS)
7. Ecoles d'ingénieurs
8. Ecoles de commerce
9. Ecoles d'architecture
10. Ecoles supérieures artistiques et culturelles
11. Ecoles paramédicales et sociales hors université
12. Ecoles de journalisme
13. Instituts d'Etudes Politiques
14. Ecoles Normales Supérieures
15. Autre école

Q7. Quel est votre statut actuel ?

1. Lycéen,
2. Étudiant ou étudiant stagiaire
3. Travailleur intérimaire
4. Salarié en CDD du public
5. Fonctionnaire ou salarié en CDI du public
6. Salarié en CDD du privé
7. Salarié en CDI du privé
8. Travailleurs indépendants, commerçant, artisan
9. Professions libérales
10. Chef d'entreprise

11. Femme/homme au foyer
12. Chômeur
13. Retraité

Q8. Vous exercez votre profession : (si vous ne travaillez pas, merci de cocher la réponse 3)

1. à temps plein
2. à temps partiel
3. vous ne travaillez pas

Q9. Dans quelle catégorie de revenus **mensuels** vous situeriez-vous (salaires + indemnités + allocations + autres revenus)?

1. Moins de 750 euros par mois
2. Entre 750 et 1100 euros par mois
3. Entre 1100 et 1300 euros par mois
4. Entre 1300 et 1500 euros par mois
5. Entre 1500 et 1900 euros par mois
6. Entre 1900 et 2500 euros par mois
7. Entre 2500 et 3100 euros par mois
8. Entre 3100 et 3700 euros par mois
9. Plus de 3700 euros par mois

Q10. Pourriez-vous s'il vous plaît, indiquer très précisément votre profession actuelle ou, si vous êtes retraité ou à la recherche d'un emploi, la dernière profession que vous avez exercée :

Q11. Dans quelle catégorie socioprofessionnelle vous situeriez-vous (si vous êtes lycéen ou étudiant non-salarié, merci de cocher la réponse 24 : inactifs divers) ?

1. Agriculteurs exploitants
2. Artisans
3. Commerçants et assimilés
4. Chefs d'entreprise de 10 salariés ou plus
5. Professions libérales et assimilés
6. Cadres de la fonction publique, professions intellectuelles et artistiques
7. Cadres d'entreprise
8. Professions intermédiaires de l'enseignement, de la santé, de la fonction publique et assimilés
9. Professions intermédiaires administratives et commerciales des entreprises
10. Techniciens
11. Contremaîtres, agents de maîtrise
12. Employés de la fonction publique
13. Employés administratifs d'entreprise
14. Employés de commerce
15. Personnels des services directs aux particuliers
16. Ouvriers qualifiés
17. Ouvriers non qualifiés
18. Ouvriers agricoles
19. Anciens agriculteurs exploitants
20. Anciens artisans, commerçants, chefs d'entreprise
21. Anciens cadres et professions intermédiaires
22. Anciens employés et ouvriers
23. Chômeurs n'ayant jamais travaillé
24. Inactifs divers (autres que retraités)

Q12. A quelle catégorie socioprofessionnelle appartient ou a appartenu (avant qu'elle ne soit à la retraite) votre mère?

1. Agriculteurs exploitants
2. Artisans
3. Commerçants et assimilés

4. Chefs d'entreprise de 10 salariés ou plus
5. Professions libérales et assimilés
6. Cadres de la fonction publique, professions intellectuelles et artistiques
7. Cadres d'entreprise
8. Professions intermédiaires de l'enseignement, de la santé, de la fonction publique et assimilés
9. Professions intermédiaires administratives et commerciales des entreprises
10. Techniciens
11. Contremaîtres, agents de maîtrise
12. Employés de la fonction publique
13. Employés administratifs d'entreprise
14. Employés de commerce
15. Personnels des services directs aux particuliers
16. Ouvriers qualifiés
17. Ouvriers non qualifiés
18. Ouvriers agricoles
19. Inactifs divers (femme au foyer)

Q13. A quelle catégorie socioprofessionnelle appartient ou a appartenu (avant qu'il ne soit à la retraite) votre père?

1. Agriculteurs exploitants
2. Artisans
3. Commerçants et assimilés
4. Chefs d'entreprise de 10 salariés ou plus
5. Professions libérales et assimilés
6. Cadres de la fonction publique, professions intellectuelles et artistiques
7. Cadres d'entreprise
8. Professions intermédiaires de l'enseignement, de la santé, de la fonction publique et assimilés
9. Professions intermédiaires administratives et commerciales des entreprises
10. Techniciens
11. Contremaîtres, agents de maîtrise
12. Employés de la fonction publique
13. Employés administratifs d'entreprise

14. Employés de commerce
15. Personnels des services directs aux particuliers
16. Ouvriers qualifiés
17. Ouvriers non qualifiés
18. Ouvriers agricoles
19. Inactifs divers

Q14. Actuellement, quelle est votre situation personnelle ?

1. Vous êtes célibataire
2. Vous êtes marié(e)
3. Vous êtes pacsé(e)
4. Vous êtes en couple sans être marié(e) ni pacsé(e)
5. Vous êtes veuf(ve)
6. Vous êtes divorcé(e)

Q15. D'une façon générale, vous diriez que vous vous intéressez à la politique :

1. Beaucoup
2. Assez
3. Un peu
4. Pas du tout

Q16. Sur le plan politique, vous vous considérez comme étant :

1. Très à gauche
2. Plutôt à gauche
3. Au centre
4. Plutôt à droite
5. Très à droite
6. Ni à droite, ni à gauche
7. Vous ne savez pas
8. Autre

Q17. De quel parti politique vous sentez-vous le plus proche ou disons le moins éloigné ?

1. Lutte ouvrière (LO)
2. Nouveau parti anticapitaliste (NPA)

3. Parti ouvrier indépendant (POI)
4. Parti communiste français (PCF)
5. Parti de gauche (PG)
6. Mouvement républicain et citoyen (MRC)
7. Parti socialiste (PS)
8. Parti radical de gauche (PRG)
9. Europe Ecologie Les Verts (EELV)
10. Alliance écologiste indépendante (AEI)
11. Cap 21
12. Mouvement démocrate (MODEM)
13. Nouveau centre (NC)
14. Union pour un mouvement populaire (UMP)
15. République solidaire (RS)
16. Debout la République (DLR)
17. Parti chrétien-démocrate (PCD)
18. Front National (FN)
19. Mouvement National Républicain (MNR)
20. Solidarité et Progrès
21. Autre (*précisez svp*)_____

Q18. Jusqu'à présent, avez-vous déjà : (*Plusieurs réponses possibles*)

1. Voté
2. Signé une ou des pétitions
3. Manifesté
4. Fait grève
5. Adhéré à un syndicat
6. Adhéré à un parti politique
7. Adhéré à une ou plusieurs association(s)

Q19. A laquelle de ces élections avez-vous déjà voté ? (plusieurs réponses possibles. Si vous n'avez voté à aucune d'entre elles ou si vous n'avez pas le droit de vote ou la nationalité française, merci de cocher la réponse 8)

1. Présidentielle 2007
2. Législatives 2007
3. Municipales 2008

4. Européennes 2009
5. Régionales 2010
6. Cantonales (2008 ou 2011)
7. Primaires citoyennes 2011 (organisées par le Parti socialiste)
8. Je n'ai voté à aucune de ces élections

Q20. Pour quel candidat envisagez-vous de voter en 2012 ? (si vous n'avez pas le droit de vote ou si vous n'avez pas la nationalité française, merci de cocher la réponse 1)

1. Je n'envisage pas de voter en 2012
2. J'envisage de voter blanc ou nul
3. Nathalie Arthaud (LO)
4. Philippe Poutou (NPA)
5. Jean-Luc Mélenchon (Front de gauche)
6. François Hollande (PS)
7. Eva Joly (EELV)
8. François Bayrou (MODEM)
9. Nicolas Sarkozy (UMP)
10. Nicolas Dupont-Aignan (DLR)
11. Marine Le Pen (FN)
12. Jacques Cheminade (Solidarité et Progrès)

Q21. Enfin, acceptez-vous de faire partie d'un panel de Twittos ? En acceptant, vous vous engagez à ce que l'ensemble de vos tweets soient susceptibles d'être archivés et analysés par notre équipe de chercheurs. Néanmoins, toutes vos données seront **entièrement anonymisées**, et ne feront l'objet que de traitements statistiques. Nous nous engageons à ne **jamais divulguer ni vendre** ces données. Vous pourrez de plus quitter le panel à tout moment, en nous envoyant simplement un courrier électronique à l'adresse suivante : imagiweb@eric.univ-lyon2.fr. Si vous avez des questions ou des demandes d'information, n'hésitez pas à nous les envoyer à cette même adresse.

1. Oui j'accepte de faire partie du « Panel Twittos » et j'accepte que tous mes tweets passés et futurs soient archivés et analysés dans le cadre d'un projet de recherche universitaire.
2. Non, je refuse de faire partie de ce panel.

Annexe 3. Caractéristiques générales de la population « répondants »

Principales caractéristiques socio-démographiques et politiques des individus qui ont répondu au questionnaire. La « Base Répondants $n = 658$ » est composée des individus qui ont accepté de répondre au questionnaire. La base « Répondants Panel ($n = 608$) » est composée des individus qui ont répondu au questionnaire **ET** qui ont accepté d'intégrer le panel.

1. Sexe des répondants

Sexe	Base Répondants ($n = 658$)		Répondants Panel ($n = 608$)	
	Effectifs	%	Effectifs	%
Un homme	391	59,4	353	58,1
Une femme	267	40,6	255	41,9
Total général	658	100,0	608	100,0

2. Âge

Âge	Base Répondants ($n = 658$)		Répondants Panel ($n = 608$)	
	Effectifs	%	Effectifs	%
Âge moyen	30,6		30,6	
Âge médian	26		26	
Moins de 18 ans	53	8,1	48	7,9
18 à 25 ans	262	39,8	249	41,0
26 à 34 ans	132	20,1	116	19,1
35 à 45 ans	124	18,8	111	18,3
46 à 60 ans	61	9,3	59	9,7
61 ans et plus	26	4,0	25	4,1
Total général	658	100,0	608	100,0

3. Lieu de résidence

Lieu de résidence	Base Répondants (n = 658)		Répondants Panel (n = 608)	
	Effectifs	%	Effectifs	%
Paris intra-muros	76	11,6	69	11,3
Agglomération parisienne	91	13,8	85	14,0
Province	443	67,3	407	66,9
Etranger	48	7,3	47	7,7
Total général	658	100,0	608	100,0

4. Taille de l'agglomération habitée

Taille agglomération	Base Répondants (n = 658)		Répondants Panel (n = 608)	
	Effectifs	%	Effectifs	%
Moins de 2 000 habitants	61	9,3	57	9,4
De 2 000 à 20 000 habitants	146	22,2	133	21,9
De 20 000 à 100 000 habitants	143	21,7	131	21,5
De 100 000 à 500 000 habitants	139	21,1	131	21,5
De plus de 500 000 habitants	169	25,7	156	25,7
Total général	658	100,0	608	100,0

5. Dernier diplôme obtenu

Diplôme	Base Répondants (n = 658)		Répondants Panel (n = 608)	
	Effectifs	%	Effectifs	%
Je ne possède aucun diplôme	7	1,1	7	1,2
Certificat d'études	4	0,6	4	0,7
Brevet des collèges ou BEPC	54	8,2	49	8,1
CAP ou BEP	32	4,9	30	4,9
Baccalauréat pro.	19	2,9	17	2,8
Baccalauréat technologique	18	2,7	16	2,6
Baccalauréat général	120	18,2	115	18,9
Deug, BTS ou équivalent Bac+2	85	12,9	78	12,8
Licence ou équivalent Bac+3	93	14,1	88	14,5
Maîtrise ou équivalent Bac+4	67	10,2	63	10,4
Master ou équivalent Bac+5	145	22,0	130	21,4
Doctorat	14	2,1	11	1,8
Total général	658	100,0	608	100,0

6. Dernier établissement d'études supérieures fréquenté

Dernier établissement d'études sup. fréquenté	Base Répondants (n = 658)		Répondants Panel (n = 608)	
	Effectifs	%	Effectifs	%
Je n'ai pas suivi d'études supérieures	138	21,0	128	21,1
Universités	281	42,7	265	43,6
Instituts universitaires de formation des maîtres	10	1,5	9	1,5
Instituts universitaires de technologie (IUT)	24	3,6	21	3,5
Classes préparatoires aux grandes écoles (CPGE)	16	2,4	15	2,5
Sections de techniciens supérieurs (STS)	14	2,1	14	2,3
Ecoles d'ingénieurs	36	5,5	32	5,3
Ecoles de commerce	37	5,6	29	4,8
Ecoles d'architecture	3	0,5	3	0,5
Ecoles supérieures artistiques et culturelles	10	1,5	9	1,5
Ecoles paramédicales et sociales hors université	14	2,1	13	2,1
Ecoles de journalisme	16	2,4	14	2,3
Instituts d'Etudes Politiques	9	1,4	9	1,5
Ecoles Normales Supérieures	6	0,9	6	1,0
Autre école	44	6,7	41	6,7
Total général	658	100,0	608	100,0

7. Statut socio-professionnel

Statut	Base Répondants (n = 658)		Répondants Panel (n = 608)	
	Effectifs	%	Effectifs	%
Lycéen	58	8,8	54	8,9
Etudiants	222	33,7	212	34,9
Travailleur intérimaire	4	0,6	2	0,3
Salarié en CDD du public	27	4,1	25	4,1
Fonctionnaire ou salarié en CDI du public	74	11,2	67	11,0
Salarié en CDD du privé	18	2,7	16	2,6
Salarié en CDI du privé	133	20,2	120	19,7
Travailleurs ind., commerçant, artisan	22	3,3	18	3,0
Professions libérales	24	3,6	23	3,8
Chef d'entreprise	15	2,3	13	2,1
Femme/homme au foyer	7	1,1	7	1,2
Chômeur	26	4,0	24	3,9
Retraité	28	4,3	27	4,4
Total général	658	100,0	608	100,0

8. PCS

PCS niveau 2	Base Répondants (n = 658)		Répondants Panel (n = 608)	
	<i>Effectifs</i>	%	<i>Effectifs</i>	%
Agriculteurs exploitants	1	0,2	1	0,2
Artisans	2	0,3	2	0,3
Commerçants et assimilés	2	0,3	1	0,2
Chefs d'entreprise de 10 salariés ou plus	5	0,8	4	0,7
Professions libérales et assimilés	15	2,3	12	2,0
Cadres de la fonction publique, professions intellectuelles et artistiques	83	12,6	72	11,8
Cadres d'entreprise	91	13,8	83	13,7
Professions intermédiaires de l'enseignement, de la santé, de la fonction publique et assimilés	47	7,1	46	7,6
Professions intermédiaires administratives et commerciales des entreprises	31	4,7	28	4,6
Techniciens	15	2,3	14	2,3
Contremaîtres, agents de maîtrise	4	0,6	1	0,2
Employés de la fonction publique	18	2,7	16	2,6
Employés administratifs d'entreprise	14	2,1	13	2,1
Employés de commerce	5	0,8	4	0,7
Personnels des services directs aux particuliers	3	0,5	3	0,5
Ouvriers qualifiés	8	1,2	7	1,2
Ouvriers non qualifiés	1	0,2	1	0,2
Anciens artisans, commerçants, chefs d'entreprise	1	0,2	1	0,2
Anciens cadres et professions intermédiaires	14	2,1	14	2,3
Anciens employés et ouvriers	7	1,1	7	1,2
Chômeurs n'ayant jamais travaillé	2	0,3	2	0,3
Inactifs divers (autres que retraités)	289	43,9	276	45,4
Total général	658	100,0	608	100,0

9. Situation matrimoniale

Situation matrimoniale	Base Répondants (n = 658)		Répondants Panel (n = 608)	
	<i>Effectifs</i>	%	<i>Effectifs</i>	%
Vous êtes célibataire	317	48,2	294	48,4
Vous êtes marié(e)	115	17,5	102	16,8
Vous êtes pacsé(e)	29	4,4	28	4,6
Vous êtes en couple sans être marié(e) ni pacsé(e)	177	26,9	165	27,1
Vous êtes veuf(ve)	1	0,2	1	0,2
Vous êtes divorcé(e)	19	2,9	18	3,0
Total général	658	100,0	608	100,0

10. Intérêt pour la politique

	Base Répondants (n = 658)		Base Panel (n = 608)	
	<i>Effectifs</i>	%	<i>Effectifs</i>	%
Intérêt pour la politique				
Beaucoup	366	55,6	341	56,1
Assez	202	30,7	185	30,4
Un peu	73	11,1	68	11,2
Pas du tout	17	2,6	14	2,3
Total général	658	100,0	608	100,0

11. Appartenance politique

	Base Répondants (n= 658)		Base Panel (n=608)	
	<i>Effectifs</i>	%	<i>Effectifs</i>	%
Appartenance politique				
Très à gauche	125	19,0	117	19,2
Plutôt à gauche	242	36,8	223	36,7
Au centre	64	9,7	60	9,9
Plutôt à droite	120	18,2	113	18,6
Très à droite	31	4,7	25	4,1
Ni à droite, ni à gauche	41	6,2	37	6,1
Vous ne savez pas	18	2,7	17	2,8
Autre	17	2,6	16	2,6
Total général	658	100,0	608	100,0

12. Vote au premier tour de l'élection présidentielle 2012

	Base Répondants (n= 658)		Base Panel (n=608)	
	<i>Effectifs</i>	%	<i>Effectifs</i>	%
Vote premier tour présidentielle 2012				
Je suis inscrit sur les listes électorales mais je n'ai pas voté	41	6,2	37	6,1
J'ai voté blanc ou nul	18	2,7	9	1,5
Philippe Poutou (NPA)	4	0,6	12	2,0
Jean-Luc Mélenchon (Front de gauche)	102	15,5	1	0,2
François Hollande (PS)	168	25,5	101	16,6
Eva Joly (EELV)	33	5,0	16	2,6
François Bayrou (MODEM)	65	9,9	4	0,7
Nicolas Sarkozy (UMP)	94	14,3	93	15,3
Nicolas Dupont-Aignan (DLR)	9	1,4	157	25,8
Marine Le Pen (FN)	15	2,3	30	4,9
Jacques Cheminade (Solidarité et Progrès)	1	0,2	60	9,9
Je ne suis pas inscrit sur les listes électorales	108	16,4	88	14,5
Total général	658	100,0	608	100,0

13. Parti le plus proche

	Base Répondants (n= 658)		Base Panel (n=608)	
	<i>Effectifs</i>	%	<i>Effectifs</i>	%
Parti le plus proche				
Lutte ouvrière (LO)	11	1,7	9	1,5
Nouveau parti anticapitaliste (NPA)	15	2,3	12	2,0
Parti communiste français (PCF)	19	2,9	19	3,1
Parti de gauche (PG)	70	10,6	64	10,5
Mouvement républicain et citoyen (MRC)	4	0,6	3	0,5
Parti socialiste (PS)	216	32,8	203	33,4
Parti radical de gauche (PRG)	4	0,6	4	0,7
Europe Ecologie Les Verts (EELV)	39	5,9	35	5,8
Alliance écologiste indépendante (AEI)	3	0,5	3	0,5
Cap 21	4	0,6	4	0,7
Mouvement démocrate (MODEM)	64	9,7	60	9,9
Parti radical (valoisien) (PR)	5	0,8	5	0,8
Nouveau centre (NC)	8	1,2	7	1,2
Union pour un mouvement populaire (UMP)	114	17,3	106	17,4
République solidaire (RS)	3	0,5	3	0,5
Debout la République (DLR)	7	1,1	6	1,0
Parti chrétien-démocrate (PCD)	5	0,8	5	0,8
Front National (FN)	14	2,1	11	1,8
Autre	53	8,1	49	8,1
Total général	658	100,0	608	100,0

14. Pratiques politiques

	Base Répondants (n= 658)		Base Panel (n=608)	
	<i>Effectifs</i>	%	<i>Effectifs</i>	%
Pratiques politiques				
Voté	572	86,9	530	87,2
Signé une pétition	531	80,7	489	80,4
Manifesté	402	61,1	369	60,7
Fait grève	213	32,4	200	32,9
Adhéré à un syndicat	104	15,8	92	15,1
Adhéré à un parti politique	201	30,5	190	31,3
Adhéré à une ou plusieurs association(s)	319	48,5	299	49,2

15. Participations aux élections de 2007-2012

Participations aux élections	Base Répondants (n= 658)		Base Panel (n=608)	
	<i>Effectifs</i>	%	<i>Effectifs</i>	%
Présidentielle 2007	346	52,6	315	51,8
Législatives 2007	317	48,2	288	47,4
Municipales 2008	364	55,3	332	54,6
Européennes 2009	304	46,2	278	45,7
Régionales 2010	332	50,5	303	49,8
Cantoniales (2008 ou 2011)	311	47,3	284	46,7
Primaires PS 2011 (organisées par le PS)	179	27,2	163	26,8
Présidentielle 2012	550	83,6	423	69,6
N'a jamais voté à une de ces élections	120	18,2	120	19,7

Annexe 4. Caractéristiques générales de la population « non répondants »

Principales caractéristiques socio-démographiques et politiques des individus qui n'ont pas répondu au questionnaire.

1. Sexe des non-répondants

Sexe	<i>Effectifs</i>	<i>%</i>
Non identifié	30	4,8
Homme	392	63,2
Femme	198	31,9
Total général	620	100,0

2. Âge

Âge estimé	<i>Effectifs</i>	<i>%</i>
Non identifié	380	61,3
Moins de 18 ans	47	7,6
18 à 25 ans	110	17,7
26 à 34 ans	34	5,5
35 à 45 ans	35	5,6
46 à 60 ans	11	1,8
61 ans et plus	3	0,5
Total général	620	100,0

3. Lieu de résidence

Lieu de résidence	<i>Effectifs</i>	<i>%</i>
Non identifié	247	39,8
Paris intra-muros	134	21,6
Agglomération parisienne	41	6,6
Province	163	26,3
Etranger	35	5,6
Total général	620	100,0

4. Niveau de diplôme

Diplôme	<i>Effectifs</i>	%
Non identifié	395	63,7
Inférieur au Bac	49	7,9
Niveau Bac	60	9,7
Deug, BTS ou équivalent Bac+2	21	3,4
Licence ou équivalent Bac+3	16	2,6
Maîtrise ou équivalent Bac+4	25	4,0
Master (ex. DEA ou DESS) ou équivalent Bac+5	50	8,1
Doctorat	4	0,6
Total général	620	100,0

5. PCS

PCS niveau 2	<i>Effectifs</i>	%
Non identifiée	294	47,4
Agriculteurs exploitants	0	0,0
Artisans	1	0,2
Commerçants et assimilés	5	0,8
Chefs d'entreprise de 10 salariés ou plus	2	0,3
Professions libérales et assimilés	4	0,6
Cadres de la fonction publique, professions intellectuelles et artistiques	68	11,0
Cadres d'entreprise	71	11,5
Professions intermédiaires de l'enseignement, de la santé, de la fonction publique et assimilés	6	1,0
Professions intermédiaires administratives et commerciales des entreprises	19	3,1
Techniciens	10	1,6
Contremaîtres, agents de maîtrise	0	0,0
Employés de la fonction publique	1	0,2
Employés administratifs d'entreprise	4	0,6
Employés de commerce	2	0,3
Personnels des services directs aux particuliers	0	0,0
Ouvriers qualifiés	1	0,2
Ouvriers non qualifiés	1	0,2
Anciens artisans, commerçants, chefs d'entreprise	0	0,0
Anciens cadres et professions intermédiaires	1	0,2
Anciens employés et ouvriers	1	0,2
Chômeurs n'ayant jamais travaillé	3	0,5
Inactifs divers (autres que retraités)	126	20,3
Total général	620	100,0

6. Statut

Statut	<i>Effectifs</i>	%
Non identifié	289	46,6
Lycéen	49	7,9
Etudiants	75	12,1
Actifs en activité	193	31,1
Actifs en recherche d'activité	10	1,6
Inactifs	4	0,6
Total général	620	100,0

7. Appartenance politique

Appartenance politique	<i>Effectifs</i>	%
Non identifiée	356	57,4
Gauche non-socialistes et extrême-gauche	36	5,8
Gauche socialiste	118	19,0
Centre	14	2,3
Droite	88	14,2
Extrême-droite	8	1,3
Total général	620	100,0

Annexe 5. Caractéristiques générales du panel « Twitter audience politique »

1. Catégorie des comptes panélisés

Catégorie	Effectifs	%
Journalistes	189	18,2
Blogueurs	113	10,9
Médias <i>pureplayer</i>	83	8,0
Autres	83	8,0
Presse écrite	73	7,0
Artistes et sportifs	53	5,1
Députés	42	4,0
Présentateurs TV et radio	41	3,9
Chefs de service	35	3,4
Entrepreneurs	35	3,4
Experts	29	2,8
Chroniqueurs	27	2,6
Directeurs de rédaction	24	2,3
Cadres d'organisation	24	2,3
Ministres	22	2,1
Chaîne de télévision	22	2,1
Associations	18	1,7
Radios	14	1,3
Partis politique	13	1,2
Députés européens	12	1,2
Entreprises	12	1,2
Maires	12	1,2
Sénateurs	11	1,1
Institutions publiques	11	1,1
Partis politiques	9	0,9
Conseillers municipaux	8	0,8
Conseillers régionaux	8	0,8
Editorialistes	7	0,7
Présidents de parti politique	6	0,6
Association de défense de cause	3	0,3
Porte-parole de parti politique	2	0,2
Total général	1041	100,0

2. Partis politique associés aux comptes

Parti politique	<i>Effectifs</i>	%
Non identifié / non attribué	849	81,6
PS	86	8,3
UMP	52	5,0
Autres	13	1,2
EELV	12	1,2
UDI	8	0,8
Parti de Gauche	7	0,7
PCF	5	0,5
Modem	5	0,5
FN	4	0,4
Total général	1041	100,0

3. Orientation politique supposée des comptes

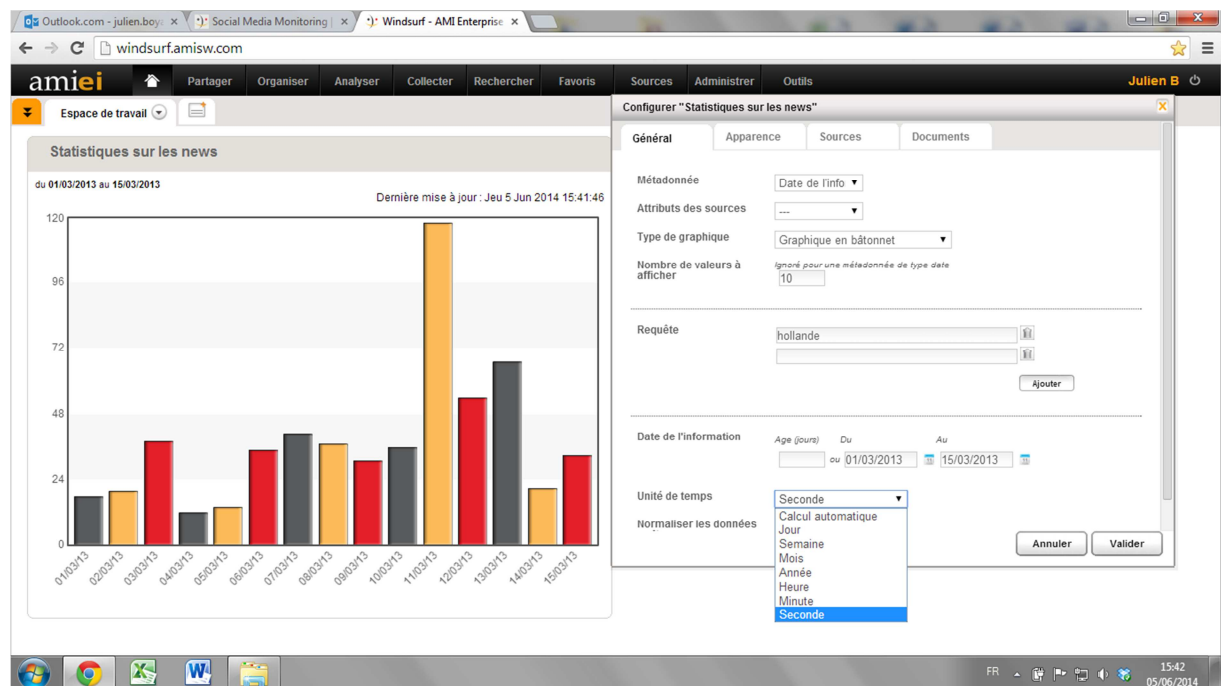
Orientation politiques	<i>Effectifs</i>	%
Non identifiée / non attribuée	609	58,5
Gauche	289	27,8
Droite	121	11,6
Centre	17	1,6
FN	5	0,5
Total général	1041	100

Annexe 6. Présentation du logiciel AMI Opinion Tracker

AMI Opinion Tracker est un logiciel de *web-tracking* qui permet de collecter de façon automatisée des données numériques sur différents espaces du web (blogs, sites d'information, Facebook, Twitter, etc.), de les archiver, les classer, les catégoriser et les analyser à l'appui d'outils lexicométriques ou cartographiques.

Nous proposons ici de présenter quelques-unes de ces fonctions, celles qui ont été les plus utiles à l'analyse de nos données mais également celles qui nous semblent les plus prometteuses méthodologiquement.

1. Le widget « Statistiques sur les news » : un outil d'analyse volumétrique des données



Le widget « Statistiques sur les news » permet de mesurer le volume de messages publiés sur une période de temps donnée et de représenter cette information sous forme graphique (histogramme, courbes, radar, etc.). L'utilisateur doit pour cela paramétrer différentes options :

- **La métadonnée** : il s'agit de l'information qui figurera en abscisse de l'histogramme. Il peut s'agir d'une information temporelle (année, mois, semaine, jour, heure, minute, seconde) – on

analysera alors la distribution du nombre de messages dans le temps – ou bien les différentes sources analysées – on pourra alors comparer le nombre de messages citant différents candidats par exemple.

- **La plage temporelle étudiée** : quelle que soit la métadonnée retenue, l'utilisateur doit définir la plage de temps qu'il souhaite étudier. L'utilisateur peut ainsi mesurer l'évolution du nombre de messages publiés durant une heure, un jour ou plusieurs jours.
- **Le type de graphique** : il peut s'agir d'un histogramme, d'une courbe, d'un radar, de bulles, etc.
- **Les sources sélectionnées** : l'utilisateur doit sélectionner la ou les sources qu'il souhaite interroger, c'est-à-dire la ou les sources à l'origine des messages collectés et mesurés par le *widget*. Il peut s'agir, dans le cas de Twitter, de sources de type *search* ou des sources de type *count*.
- **La définition de requête** : l'utilisateur peut, s'il le souhaite, définir une ou plusieurs requêtes, c'est-à-dire filtrer en amont les messages collectés par les sources interrogées en fonction d'un mot-clé ou d'une combinaison de mots-clés. Si l'on sélectionne par exemple la source de type *search* « Hollande » et que l'on rentre comme requête « chômage », ne seront alors comptabilisés que les tweets contenant l'occurrence « Hollande » **ET** l'occurrence « chômage ».

2. Le *widget* « lecteur de news » : lire et sélectionner les documents que l'on souhaite analyser

The screenshot shows the Amii Social Media Monitoring interface. The main window displays a list of news items under the 'Lecteur de News' widget. The configuration dialog 'Configurer "Lecteur de News"' is open, showing the 'Général' tab. The 'Ordre de tri' dropdown is set to 'Échantillon Aléatoire'. The 'Requête' field contains 'Échantillon Aléatoire'. The 'Avec hypothèses' checkbox is unchecked. The 'Age des documents (jours)' is set to '0'. The 'Masquer les documents similaires' section has 'Même URL' and 'Même titre' checkboxes unchecked. The 'Annuler' and 'Valider' buttons are visible at the bottom of the dialog.

Le « lecteur de *news* » permet de lire et trier les documents collectés par le logiciel et de les classer dans un dossier qui pourra être ensuite soumis à une analyse lexicométrique ou cartographique. Comme pour le *widget* « Statistiques sur les *news* », l'utilisateur doit sélectionner les sources qu'il souhaite interroger, éventuellement les filtrer à l'aide d'une ou plusieurs « requêtes » et définir une période de temps d'analyse. L'utilisateur est amené également à choisir entre trois ordres de tris : le tri par date, le tri par pertinence et le tri aléatoire. Le tri par date permet de faire apparaître les documents par ordre chronologique, selon l'heure de publication des messages. Le tri par pertinence ne concerne que les billets de blogs et repose sur un calcul de fréquence d'occurrences, nous ne nous y arrêtons pas ici. Le tri aléatoire a quant à lui été instauré à notre demande. Il permet d'afficher selon un ordre totalement aléatoire les documents correspondants à la recherche paramétrée. Cela peut être utile si l'utilisateur souhaite analyser un échantillon représentatif de messages. Ainsi, si l'on souhaite par exemple connaître la tonalité des tweets parus à propos d'un candidat sur une journée donnée, plutôt que de coder l'intégralité des messages (qui se comptent parfois en millions), il est possible de ne sélectionner qu'un échantillon aléatoire de 1000 documents. Les résultats obtenus sur ces 1000 documents seront alors extrapolables, en suivant la loi des grands nombres, à l'ensemble du corpus.

3. « Mon Twitter audience politique » : classer les tweets en fonction de leur nombre de reprises et cartographier les logiques de circulation des messages sur le réseau social Twitter

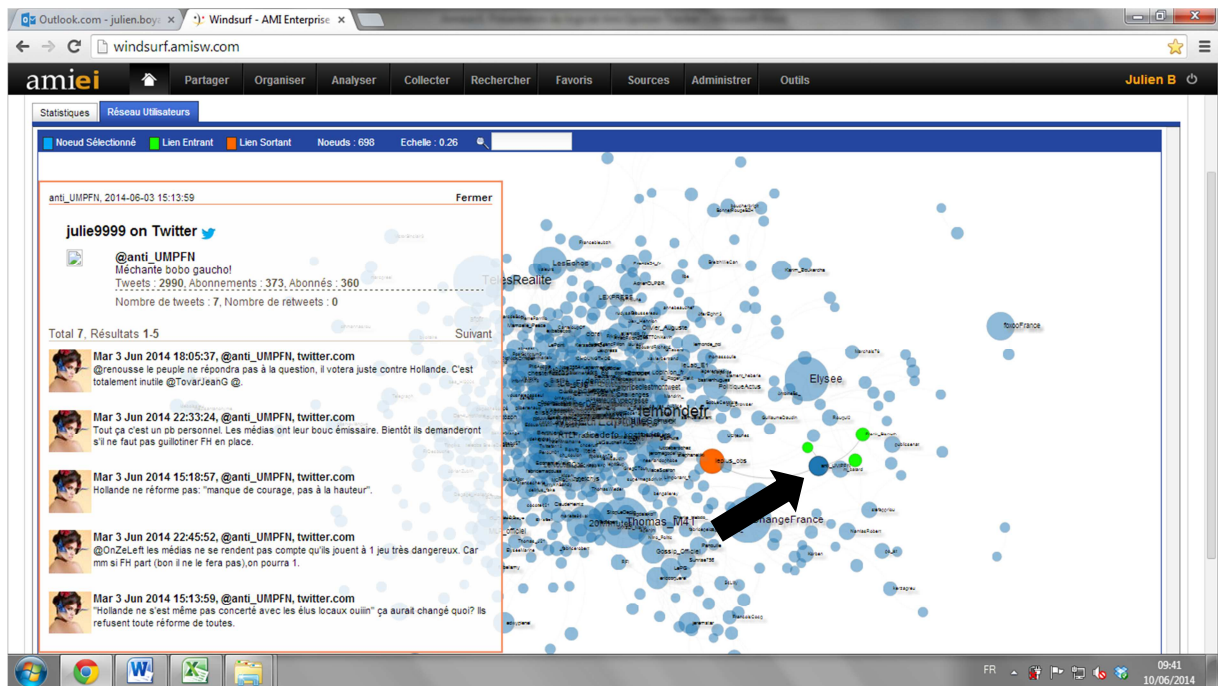
Contrairement aux deux *widgets* précédents, « Mon Twitter audience politique » est encore, en 2014, à l'état de prototype. Nous y avons cependant accès à titre expérimental depuis l'année 2012. Il permet de connaître d'une part quels sont les tweets les plus retweetés sur le réseau social durant une période de temps donnée (la fonction « Top Retweets ») et permet d'autre part de cartographier la circulation des tweets et des retweets entre les différents utilisateurs qui ont participé à la production et à la circulation des messages étudiés.

Le « Top Retweets »



Le « Top Retweets » permet, sur une période de temps définie, de classer les tweets contenant une occurrence donnée en fonction de leur nombre de reprises, de retweets sur le réseau social. L'utilisateur doit pour cela sélectionner une source Twitter, définir une période de temps d'analyse et le nombre de tweets qu'il souhaite voir figurer dans le « top ». Ainsi, à titre d'exemple, si l'on choisit de sélectionner la source Twitter de type *search* Hollande, du 01/01/2013 au 31/12/2013 et qu'on limite le nombre de tweets affichés à 10, on connaîtra alors les 10 tweets évoquant François Hollande les plus repris sur l'ensemble du réseau social sur l'année 2013. Le widget permet également d'établir quels sont les auteurs de ces messages les plus repris, ainsi que le nombre total de tweets publiés par ces comptes.

Cartographie des réseaux d'utilisateurs



En plus de hiérarchiser les tweets en fonction de leur nombre de reprises sur le réseau, le *widget* « Mon Twitter audience politique » permet de cartographier les réseaux d'utilisateurs qui ont tweeté ou retweeté les messages correspondant à la recherche paramétrée par l'utilisateur.

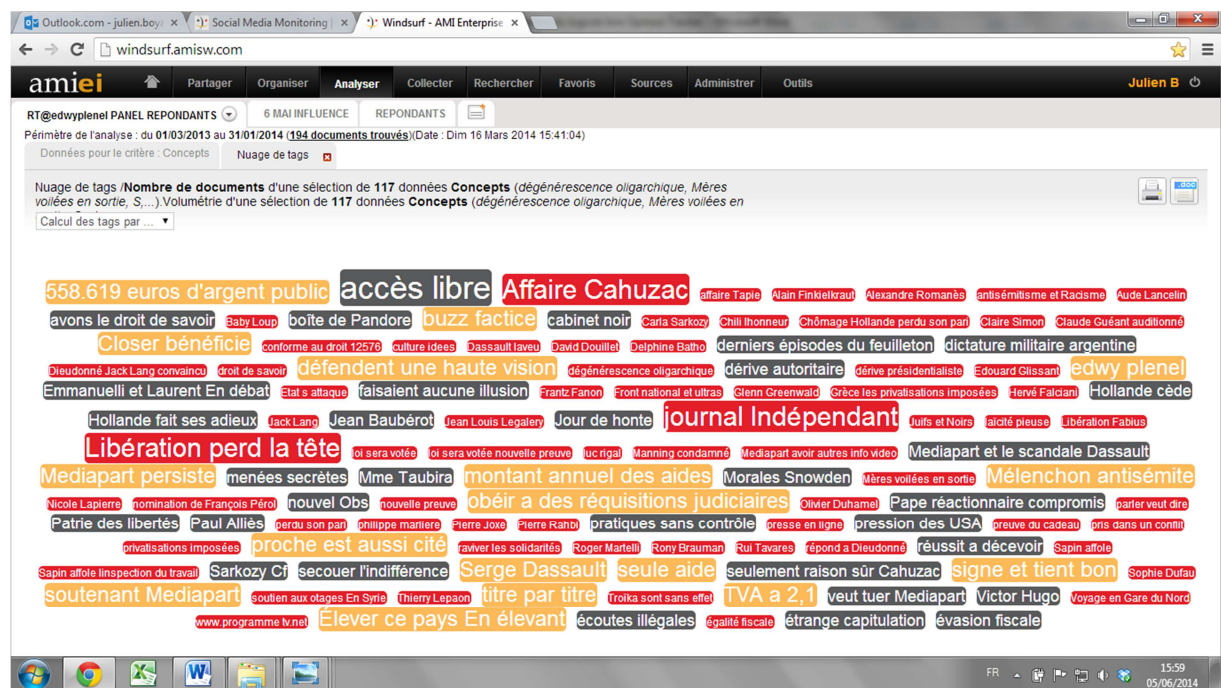
Chaque nœud correspond soit à un compte d'utilisateur dont l'un des tweets a été retweeté au moins une fois par un autre utilisateur soit à un compte d'utilisateur qui a retweeté au moins un tweet publié par un autre utilisateur (tweet correspondant à la recherche paramétrée par l'utilisateur, ici, à titre d'exemple, les tweets contenant l'occurrence « Hollande » publiés sur l'ensemble de l'année 2013). La taille des nœuds correspond au nombre total de fois où les messages publiés par le compte ont été retweetés. Plus un compte est retweeté, plus le diamètre de son nœud est grand (précisons que l'utilisateur peut également choisir d'associer au diamètre des nœuds non pas le nombre de retweets mais le nombre de tweets publiés par le compte ; les nœuds les plus « gros » correspondent alors aux comptes les plus productifs mais pas nécessairement aux plus « influents »).

Lorsque l'utilisateur pointe le curseur de la souris sur l'un de ces nœuds, celui-ci se teinte en bleu (*cf.* nœud pointé par la flèche sur le graphique ci-dessus) et une nouvelle fenêtre s'ouvre. Celle-ci contient des informations sur le compte : sa description de profil, son nombre d'abonnés, d'abonnements, le nombre de fois où ses tweets ont été retweetés ainsi que la liste des messages correspondant à la recherche paramétrée par l'utilisateur. De plus, lorsqu'un lien est sélectionné, le logiciel rend apparent les liens entrants (teintés en vert) et les liens sortants (teintés en orange) qui les lient aux autres

comptes d'utilisateurs. Les nœuds teintés en vert correspondent aux comptes qui ont retweeté un ou plusieurs messages du nœud sélectionné par l'utilisateur (teinté en bleu). Les nœuds teintés en orange correspondent aux comptes dont l'un des tweets a été retweeté par le compte sélectionné (teinté en bleu).

Enfin, l'utilisateur peut visionner l'évolution dans le temps de ce réseau d'utilisateur. Les nœuds et les liens entrants et sortants apparaîtront alors à l'écran selon un ordre chronologique, correspondant à la période de temps paramétré par l'utilisateur. À l'aide de boutons similaires à ceux que l'on trouve sur un magnétoscope (lecture, pause, avance rapide, etc.) l'utilisateur peut suivre la circulation des messages dans le temps, s'arrêter sur des moments précis, revenir en arrière, etc.

3. Le module « Analyser » : analyser les corpus de messages à l'appui d'outils lexicométriques



Contrairement aux autres outils que nous avons présentés plus haut, « Analyser » n'est pas un *widget* mais un module à part entière. Il contient à la fois des outils de mesure de volumétrie (qui ont été depuis repris dans différents *widgets*) mais aussi certains outils lexicométriques, comme notamment des nuages de *tags* (cf. graphique ci-dessus), des cartographies de co-occurrences de termes et des cartographies de réseaux de concepts. Le nuage de *tags* permet par exemple de rendre apparentes les occurrences qui reviennent le plus fréquemment dans les messages du corpus analysé en attribuant à

ces occurrences une taille de caractère proportionnelle au nombre de documents où elles se trouvent référencées.

Nous n'avons cependant pas utilisé ces outils lexicométriques pour analyser nos données, car ils ne nous paraissent pas encore suffisamment fiables. Ils peinent notamment à éliminer le « bruit », c'est-à-dire toutes les occurrences ne correspondant pas au contenu des documents analysés mais à des informations annexes (par exemple « *envoyer cet article par mail* »). Cependant, bien que leur potentiel reste encore à être exploiter, l'ensemble de ces outils nous semble prometteur.

Annexe 7. Requête de sélection des tweets politiques

Afin d'être considérés comme « politiques », les tweets archivés par notre logiciel devaient contenir l'une des 121 occurrences suivantes :

1. Des noms propres correctement orthographiés de femmes et d'hommes politiques

« *hollande* OR *ayrault* OR *valls* OR *montebourg* OR *taubira* OR *duflot* OR *moscovici* OR *belkacem* OR *peillon* OR *fabius* OR *batho* OR *touraine* OR *sapin* OR *drian* OR *foll* OR *hamon* OR *filippetti* OR *fioraso* OR *royal* OR *aubry* OR *bartolone* OR *cahuzac* OR *bayrou* OR *borloo* OR *jupe* OR *raffarin* OR *cope* OR *fillon* OR *sarkozy* OR *jacob* OR *morano* OR *wauquiez* OR *guaino* OR *barjot* OR *le pen* OR *collard* OR *maréchal* OR *mélenchon* OR *laurent* »

2. Des noms propres mal orthographiés de femmes et d'hommes politiques

« *holande* OR *ayraut* OR *ayrau* OR *airaut* OR *mélanchon* OR *sarkozi* OR *sarkosy* OR *sarkosi* OR *coppe* »

3. Des surnoms et sobriquets attribués à des hommes politiques

« *sarko* OR *flambi* OR *flamby* OR « *normal 1er* » »

4. Des noms et sigles de partis politiques

« *npa* OR *fdg* OR « *front de gauche* » OR *pcf* OR « *parti communiste* » OR *pg* OR « *parti de gauche* » OR *eelv* OR « *les verts* » OR *ps* OR « *parti socialiste* » OR *modem* OR *udi* OR « *ump* » OR *fn* OR « *front national* » »

5. Des termes relatifs à des fonctions ou à des institutions politiques

« *ministre* OR *ministres* OR *gouvernement* OR *matignon* OR *élysée* OR « *cour des comptes* » OR *loi* OR *lois* OR *décret* OR *décrets* OR *amendement* OR *amendements* OR *majorité* OR *opposition* OR

parlement OR parlementaires OR député OR députés OR sénateur OR sénateurs OR sénat OR directan OR assemblée OR parlement OR fmi OR omc OR ue OR commission OR bruxelles »

6. Des termes relatifs aux idées politiques

« politique OR politiques OR gauche OR droite OR centre OR extrême OR république OR démocratie OR capitalisme OR capitaliste OR capitalistes OR communisme OR communiste OR communistes OR socialisme OR écologie OR écologiste OR écologistes OR fascisme OR fasciste OR fascistes OR facho OR anarchisme OR anarchistes »

Annexe 8. Grille de codage des tweets politiques

Afin d'obtenir davantage d'informations sur notre population, nous avons codé manuellement un échantillon représentatif de 1041 tweets politiques publiés par notre panel « Répondants » à partir de la grille suivante :

Nature du tweet	1. Tweet
	2. Retweet sans commentaire
	3. Retweet avec commentaire
Nature du compte retweeté	1. Grand média
	2. Journaliste de grand média
	3. Petit média / blog
	4. Journaliste de petit média / blogueur
	5. Organisation politique
	6. Homme politique
	7. Twitto militant
	8. Twittos non militant
@Destinataire du tweet	0. Aucun
	1. Public indéfini
	2. Grand média
	3. Journaliste grand média
	4. Journaliste petit média / blogueur
	5. Organisation politique
	6. Homme politique
	7. Twitto militant
8. Twitto non militant	
Insertion du tweet dans une discussion	1. Le tweet ne provoque pas d'échange
	2. Provoque un échange ou une réaction
	3. Première réponse à un tweet
	4. S'insère dans un échange
Impact sur la compréhension du message	1. Indispensable
	2. Eclairant mais pas indispensable
	3. Inutile
Présence de lien hypertexte	1. Absence de lien
	2. Présence de lien
Nature du lien	1. Lien vers un grand média
	2. Lien vers un petit média
	3. Autres liens
	4. Lien vers une organisation politique
	5. Liens vers un site institutionnel
	6. Lien photo
	7. Lien Youtube ou Dailymotion

Lien commenté	1. Lien non commenté
	2. Lien commenté
Tonalité subjective du lien	1. Positif
	2. Neutre
	3. Négatif
Tonalité objective du lien	1. Positif
	2. Neutre
	3. Négatif
Impact sur la compréhension du message	1. Indispensable
	2. Eclairant mais pas indispensable
	3. Inutile
Niveau de langue	1. Familier
	2. Courant
	3. Soutenu
Niveau orthographe	1. Aucune faute
	2. Peu de fautes
	3. Beaucoup de fautes
Contexte d'émission du tweet	1. Actualité
	2. Réactions à des déclarations
	3. Prestations télévisées
	4. Hors-actualité immédiate
	5. Hors-actualité
Registre	1. Injonction
	2. Polémique, critique
	3. Ironie, humour
	4. Informatif
	5. Soutien
	6. Argumentatif
Technicité termes politiques	1. Elevée
	2. Moyenne
	3. Faible
Mode de production	1. Ethos
	2. Logos
Tonalité subjective du tweet	1. Positif
	2. Neutre
	3. Négatif
	4. Ambigu
Tonalité objective du tweet	1. Positif
	2. Neutre
	3. Négatif
	4. Ambigu

Nature de l'objet du message	1. Attributs politiques d'un acteur politique
	2. Attributs non-politiques d'un acteur politique
	3. Appréciation politique d'un output politique
	4. Appréciation non-politique d'un output politique
	5. Idéel, idéologie / abstraction politique

Vu et permis d'imprimer

Montpellier, le

Le Président de l'Université Montpellier I

Philippe AUGE

Résumé en français :

Les milliers de messages publiés quotidiennement sur internet constituent autant d'indices de pratiques, d'attitudes et d'opinions exprimées sur de nombreux sujets, dont la politique. Ces messages peuvent être appréhendés comme un véritable matériau d'analyse du monde social. Ils présentent néanmoins une certaine spécificité par rapport à d'autres types de données : ils ne sont pas générés *par* et *pour* un protocole de recherche. De ce fait, le chercheur ignore bien souvent les propriétés sociologiques de leurs auteurs. Afin de pouvoir situer ces auteurs dans le monde social «réel», nous avons construit notre propre dispositif méthodologique de panélisation d'une population d'inscrits au réseau social Twitter. Les données générées par notre dispositif nous ont permis d'observer que, bien que politisés et dotés en capitaux culturels, les individus publiant des messages politiques ne le font que de façon très intermittente. Le niveau de production de messages politiques sur Twitter est en fait corrélé au niveau d'activité du champ de production de l'information et de l'opinion. On peut donc appréhender Twitter comme un observatoire du marché des opinions politiques.

Titre et résumé en anglais :**Analyzing digital political opinion**

Thousands of texts daily published on the Internet indicate practices, attitudes and opinions on plenty of issues, politics included. They can be considered a real material to analyze the social world. These digital texts are quite specific, with regard to other types of data: they are not generated *by* and *for* a research protocol. Therefore, the researcher ignores the sociological properties of their authors. In order to identify these authors in the real social world, I built my own methodological plans to panelize the members of the social network Twitter. The data generated by my method support the following idea: the individuals who publish political texts are politicized and have high cultural capital. Besides, they only do so on an occasional basis. The level of production of political tweets is in fact correlated to the level of activity of the *field* of production of information and opinion. Twitter can therefore be considered a monitoring tool of the political opinion market.

Discipline :

Science politique

Mots-clés :

Internet, Twitter, opinions politiques, opinion publique

Intitulé et adresse de l'U.F.R. ou du Laboratoire :

Centre d'Etudes Politiques de l'Europe Latine - UMR 5112 Faculté de Droit et de Science politique - 39, rue de l'Université 34060 MONTPELLIER cedex 2