



HAL
open science

Structuration des discours au sein de Twitter durant l'élection présidentielle française de 2017

Pierre Ratinaud, Nikos Smyrnaio, Julien Figeac, Guillaume Cabanac, Ophélie Fraasier, Gilles Hubert, Yoann Pitarch, Tristan Salord, Thibaut Thonet

► **To cite this version:**

Pierre Ratinaud, Nikos Smyrnaio, Julien Figeac, Guillaume Cabanac, Ophélie Fraasier, et al.. Structuration des discours au sein de Twitter durant l'élection présidentielle française de 2017. Réseaux : communication, technologie, société, 2019, Enquêter à partir des traces textuelles du web, 2 (214-215), pp.171-208. <10.3917/res.214.0171>. <halshs-02161154>

HAL Id: halshs-02161154

<https://shs.hal.science/halshs-02161154v1>

Submitted on 20 Jun 2019

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.



HAL Authorization

STRUCTURATION DES DISCOURS AU SEIN DE TWITTER DURANT L'ÉLECTION PRÉSIDENTIELLE FRANÇAISE DE 2017

Entre agenda politique et représentations sociales

Pierre Ratinaud, Nikos Smyrnaiois, Julien Figeac, Guillaume Cabanac, Ophélie Fraisier, Gilles Hubert, Yoann Pitarch, Tristan Salord et Thibaut Thonet

La Découverte | « Réseaux »

2019/2 n° 214-215 | pages 171 à 208

ISSN 0751-7971

ISBN 9782348043581

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-reseaux-2019-2-page-171.htm>

Distribution électronique Cairn.info pour La Découverte.

© La Découverte. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

STRUCTURATION DES DISCOURS AU SEIN DE TWITTER DURANT L'ÉLECTION PRÉSIDENTIELLE FRANÇAISE DE 2017

Entre agenda politique et représentations sociales ¹

Pierre RATINAUD
Nikos SMYRNAIOS
Julien FIGEAC
Guillaume CABANAC
Ophélie FRAISIER
Gilles HUBERT
Yoann PITARCH
Tristan SALORD
Thibaut THONET

1. Ce travail a été réalisé dans le cadre du projet ANR LisTIC portant la référence ANR-16-CE26-0014-01 et du LABEX SMS portant la référence ANR-11-LABX-0066.

Si les technologies numériques ont pour certains profondément modifié la vie quotidienne d'une partie de l'humanité, il convient de souligner qu'elles n'ont pas eu le même effet en ce qui concerne les théories des sciences humaines et sociales. Même si l'on admet que les mesurables sociaux et cognitifs sont fortement régulés par leur contexte de développement et d'expression, il apparaîtrait abusif de soutenir que les processus psycho-sociaux mis au jour avant l'apparition d'internet sont aujourd'hui caducs. En revanche, on peut aisément admettre que les quantités phénoménales de données produites chaque jour par des humains ou par les outils qu'ils utilisent pour « se connecter », participent à offrir de nouveaux horizons de recherche et/ou proposent de nouvelles voies d'investigation et de validation/réfutation des modèles antérieurs. Le champ des analyses lexicales intègre ce mouvement et la disponibilité des corpus générés par le développement des technologies numériques s'accompagne du développement de méthodes et d'études qui échappent parfois aux chercheurs en sciences humaines (Cointet et Parasie, 2018). Dans ce travail, nous nous intéressons aux discours produits sur Twitter pendant la campagne électorale pour la présidentielle française de 2017. Nous serons amenés à analyser une quantité importante de données textuelles dans le cadre d'un protocole hérité de la tradition des études sur les représentations sociales.

Le réseau socionumérique Twitter fait partie des dispositifs qui ont très rapidement trouvé une place de choix dans la littérature (Weller *et al.*, 2013) en raison de la particularité des modes de production et de diffusion des messages, qui tient en fait dans la nature « sociale » et publique de la circulation de ces informations. En effet, un compte Twitter peut avoir des « followers », correspondant aux autres comptes Twitter qui s'abonnent à ce dernier pour en « suivre » les publications. Par ailleurs, chaque compte peut choisir de devenir le « follower » d'un autre compte, et donc recevoir l'intégralité de ses productions. La modélisation de ce système de comptes « suiveurs » et de comptes que l'on suit permet alors d'étudier les réseaux de relations qu'entretiennent les comptes entre eux, uniquement à partir de l'étude des abonnements et des abonnés (Golbeck et Hansen, 2014). Une étude de ces relations réalisée à l'échelle d'un pays comme l'Australie permet de conclure que la

structuration de ce réseau d'interactions est principalement guidée par un processus d'homologie de centres d'intérêt (Bruns, Moon, Münch et Sadkowsky, 2017). Autrement dit, les utilisateurs de Twitter ont tendance à s'abonner à des comptes dont les auteurs partagent un de leurs centres d'intérêt, qu'il s'agisse de sport, de musique ou de gastronomie.

Si l'on porte un regard moins « macro » sur les informations circulant sur Twitter et que l'on s'intéresse à des discussions particulières qui ont une dimension politique, un autre étalon peut être utilisé pour mesurer les interactions : le retweet (Conover *et al.*, 2011). La pratique du retweet consiste à faire suivre aux abonnés d'un compte les tweets reçus par le biais d'un abonnement. Le retweet est donc une action volontaire des utilisateurs qui a pour objectif de transmettre une information auprès des utilisateurs qui suivent leur compte. Il devient donc possible d'étudier les interactions entre utilisateurs en s'intéressant à « qui retweete qui ». Les études menées sur cette base dessinent un autre espace de sociabilité : la structuration des réseaux de relations qu'elles mettent au jour repose clairement sur un principe d'homologie axiologique (Smyrniaos et Ratinaud, 2014, 2017 ; Ratinaud, 2016). Quand la discussion traite d'un sujet proprement politique, la tendance qui se dessine est celle d'une pratique du retweet très majoritairement guidée par le fait de retweeter des informations provenant d'un compte dont les positionnements idéologiques sont congruents avec ceux du retweeteur. En accord avec le modèle de la pensée sociale (Rouquette, 1973, 2009) et particulièrement avec l'approche socio-dynamique des représentations sociales (Doise, 1989), définies dans ce cadre comme des « principes générateurs de prise de position » (Doise, 1985, p. 245), on peut alors faire l'hypothèse que le contenu informationnel qu'un utilisateur fera circuler dans ce contexte est lui aussi en adéquation avec ses opinions politiques. Cette redondance axiologique dans les informations auxquelles accèdent les utilisateurs ne sera alors plus interprétée comme l'apparition d'une « bulle de filtre » (Parisier, 2011 ; Flaxman, Goel et Rao, 2016) qui isole algorithmiquement les sujets dans un environnement informationnel confiné et peu hétérogène, mais comme le prolongement au sein des réseaux socionumériques des processus psycho-sociologiques qui régulent les comportements humains, par exemple la tendance des acteurs à éviter les sujets de discussion susceptibles de les exposer à des opinions politiques non alignées aux leurs (Muntz, 2000). Cette lecture consiste donc à considérer que les logiques sociales qui guident les pratiques sur Twitter (pratique de l'abonnement ou du retweet) ne sont pas fondamentalement différentes de celles qui présidaient à la consommation de la presse dans les années 1970 (Bourdieu, 1977).

Cette étude a pour objectif d'étudier le contenu des échanges au sein de Twitter des partisans des cinq principaux candidats à l'élection présidentielle de 2017. La spécificité de ce travail et son apport à ce dossier « enquêter sur les traces textuelles du web » consistent à proposer une méthode qui s'inscrit dans le cadre théorique des représentations sociales. Après avoir rappelé les principes de ce cadre théorique, nous montrerons que ce travail repose sur un protocole méthodologique hérité de cette tradition de recherche qui consiste à investiguer les discours en fonction des appartenances des sujets à l'une ou l'autre des communautés politiques repérées. Ce protocole se décompose en trois étapes. Dans un premier temps, il s'agit de déterminer les différentes communautés politiques à partir de l'étude du graphe des retweets obtenus sur un large corpus de tweets en rapport avec la campagne. Dans un deuxième temps, nous produisons un corpus du contenu textuel des tweets de chacune des communautés que nous analysons avec la méthode Reinert (Reinert 1983, 1990 ; Ratinaud et Marchand, 2012) dont l'objectif est d'établir les différentes thématiques abordées. La dernière étape consiste à comparer les thématiques entre les différentes communautés politiques identifiées pour rendre saillants les points de convergence et les spécificités des regards sur certains objets.

IDÉOLOGIE, REPRÉSENTATIONS SOCIALES ET DISCOURS

L'analyse de données textuelles proposée dans cette contribution s'inscrit dans le cadre théorique des représentations sociales. Une représentation sociale peut être définie comme « une forme de connaissance, socialement élaborée et partagée ayant une visée pratique et concourant à la construction d'une réalité commune à un ensemble social » (Jodelet, 1989, p. 36). Les représentations sociales renvoient donc à la définition de réalités communes à des groupes sociaux. Elles concernent des objets complexes, polymorphes, qui peuvent ne pas avoir d'existence physique (le travail, la santé, la folie, l'éducation...) et qui ont une valeur d'enjeux pour certains groupes sociaux. Ces objets font partie d'une dynamique sociale qui conduit à une mise en concurrence entre les groupes. Parce qu'elles permettent de se distinguer des autres et de reconnaître les siens, les représentations sociales participent à définir les identités. Ces réalités collectives ont également pour fonction de donner une signification et une compréhension de l'environnement des groupes et donc de participer à orienter et à justifier les conduites de leurs membres. Deux processus en interaction sont à l'origine de leur formation : l'objectivation

et l'ancrage. L'objectivation est un processus de sélection d'informations et de simplification de l'objet. Il rend concret ce qui apparaît abstrait en première intention. L'ancrage est un processus d'incorporation d'un nouvel objet dans un « déjà-là pensé » (Jodelet, 1989), c'est-à-dire dans un ensemble déjà constitué de connaissances.

Dans son étude princeps, Moscovici (1961) a montré que la façon de percevoir un objet complexe comme la psychanalyse dépend essentiellement de l'appartenance des sujets à l'un ou l'autre des groupes sociaux examinés. Dans cette étude, les groupes sociaux étaient déjà en partie distingués par leur inscription dans des idéologies différentes (communiste vs catholique par exemple). Dans le prolongement de ces travaux, le modèle de la pensée sociale pose que les processus d'ancrage et d'objectivation sont dépendants d'un processus idéologique qui régule la congruence de l'ensemble d'un système représentationnel (Rouquette, 1996).

Cette lecture particulière distingue donc clairement les idéologies comme « produits » (celles que nous pouvons nommer, comme les convictions politiques et religieuses par exemple) du processus idéologique qui régule leur fonctionnement. Il est admis que la hiérarchie entre idéologies, représentations sociales, attitudes et opinions correspond à la fois à un gradient de variabilité inter- et intra-individuel (il y a beaucoup plus d'opinions exprimées à un moment donné que d'idéologies et ces dernières sont beaucoup plus résistantes dans le temps) et d'intégration (dans le sens où les idéologies concernent des groupes sociaux plus larges que les représentations, elles-mêmes concernant des groupes plus larges que les attitudes ou les opinions) (Flament et Rouquette, 2003). Pour Van Dijk (1998), l'ensemble des processus psycho-sociaux qui interviennent dans ce modèle sont également régulés, au plus haut niveau, par les cultures dans lesquelles les groupes concernés évoluent. Rouquette (1998) propose de lire ce modèle comme une « hiérarchie de raison » qui va des idéologies vers les opinions, dans le sens où les idéologies rendent raison d'une famille de représentations sociales, celles-ci pouvant à leur tour expliquer une famille de positionnements attitudeaux, ces positionnements expliquant un ensemble d'opinions. Il a par ailleurs été montré expérimentalement que le sens des dépendances réciproques correspond à la direction envisagée par le modèle : modifier expérimentalement la représentation d'un objet conduit les sujets à revoir leur positionnement attitudeal alors que l'inverse n'est pas vrai (Tafari, 1997) ; de la même façon, changer l'inscription idéologique d'un objet conduit à le

concevoir différemment, alors qu'un groupe peut reconnaître un objet que l'on décrit sans certaines de ses caractéristiques les plus importantes si on inscrit cet objet dans une idéologie congruente avec celle du groupe étudié (Rateau, 2000).

L'expression d'une représentation sociale est donc constituée d'éléments permettant la définition de l'objet, nous parlerons ici du contenu de la représentation, mais également de positionnements attitudeux (Moliner, 1994). Les discours sur les objets sont empreints d'éléments qui spécifient les positionnements des locuteurs et qui peuvent permettre la reconnaissance de leur inscription idéologique ou de la composante attitude de leur rapport à l'objet. Dans l'approche sociologique de l'étude de l'opinion politique sur Twitter, c'est bien souvent uniquement cette dernière composante qui est envisagée (voir Severo et Lamarche-Perrin (2018) pour une revue de la littérature abordant cette question). Les travaux ont ici régulièrement recours à des méthodes de « *sentiment analysis* » qui permettent l'étude du rapport des sujets aux objets en termes de pour/contre ou de favorable/neutre/défavorable, mais ne disent rien, ou pas grand-chose, sur les contenus qui leur sont associés.

Dès lors, cet article va tout d'abord chercher à identifier et à déterminer les différents thèmes dont les utilisateurs de Twitter ont pu débattre lors de la précédente élection présidentielle. Ce travail se propose ainsi de contribuer à démontrer que les échanges tendent à être organisés autour de thèmes clés, structurants et clivants, qui ont émergé au fil des événements de la campagne. Ces événements, tels que le Pénélope Gate, sont bien souvent des événements « disruptifs » (Katz et Liebes, 2007), non programmés, qui viennent faire la Une de l'actualité journalistique. Ils viennent alors s'inscrire dans un traitement journalistique de l'actualité politique qui rythme à bien des égards, et c'est là l'une des principales hypothèses de ce travail, les volumes des échanges et les thèmes débattus à un instant T dans Twitter. Cet article montrera ensuite que la réception de ces événements, puis la manière dont ils vont être traités dans Twitter, varient considérablement d'un groupe d'acteurs à l'autre, en fonction de leur appartenance politique. C'est du moins l'hypothèse que nous faisons. Les opinions et les représentations sociales que les acteurs ont construites, ou réactualisées, au fil des événements qui ont rythmé la campagne, dépendent de leurs filiations politiques et de leur choix de candidat. Twitter forme alors un parfait observatoire pour comprendre, sur la base de la collecte massive de discours et leurs analyses assistées par ordinateur, comment des groupes de sympathisants politiques ont produit des représentations sociales différentes de certains faits, de certains personnages politiques,

comme Emmanuel Macron, ou plus généralement des hommes politiques et de leurs manières d'exercer le pouvoir.

De façon à pouvoir déterminer ce qui rapproche et différencie les partisans de chacun des candidats, c'est-à-dire de façon à pouvoir décrire en quoi leur discours est potentiellement idéologiquement orienté, nous réaliserons nos analyses groupe par groupe. Nous utiliserons une méthode lexicale qui a effectivement pour objectif de déterminer les différentes thématiques abordées. Dans ce cadre, une thématique est définie comme « un ensemble de formes pleines contextuelles liées entre elles par leur objet et leur contexte » (Ratinaud et Marchand, 2015, p. 57). L'inscription d'un objet dans la pensée sociale particulière d'un groupe s'exprime donc par l'utilisation d'un lexique spécifique à son propos, ce que traduit le concept de « marqueur socio-langagier » (Scherer et Giles, 1979 ; Mange, 2016). Quand l'orientation idéologique des groupes considérés est forte, leur perception de certains objets est très différente. Ces différences se manifestent dans les discours par l'emploi d'un lexique propre à chaque groupe et/ou par l'ancrage de l'objet dans des univers lexicaux distincts. Par ailleurs, une approche par thématique permet également de comparer ce qui fait « enjeu » pour chacun des groupes en repérant des thématiques ou des objets qui sont abordés par certains, mais pas par d'autres.

DÉMARCHE MÉTHODOLOGIQUE

La démarche méthodologique que nous suivrons s'inscrit donc dans ce cadre théorique. Chacune des étapes que nous présenterons (l'échantillonnage, la recherche des communautés de partisans des cinq principaux candidats à l'élection et la comparaison des contenus de leurs tweets) est guidée par la volonté de mettre en évidence des proximités et des différences dans les objets mobilisés et dans leur description. Nous aborderons les différentes méthodologies utilisées dans chacune de ces étapes au fur et à mesure de leur emploi, de façon à ne pas alourdir la lecture. Nous souhaitons simplement résumer ici ces étapes en rappelant la logique qui les unit et l'intérêt qu'elles présentent pour l'étude des représentations sociales et de leur inscription idéologique. Classiquement, ce type d'étude se réalise sur la base de questionnaires ou d'entretiens semi-directifs qui servent à déterminer le contenu des représentations sociales étudiées et les positionnements des sujets sur ces contenus. Les résultats peuvent alors être contrastés en fonction des réponses des sujets à des

échelles de positionnement politique par exemple (Mari *et al.*, 2017), ou par l'appartenance déclarée des sujets à des groupes que le chercheur considère comme idéologiquement différents (Moscovici, 1961). Il est aussi possible de déterminer des groupes aux positionnements axiologiques antagonistes à partir de réponses à des opinions catégorisées *a priori* comme relevant de telle ou telle idéologie (Ratinaud, 2003).

Pour étudier les discours diffusés via Twitter pendant la campagne présidentielle de 2017, nous avons débuté par l'indexation d'un corpus de tweets relatifs à cette campagne (voir Échantillonnage). Twitter permet de mettre un peu plus à disposition un recueil de données qui présente des caractéristiques originales en comparaison des méthodes classiques d'étude des représentations sociales. Sans prendre en considération les éventuelles difficultés techniques liées à ce type d'étude, ce mode de recueil largement automatisé apparaît beaucoup moins chronophage que les enquêtes par questionnaires et par entretiens. Il permet en fait de conserver le temps nécessaire à ces enquêtes tout en proposant un regard complémentaire sur les objets étudiés. Dit autrement, il nous semble que le recueil de données sur ce réseau socio-numérique représente un outil pertinent dans des démarches de triangulations méthodologiques appliquées à l'étude des représentations sociales (Apostolidis, 2003 ; Caillaud et Flick, 2016). Par ailleurs, ce type de recueil lève également une limite inhérente aux enquêtes classiques : quand les méthodes classiques proposent une photographie de la représentation de l'objet portée par les groupes au moment de l'enquête, Twitter permet d'avoir l'intégralité des images du film sur l'intégralité de la période considérée.

Dans un premier temps, nous chercherons à identifier au sein du corpus de tweets les différentes communautés de soutien de chacun des cinq principaux candidats à l'élection (voir Détermination des communautés). C'est ici un second avantage majeur de ce type de données sur les méthodes classiques. La théorie des représentations sociales postule en effet que les représentations se construisent majoritairement dans les communications interpersonnelles. Cette dimension est généralement postulée, tout comme l'intérêt des enquêtés pour l'objet. L'appartenance du sujet à la catégorie sociale cible est souvent le seul indicateur utilisé comme marqueur de l'appartenance à une idéologie. Dans le cas de cette étude, les groupes que nous considérons sont des groupes « constatés » sur la base d'interactions réelles représentées par la pratique du retweet.

L'étape suivante consiste à réunir les tweets originaux de ces cinq communautés pour former cinq corpus lexicaux que nous soumettons à une analyse lexicométrique avec la méthode Reinert (voir Analyse lexicale du contenu des tweets). La spontanéité de ces textes, comparée aux données obtenues par questionnaires et entretiens, est à souligner. C'est également l'un des avantages majeurs de ce type de recueil. Dans les enquêtes classiques, la plupart du temps, le chercheur provoque les données qu'il analyse. Elles sont le fruit de la sollicitation par le chercheur d'un individu et de l'acceptation par cet individu de participer à l'enquête. Le contrat de communication qui lie alors le chercheur et l'enquêté est d'une nature particulière (Salès-Wuillemin, 2011), susceptible d'orienter une partie de son discours. Les travaux sur les « zones muettes » des représentations sociales attestent aussi du poids de l'interaction « enquêteur/enquêté » dans les éléments mobilisés et/ou verbalisés (Guimelli et Deschamp, 2000). De ce point de vue, les données provenant de Twitter ont une spontanéité qui doit être perçue comme un gage de qualité. Personne n'a sollicité les textes indexés : leurs auteurs les ont produits spontanément. Ils n'ont par ailleurs pas conscience que ces propos participent à une enquête. Cependant, il ne faut pas perdre de vue le fait que les locuteurs qui s'expriment sur Twitter ont conscience d'être sur un espace d'autopublication et de mise en scène de soi, ce qui influe sur la forme et le contenu de leur discours (Granjon et Denouël, 2010).

Nous comparerons ensuite, sur la base de quelques exemples, les résultats des analyses lexicales entre les communautés. Enfin, nous proposerons une lecture globale de l'ensemble des thématiques, d'une part sous leur aspect chronologique (voir Chronologie des classes de discours), et d'autre part sous leur aspect lexical (voir Résumé des relations lexicales entre les classes). Les méthodes que nous utilisons pour déterminer les communautés et les thématiques relèvent pour nous de ce que Venturini et Latour (2010) nomment une approche « quali-quantitative », dans le sens où elles font intervenir des méthodes statistiques sur des grands ensembles de données, mais sont aussi dépendantes, en dernière instance, de l'interprétation par les chercheurs des résultats qu'elles produisent.

ÉCHANTILLONNAGE

Dans l'objectif d'étudier les informations qui circulent sur Twitter pendant la campagne présidentielle, nous avons opté pour un échantillonnage par

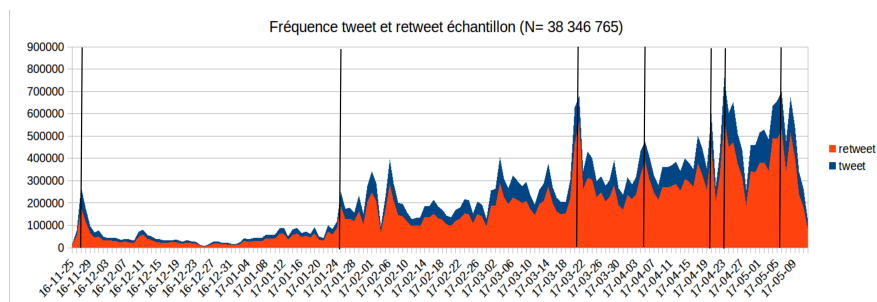
mots clés afin de sélectionner et de collecter un corpus de tweets à l'aide du logiciel libre DMI-TCAT (Borra et Rieder, 2014) qui a été installé sur un des serveurs de l'IRIT². DMI-TCAT utilise l'API de Twitter. Cette stratégie de collecte de tweets ne permet pas de garantir l'exhaustivité du corpus (González-Bailón *et al.*, 2012) pour autant elle est très répandue et les conséquences de biais éventuels sont compensées par notre approche comparative. L'étude s'est centrée sur les cinq principaux candidats (Jean-Luc Mélenchon (LFI), Benoît Hamon (PS), Emmanuel Macron (EM), François Fillon (LR) et Marine Le Pen (FN)) qui ont réuni plus de 90 % des suffrages au premier tour. L'indexation de ces tweets a commencé le 25 novembre 2016 à partir de mots clés renvoyant aux noms des candidats déclarés ou supposés et aux noms de leur parti politique d'appartenance. Cette liste de mots clés ou de hashtags a ensuite été constamment tenue à jour en fonction des événements de la campagne (résultats des primaires, événements médiatiques...). Nous avons arrêté l'indexation de ces tweets le 12 mai 2017, soit cinq jours après le second tour de l'élection. Ce premier échantillon, qui court sur 168 jours, contenait 42 251 431 tweets. Nous avons réduit notre collection en effectuant une requête qui ne comprenait que les noms des candidats, de leur parti ou certains *hashtags* liés à la campagne, comme les slogans des candidats ou des *hashtags* généralistes comme « présidentielle2017 » (avec ou sans accent, au singulier et au pluriel)³. Cette démarche permet de nous assurer que les cinq candidats reçoivent un traitement identique dans notre analyse en venant rééquilibrer une liste de mots-clés qui pouvait présenter des disparités dues à la disproportion de l'actualité de chacun des candidats pendant la campagne (par exemple, l'affaire de l'emploi fictif de Pénélope Fillon a généré beaucoup de *hashtags* spécifiques). Après cette requête, l'échantillon se compose de 38 346 765 tweets émis par 2 163 812 comptes différents.

2. La collecte de données a été effectuée à l'aide de la plateforme OSIRIM qui est administrée par l'Institut de Recherche en Informatique de Toulouse (IRIT) et soutenue par le CNRS, la région Occitanie, le gouvernement français et le FEDER (<http://osirim.irit.fr/site/>, consulté le 9 avril 2019).

3. La liste retenue est la suivante : “macron”, “enmarche”, “lafrancenmarche”, “fillon”, “lesr_publicains”, “#lr”, “mlp”, “lepen”, “le pen”, “marine2017”, “#fn”, “aunomdupesuple”, “francebleumarine”, “franceenordre”, “hamon”, “#ps”, “futurdesirable”, “fairebattrelecoeurdelafrance”, “m_lenchon”, “jlm”, “avenirencommun”, “humaindabord”, “#insoumis”, “#fi”, “#lfi”, “pr_sidentielle2017”, “pr_sidentielles2017”; le tiret bas () permet d'envisager les mots clés concernés avec et sans accent.

Le graphique 1 présente la fréquence des tweets et des retweets par jour d'indexation :

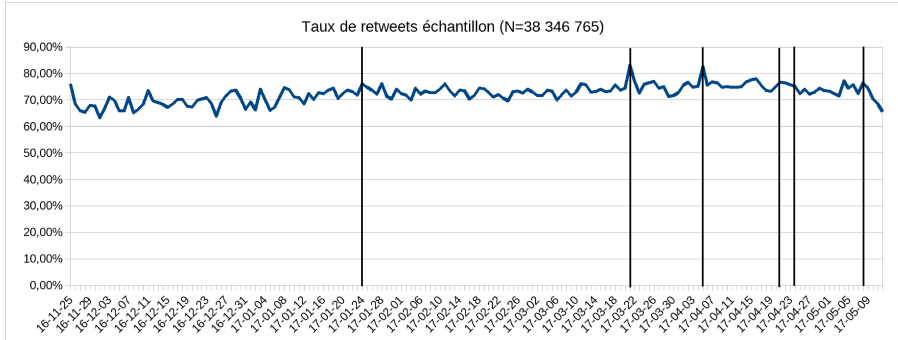
Graphique 1. Fréquence des tweets (gris foncé) et des retweets (gris clair) par jour



Source : auteurs.

Nous avons indiqué dans cette distribution chronologique quelques événements représentés par des lignes verticales noires. La première correspond au second tour de la primaire de la droite et du centre qui a vu la victoire de François Fillon ; la deuxième au second tour de la primaire citoyenne qui a vu la victoire de Benoît Hamon ; les trois suivantes correspondent à des débats télévisés et les deux dernières aux premier et second tours de l'élection. Cette distribution permet de constater que sur Twitter, la campagne semble se lancer à la fin du mois de janvier, après la publication par *Le Canard enchaîné* d'un article concernant Pénélope Fillon (le 24 janvier) et la primaire citoyenne (les 22 et 29 janvier). Elle montre également que les événements répertoriés correspondent à des pics de communication et que la production de tweets augmente régulièrement jusqu'au second tour de l'élection.

Dans le graphique 2, les lignes verticales noires sont positionnées aux mêmes dates que dans le graphique 1. Nous constatons ici que le taux de retweets augmente également légèrement avec le nombre de tweets (il varie de 65 % en début de campagne à 75 % en fin de campagne, avec des pics à plus de 80 % lors des débats télévisés). Le taux moyen de retweets sur l'ensemble de l'échantillon est de 73,79%.

Graphique 2. Taux de retweets par jour

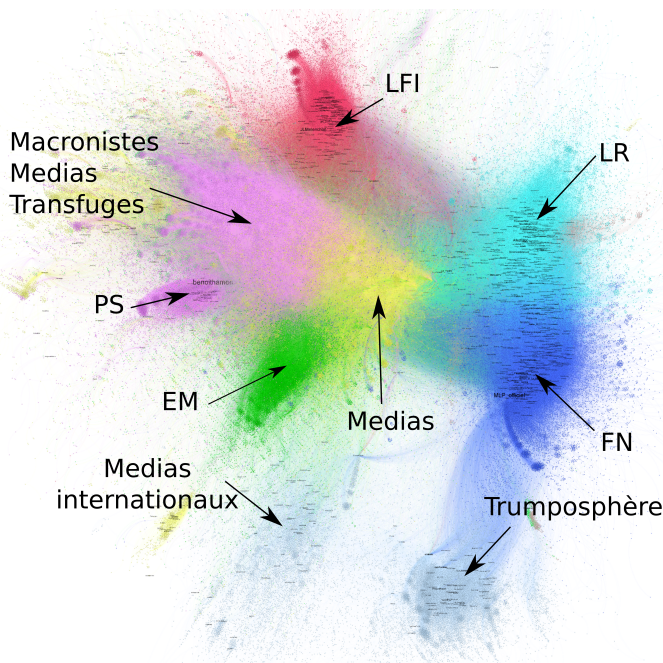
Source : auteurs.

Détermination des communautés

La première étape de notre analyse consiste à construire le graphe dirigé des retweets à partir de l'échantillon. Dans ce processus, les comptes Twitter sont traités comme les nœuds d'un graphe et lorsqu'un compte en retweete un autre, nous générons une arête du graphe (c'est-à-dire un lien entre les deux comptes). Ce lien est dirigé, dans le sens où un retweet d'un compte A concernant un tweet émis par un compte B est distinct d'un retweet du compte B concernant un tweet du compte A. À chacun de ces liens est associée une valeur qui correspond au nombre de retweets entre les deux comptes dans chacune des deux directions possibles. En raison de limitations matérielles (les graphes générés sur de telles quantités de tweets sont énormes), nous avons dû restreindre cette analyse aux comptes ayant émis au moins 10 tweets ou retweets dans l'échantillon, ce qui nous permet de nous focaliser sur les comptes les plus actifs dans les débats autour de la campagne présidentielle. Par ailleurs, dans le graphe que nous présentons, les liens d'une valeur inférieure à 3 ont été retirés. Ce choix de seuil a également pour objectif de réduire la taille du graphe pour qu'il soit concrètement analysable sur un ordinateur « classique ». Une fois le graphe calculé, il est visualisé à l'aide du logiciel libre Gephi (Bastian, Heymann et Jacomy, 2009). Nous lui avons appliqué l'algorithme de positionnement ForceAtlas2 (Jacomy *et al.*, 2014) pour obtenir une spatialisation des comptes qui prenne en considération la force des liens entre les sommets. La topologie du graphe est ainsi le résultat de l'intensité de l'interaction entre les comptes (plus deux comptes

sont caractérisés par une communication bidirectionnelle intense entre eux, plus ils sont proches dans le graphe). La dernière partie de cette première étape consiste à rechercher les différentes communautés à l'intérieur de ce graphe à partir de l'algorithme de Louvain implémenté dans Gephi (Blondel *et al.*, 2008). Cette procédure détermine des ensembles de sommets qui ont tendance à être fortement reliés entre eux. Autrement dit, la procédure signale des groupes de comptes Twitter qui se retweetent beaucoup mutuellement. Ce sont ces groupes que nous nommons *communautés*. Enfin, nous avons associé ces communautés de comptes aux partis politiques de chaque candidat en étudiant manuellement les comptes les plus retweetés en leur sein. Les résultats de ce traitement sont présentés dans le graphique 2. Le graphe représenté contient 159 325 comptes et 1 452 899 liens.

Graphique 3. Graphe des retweets obtenu à partir des comptes ayant émis au moins 10 tweets ou retweets ; les liens d'une valeur < 3 ont été éliminés.



Source : auteurs.

Cette manière de procéder pour qualifier politiquement des ensembles de comptes Twitter à partir d'une caractérisation manuelle des comptes les plus cités est évidemment critiquable d'un point de vue méthodologique. Cependant, elle a montré sa pertinence dans les cas que nous avons étudiés par le passé (par exemple Smyrniotis et Ratinaud, 2014, 2017). Plus important encore, elle a été confirmée par une analyse parallèle, et effectuée de manière indépendante, qui a consisté à identifier manuellement, de manière qualitative et en suivant un protocole contrôlé, l'appartenance politique de plusieurs milliers de comptes issus notre échantillon initial (Fraisier *et al.*, 2018). Le tableau 1 présente le croisement entre cette catégorisation manuelle et la catégorisation proposée par l'algorithme ForceAtlas2 :

Tableau 1. Croisement entre les communautés indexées manuellement et les communautés détectées automatiquement

	em	fn	fi	lr	ps
taille communauté	6357	17477	12615	14150	13628
indexés manuellement	4512	3987	5886	4733	2195
em	2328	5	6	22	10
fn	25	2697	70	135	10
fi	11	10	3088	10	18
lr	30	72	17	3084	0
ps	33	2	76	5	1355
taux de recouvrement de l'ensemble des comptes catégorisés manuellement	51.60%	67.64%	52.46%	65.16%	61.73%
taux de recouvrement des comptes Catégorisés manuellement présents dans le graphe	95.92%	96.81%	94.81%	94.72%	97.27%

Source : auteurs.

Les communautés politiques détectées par l'algorithme (ligne « taille communauté ») permettent de classer 64 227 comptes. Cet échantillon représente seulement 2,97 % des comptes du corpus total, mais ces utilisateurs sont à l'origine de 37 % de tweets originaux émis. Ce résultat souligne encore une fois que la méthode cible les utilisateurs les plus actifs dans ce contexte. La catégorisation manuelle a concerné 21 313 comptes (ligne « indexés manuellement »). Elle consiste essentiellement à parcourir les informations disponibles sur le compte pour lui attribuer une préférence politique (Fraisier *et al.*, 2018). Le tableau montre par exemple que sur les 4 515 comptes identifiés manuellement comme des soutiens d'Emmanuel Macron (ligne « em »), 2 328

sont présents dans la communauté « em » détectée par l'algorithme, contre 25 dans la communauté FN ou 33 dans la communauté PS. Cela correspond à un taux de recouvrement de 51,6 % (ligne « taux de recouvrement de l'ensemble des comptes catégorisés manuellement »). Si l'on ne prend en considération que les comptes présents dans notre graphe réduit aux communautés politiques, nous retrouvons 95,92 % de ces comptes indexés manuellement dans la communauté de soutien « em » (ligne « taux de recouvrement des comptes catégorisés manuellement présents dans le graphe »). Pour chacune de nos communautés politiques, le taux de recouvrement entre la catégorisation politique manuelle et la détection automatique des communautés se situe entre 94,7 % et 97,2 %.

Une précision également importante quant au processus de détermination des communautés est que le graphe représenté dans le graphique 2 n'est pas celui qui a été obtenu par l'analyse en première intention. En effet, lors de l'analyse initiale, la communauté formée autour d'Emmanuel Macron se confondait totalement avec celle des médias *mainstream* (presse quotidienne nationale, chaînes et radios d'information notamment). Autrement dit, l'algorithme a détecté un tel nombre de liens (retweets) entre les comptes faisant partie du réseau du candidat d'En Marche et ceux des principaux médias français qu'il les a classés automatiquement ensemble. C'est la première fois que nous observons un tel phénomène. Dans toutes les autres analyses similaires que nous avons pu effectuer par le passé sur des questions de nature politique, la communauté des médias était facilement distinguable par l'algorithme de détection, car ceux-ci sont retweetés par ou retweetent des comptes provenant de tous les bords politiques. On peut émettre deux hypothèses qui expliquent cette proximité, la première de nature politique et la seconde de nature méthodologique. L'hypothèse de nature politique est celle d'une certaine surreprésentation médiatique d'Emmanuel Macron dans la couverture de la campagne attestée par de nombreuses sources⁴. L'hypothèse méthodologique est liée au choix d'arrêter la collecte des données cinq jours après le premier tour. Il paraîtrait alors assez logique que le vainqueur de l'élection ait vu ses tweets et ceux de ses soutiens suffisamment relayés par les médias pour générer l'effet observé. Quoi qu'il en soit, afin de bien délimiter la communauté formée autour d'Emmanuel Macron, nous avons soumis la communauté Macron-médias à l'algorithme de Louvain ce qui nous a permis de la séparer en trois

4. Pour une synthèse, voir Smyrniotis, Nikos, « How the mainstream media created the Macron phenomenon », 9 mai 2017, <http://ephemeron.eu/1981>, consulté le 9 avril 2019.

communautés, dont une correspond clairement aux médias *mainstream* et une autre aux partisans d'Emmanuel Macron. Dans la suite de ce travail, lorsque nous ferons référence à la communauté des partisans de Macron (notée EM), nous renverrons à celle présentée dans le graphique 2.

Description des communautés

D'un point de vue quantitatif, les communautés que nous avons constituées autour des cinq principaux candidats vont de 6 357 comptes pour la plus petite (EM) à 17 477 pour la plus volumineuse (FN). Les trois autres sont situées dans une fourchette assez réduite (12 615 pour LFI, 13 628 pour le PS et 14 150 pour LR). Nous noterons que la structuration de ces communautés est très proche de celle décrite par les auteurs du Politoscope (Gaumont, Panah et Chavalarias, 2018 ; Chavalarias, Gaumont et Panahi, 2019). Il faut toutefois souligner que l'échantillonnage utilisé dans ces travaux diffère sensiblement de celui que nous présentons. D'une part, la période d'indexation du Politoscope est plus longue (11 mois contre environ 6 dans cette étude), d'autre part, nous n'avons utilisé que des mots clés pour sélectionner les tweets alors que les données du Politoscope proviennent à la fois du suivi des comptes de plus de 3 000 politiciens et d'une indexation par une douzaine de mots clés.

Comme expliqué précédemment, la particularité de la communauté initiale autour d'Emmanuel Macron était la présence en tête du classement des retweets des principaux médias français (BFMTV, Le Monde, Europe 1, France Info, l'AFP, Le Journal du Dimanche, RMC, TF1, LCI, Le Parisien, L'Express, RTL, Marianne, etc.) aux côtés des comptes officiels du candidat Macron. Si on extrait de cette communauté les comptes médiatiques, on s'aperçoit que le candidat Macron est essentiellement soutenu sur Twitter pendant la campagne par une multitude de comptes satellites de son mouvement (Jeunes Macron, TeamMacron2017, enmarchefr, E_Macron2017), ce qui peut être le signe d'une stratégie de communication bien réfléchie et exécutée professionnellement ou le signe d'une certaine forme d'isolement politique de ce candidat. Ces comptes sont complétés par ceux d'acteurs politiques à la fois en provenance du centre, de la droite et du PS (François Bayrou, Christophe Castaner, Richard Ferrand, Aurore Bergé par exemple) qui intégreront par la suite le gouvernement d'Édouard Philippe. En revanche, comme pour le groupe de François Fillon, la présence de militants et de citoyens lambda parmi les comptes très retweetés est faible.

La communauté de LR se structure autour des comptes officiels du candidat François Fillon et des Républicains qui sont accompagnés par des soutiens politiques de premier plan qui ont une forte présence sur Twitter (Éric Ciotti, Bruno Retailleau, Valérie Boyer, Thierry Mariani, etc.). Ce qui est frappant dans cette communauté c'est la présence des comptes de médias classés à droite (Le Figaro, L'Opinion, Valeurs Actuelles, Atlantico) ainsi que des journalistes avec une forte présence dans les médias audiovisuels (Bruno Jeudy, Eugénie Bastié, Judith Waintraub). Ces observations montrent que l'inscription politique d'un média ou d'un journaliste peut être suffisamment forte pour que son compte soit associé à ceux des autres militants du parti. Nous noterons également que la présence de militants et de citoyens lambda parmi les comptes très retweetés est faible.

La communauté du PS est structurée autour du compte officiel du candidat Benoît Hamon et de ses soutiens (Christiane Taubira, Anne Hidalgo, Gérard Filoche, Yves Jadot, Najat Vallaud-Belkacem, Cécile Duflot, Jérôme Guedj, etc). Dans le centre du graphe, on trouve également des comptes officiels créés à l'occasion de la campagne (Jeunes avec Hamon, Avec Hamon 2017) ou affiliés au PS (Parti socialiste) ainsi que des militants ou des sympathisants. Enfin, un certain nombre de médias du service public ou estampillés de gauche sont également liés à cette communauté (France Inter, Francetvinfo, Alternatives Économiques, France 2). Ce qui est assez frappant c'est l'absence des cadres du Parti Socialiste, notamment des anciens ministres de François Hollande, à quelques exceptions près. Cette absence traduit concrètement le manque de soutien à la candidature de Hamon d'une partie de l'appareil socialiste qui a rejoint la candidature d'Emmanuel Macron.

La communauté formée autour de Marine Le Pen est la plus conséquente. En son centre nous trouvons le compte officiel de Marine Le Pen qui détient le plus grand nombre de retweets (318 153) suivi par le compte de Florian Philippot (104 923) et les autres cadres du FN (Marion Le Pen, David Rachline, Nicolas Bay, Gilbert Collard, Robert Ménard, etc.). Parmi les comptes les plus retweetés se trouvent des supporteurs déclarés du FN (Fandetv, Fdesouche), des comptes situés à la droite extrême, mais ne s'identifiant pas uniquement au FN (PtdArcole, David_VanH, Damocles_fr), des hommes politiques de la droite souverainiste (Nicolas Dupont-Aignan) et gaulliste (Henri Guaino). La communauté FN regroupe également des comptes populaires engagés à l'extrême droite au ton satirique/humoristique comme KimJongUnique (soutien du régime russe) et Tprincedelamour, qui s'est illustré notamment en lançant sur Twitter une fausse nouvelle parmi les plus répandues du début de la

campagne⁵. Enfin, cette communauté contient également des individus et des organisations qui, bien que classés politiquement à l'extrême droite, ne sont pas totalement alignés avec le Front National, ainsi que des médias liés à la Russie : Russia Today en français ; Boulevard Voltaire, site complotiste fondé par Thierry Meyssan ; Sputnik (agence russe d'information en ligne) ; Riposte Laïque (organisation au discours islamophobe) et EetR_national (Égalité et Réconciliation, l'organisation d'Alain Soral).

Nous avons également signalé dans le graphique 2 trois autres groupes de comptes remarquables. La communauté notée « Macronistes, médias, transfuges » s'organise autour de soutiens tardifs du candidat d'En Marche ainsi que de certains comptes médiatiques (elle faisait partie de la grande communauté Macron-médias dans notre première analyse). Les groupes notés « médias internationaux » et « Trumposphère » sont détectés par l'algorithme comme une seule communauté. La « Trumposphère » est constituée du compte de Donald Trump et de comptes de l'extrême droite américaine.

Analyse lexicale du contenu des tweets

Pour chacune de ces cinq communautés, nous avons construit un corpus des tweets originaux (ce qui revient à éliminer les retweets). Le tableau 1 décrit succinctement chacun de ces corpus ainsi que les paramètres retenus pour l'analyse. Au total, 3,7 millions de tweets sont analysés, ce qui représente environ 37 % des tweets originaux présents dans l'échantillon.

Tableau 2. Description des corpus de tweets pour les analyses lexicales

	tweets originaux	classes phase 1	formes	taille classe min.	nombre de classes	% segments
lr	1095102	200	10000	20000	12	84.35%
fn	924890	200	10000	20000	10	78.95%
lfi	918491	200	10000	20000	13	79.71%
ps	405517	100	10000	10000	11	92.97%
em	385046	100	10000	10000	10	93.92%

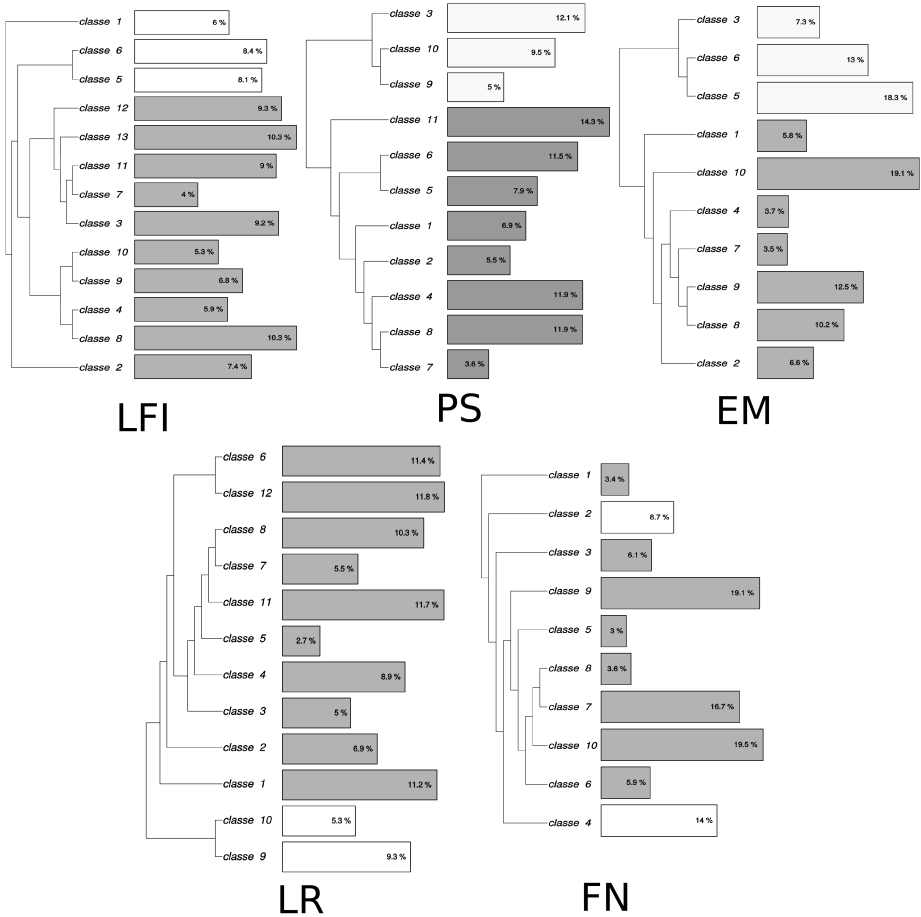
Source : auteurs.

5. <https://crosscheck.firstdraftnews.com/checked-french/etranger-hopital-agression/>, consulté le 9 avril 2019.

Ces corpus sont analysés avec la méthode Reinert (Reinert, 1983, 1990) implémentée dans le logiciel libre IRaMuTeQ (Ratinaud, 2014 ; Ratinaud et Marchand, 2012). Cette méthode permet de déterminer les différentes thématiques qui structurent un corpus textuel. Elle repose sur une classification hiérarchique descendante qui peut être décrite comme une succession de bipartitions reposant sur une analyse factorielle des correspondances. La première étape de cette analyse est un prétraitement lexical. Nous avons dans un premier temps retiré les URL des tweets. Puis le contenu textuel a été lemmatisé afin de réduire la variabilité lexicale. Ce traitement consiste à ramener les mots à la forme qu'ils auraient dans un dictionnaire : les noms sont mis au singulier, les adjectifs au masculin singulier et les verbes à l'infinitif. Dans la suite de l'analyse, seules les formes pleines (les noms, les adjectifs, les adverbes et les verbes) seront utilisées. Les mots outils (les pronoms, les auxiliaires, les conjonctions...) seront utilisés en variables supplémentaires et donc réinjectés dans les résultats après la classification. La classification proprement dite intervient sur la matrice de présence/absence qui croise les tweets et les formes pleines. Cette matrice présente donc les tweets en ligne et les formes pleines en colonne. On note 1 dans la matrice lorsqu'une forme est présente dans un tweet et 0 lorsqu'elle est absente. Le processus de classification consiste à opérer une série de coupures en deux des lignes de cette matrice de façon à réunir les tweets qui ont tendance à contenir les mêmes mots. Cette étape repose sur une analyse factorielle des correspondances dont le premier facteur est utilisé pour déterminer deux groupes de tweets qui maximisent l'homogénéité interne des groupes et l'hétérogénéité entre les deux groupes. Après chaque partition, la matrice correspondant au plus grand des groupes déjà constitués va subir le même traitement. Ces opérations se répètent jusqu'à obtention du nombre de classes terminales paramétré par l'utilisateur. Une fois les classes terminales obtenues, elles sont décrites à partir du lexique qui les caractérise. Ce lexique est constitué des mots qui sont significativement surreprésentés dans la classe si on la compare à l'ensemble des autres classes (sur la base d'un Chi2). Les « mondes lexicaux » (Reinert, 1990) qui se dégagent présentent alors les différentes thématiques abordées dans le corpus.

Résultats des analyses lexicales

Graphiques 4, 5, 6, 7 et 8. Dendrogrammes et tailles des classes pour les corpus de la France Insoumise, du Parti Socialiste, d'En Marche, des Républicains et du Front National



Source : auteurs.

Le premier élément d'interprétation à la lecture de ces résultats concerne la structuration de ces analyses. Quatre de ces dendrogrammes présentent une coupure principale qui sépare des classes de discours liées à la présentation des programmes et des agendas médiatiques des candidats (en blanc) de toutes les autres (en gris)⁶. Pour LFI, il s'agit des classes 1, 6 et 5 ; pour le Parti Socialiste ce sont les classes 3, 10 et 9 ; pour En Marche les classes 3, 6 et 5 et pour Les Républicains, les classes 9 et 10. Le corpus du Front National présente bien des classes lexicales relevant de ces thématiques (les classes 2 et 4), mais elles n'ont pas le même positionnement.

Une autre caractéristique de ces classes, commune à l'ensemble de l'analyse, est le fait qu'elles portent toutes une surreprésentation des *hashtags* « officiels » de chacun des candidats. Ces classes sont les seules à aborder des éléments liés aux programmes. Nous constatons donc que les informations qui circulent sur Twitter en temps de campagne n'abordent le contenu des programmes politiques que de façon marginale. La majorité des conversations porte sur autre chose. Comme nous ne pouvons pas présenter l'intégralité du contenu de ces 56 classes, nous centrerons notre interprétation sur quelques exemples caractéristiques.

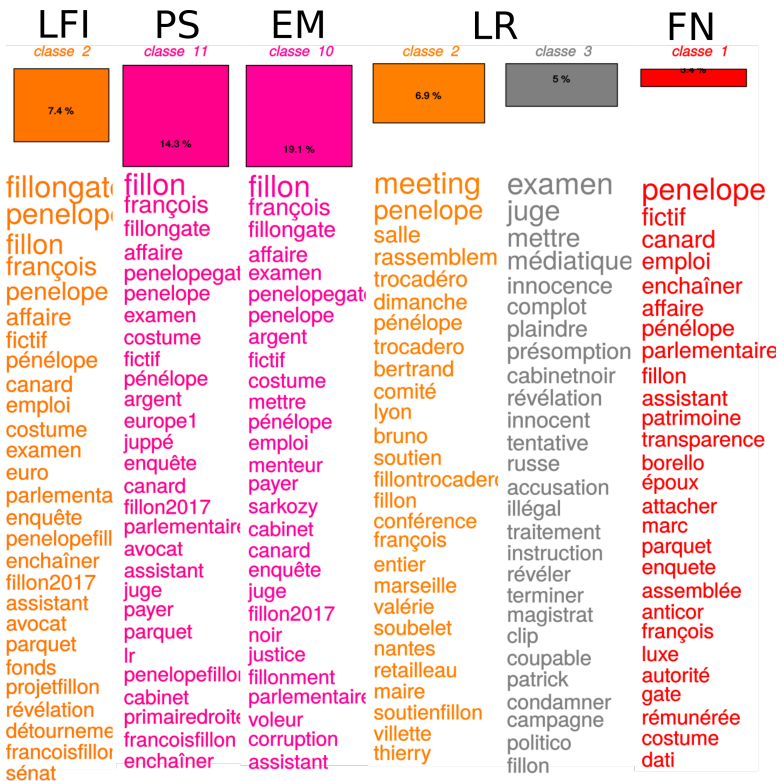
Le PénélopeGate

À l'exception du corpus des Républicains, toutes les analyses contiennent une classe qui concerne les révélations du *Canard enchaîné* au sujet des emplois fictifs dont ont bénéficié l'épouse et les enfants de François Fillon. Rappelons que ce journal a publié le 24 janvier 2017 un article qui accusait François Fillon d'avoir rémunéré son épouse comme attachée parlementaire alors qu'elle ne se serait jamais acquittée de cette tâche. L'ouverture d'une enquête à partir de ces révélations a conduit à la mise en examen de François Fillon. Ces événements ont rapidement été référencés au sein de Twitter par les appellations PénélopeGate et FillonGate et ont poursuivi ce candidat pendant toute la campagne. L'analyse montre ici que toutes les communautés se sont emparées de cette thématique autour d'un lexique relativement homogène. Comme un écho à ces classes, nous retrouvons dans l'analyse du corpus des Républicains des classes qui lui sont spécifiques et qui concernent les différents événements mis en place par l'équipe de campagne de François Fillon pour amortir ce choc, comme le grand meeting du Trocadéro auquel participa

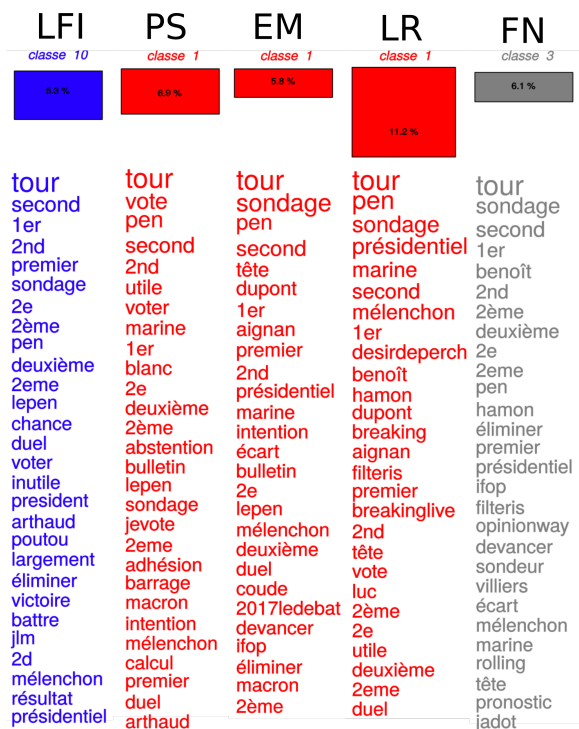
6. La numérotation des classes est indépendante de leur contenu. Ces numérotations ne peuvent pas être comparées entre différentes classifications.

Pénélope Fillon (classe 2) ; ou des éléments de communication dans lesquels les Républicains dénoncent un complot et mettent en avant l'innocence de leur candidat (classe 3). Ce premier exemple montre clairement une différence entre les représentations sociales que les adversaires et les partisans de François Fillon ont développées de cet événement. Le discours des adversaires, de gauche et de droite, est ici relativement homogène et ne fait qu'amplifier sur Twitter une polémique qui existe par ailleurs au même moment dans la presse traditionnelle et dans les journaux télévisés. Les partisans perçoivent bien évidemment la situation tout autrement : ils décrivent François Fillon comme la victime innocente d'un complot politico-médiatique. Ce constat forme donc un exemple de l'effet d'un positionnement idéologique sur la construction de la représentation sociale d'un événement médiatique.

Graphique 9. Profils des classes en rapport avec l'affaire de l'emploi fictif de l'épouse de François Fillon



Source : auteurs.

*Les commentaires des sondages et des résultats***Graphique 10. Profils des classes qui commentent les sondages et les résultats**

Source : auteurs.

Nous retrouvons dans chacune des analyses une classe de discours liée aux commentaires des sondages d'opinion et aux résultats des primaires et du premier tour de l'élection. Nous verrons que ces discours sont fortement régulés chronologiquement, dans le sens où ils elles sont surreprésentées au même moment dans chacune des analyses, soit après les primaires, soit après le premier tour. C'est un exemple du poids du contexte dans les discours produits : une partie importante des thématiques que nous mettons en évidence ne sont que la conséquence d'événements extérieurs commentés sur Twitter. Par ailleurs, la systématité de cette thématique, qui apparaît dans toutes les communautés, montre comment les sondages et les résultats d'élection participent à entretenir les flux conversationnels en jouant un rôle de déclencheur dans la production de message.

Insultes et dénigrements

Graphique 11. Profils des classes « Insultes et dénigrement »



Source : auteurs.

Ces classes de discours présentent également un lexique très proche. Leur profil contient le verbe *aller* qui est particulièrement caractéristique du discours oral auquel on peut comparer certaines communications écrites en ligne (Matuszak, 2007). On notera également la présence de la forme *ca* qui est une variation orthographique du pronom « ça » (et donc non reconnu par le logiciel) et qui est assurément la conséquence d'un usage non professionnel du clavier. On note également dans ces profils des marqueurs de la culture du SMS et des communications en ligne, avec par exemple les expressions *mdr* et *lol*. Mais ce qui caractérise le plus ces classes s'avère être la présence d'un lexique familier ou vulgaire (*foutre*, *putain*, *con*, *gueule*, *merde*, *cul*...) et des verbes ou des

adjectifs que l'on retrouve dans des expressions comme « *prendre les gens pour des cons* », « *se foutre de notre gueule* » ou « *gros con* ». Le marquage idéologique de cette forme de communication est principalement perceptible dans les cibles de ses messages : *mlp* (Marine Le Pen) pour les partisans d'En Marche et du Parti Socialiste, *macron* (Emmanuel Macron) pour le Front National. Ce type de discours est probablement une des constantes des discussions politiques sur Twitter (et probablement sur tous les réseaux socio-numériques) : nous avons retrouvé des classes équivalentes dans chacune des études qui mobilisent ce type de méthode, y compris dans des corpus en anglais (Smyrniaios et Ratinaud, 2014, 2017 ; Ratinaud, 2016). Ces discours témoignent également d'un certain rapport à la politique. Au-delà du signalement de l'adhésion à un candidat par le rejet de l'autre, ils sont probablement les marqueurs d'une représentation de la politique, des politiciens ou de l'exercice du pouvoir principalement négative et orientée vers la suspicion et le déni de légitimité.

Emmanuel Macron comme objet de représentation sociale

Graphique 12. Profils des classes sur Emmanuel Macron



Source : auteurs.

Dans chacune de nos analyses, à l'exception évidente des partisans d'En Marche, nous trouvons une classe de discours qui décrit Emmanuel Macron. Ces classes forment une parfaite illustration des conséquences d'une inscription idéologique d'un objet sur la genèse de sa représentation. Rappelons qu'avant son accession à la présidence de la République, Emmanuel Macron était perçu comme un acteur politique de second plan. Autrement dit, à moins d'avoir une appétence particulière pour la politique, la plupart des Français le connaissaient essentiellement comme un jeune ministre de l'économie d'un des gouvernements de François Hollande. Durant la campagne, au fur et à mesure que les instituts de sondage commençaient à pointer sa potentielle victoire, tous les groupes concernés ont dû se forger une représentation de ce personnage. Ces classes de discours montrent qu'ils n'en ont visiblement pas construit la même représentation sociale.

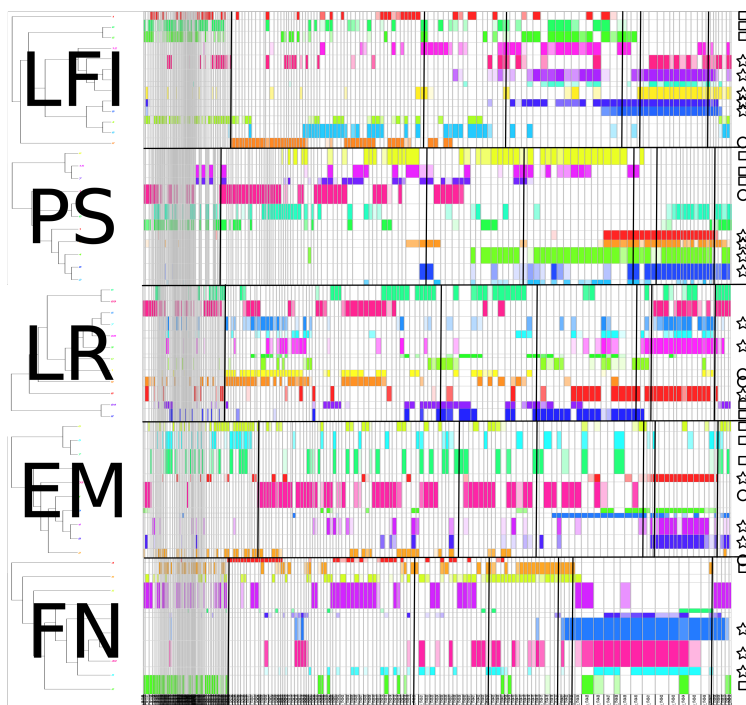
Dans le discours des partisans de Jean-Luc Mélenchon (LFI), Macron est décrit comme un objet *marketing* fabriqué par les *médias*. Il est dépeint comme *vide*, *creux*, à l'image d'une *bulle* ou d'un ballon de *baudruche* qui ne produit que du *vent*. Pour les partisans de Benoît Hamon (PS), Emmanuel Macron est avant tout un homme de *droite*. L'ancrage du personnage de ce côté de l'échiquier politique s'accompagne d'un lexique lié au *libéralisme* : *finance*, *libéral*, *medef*, *banquier*, *patron* par exemple. De façon à insister sur cette lecture du personnage, les partisans du PS ont popularisé un hashtag que nous retrouvons dans le lexique caractéristique de cette classe : *#nidegauchenidegauche*. Dans les résultats du Front National, deux classes de discours font référence à Macron. La première (classe 10) reprend en partie une description proche de celle de LFI : Macron serait un personnage *creux*. Cependant, cette classe insiste beaucoup plus sur les *dangers* associés au personnage : il faut en avoir *peur*. La seconde classe du Front National qui aborde la personnalité de Macron est liée aux « macronleaks » (classe 5), c'est-à-dire à la fuite sur internet d'une partie des mails de certains responsables du mouvement En Marche. Dans cette classe, Emmanuel Macron est décrit comme un *homo-sexuel* qui a des comptes offshores aux *Bahamas* et aux îles *Caimans*. Marine Le Pen a fait référence à ces rumeurs à la fin du débat télévisé de l'entre-deux tours. Il est intéressant de noter que les sympathisants PS classent Emmanuel Macron à droite alors que les partisans de François Fillon (LR) le décrivent comme un *ministre socialiste*, c'est-à-dire comme un homme de gauche. Il est dépeint comme le *fils* ou *l'héritier* de François Hollande et de *Ségolène Royal* qui s'inscrira dans la *continuité* d'un *quinquennat* jugé *catastrophique* ou *désastreux*. On perçoit bien, en filigrane de ces lectures très différentes de

la personnalité d'Emmanuel Macron, le poids de l'ancrage idéologique des locuteurs dans la construction de cette représentation sociale.

Cette première partie de l'analyse lexicale permet donc de mettre en évidence les thématiques abordées pendant la campagne. Elle souligne les différences de perception de certains objets ou événements. Mais comme nous le signalions à propos des classes sur les résultats et les sondages, tous ces discours s'inscrivent dans une temporalité qui semble en grande partie régulée, d'un point de vue quantitatif, par les événements médiatiques et institutionnels qui ont rythmé la campagne.

Chronologie des classes de discours

Le graphique 13 présente une lecture chronologique de la surreprésentation des différentes classes de discours à chacun des jours de l'indexation



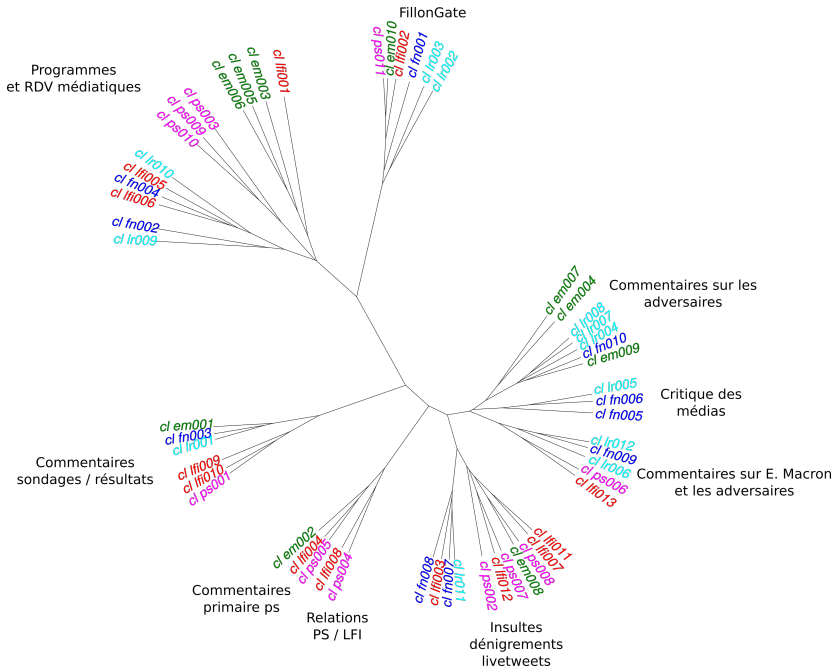
Source : auteurs.

Dans ce graphique, chaque ligne représente l'une des classes de discours des analyses. La hauteur des lignes est proportionnelle à la taille des classes et la largeur des colonnes est proportionnelle au nombre de tweets émis le jour concerné. Lorsqu'une case est grisée (ou coloriée) cela signifie que le type de discours de cette classe est significativement surreprésenté ce jour-là (toujours au sens du χ^2). Les barres noires signalent les mêmes événements que ceux rapportés dans le graphique 1, à l'exception de la première, positionnée le jour de la publication des révélations du *Canard enchaîné* (24 janvier 2017). En première intention, ce graphique traduit l'augmentation du nombre de tweets au fur et à mesure de l'avancée de la campagne (les colonnes sont de plus en plus larges). On note également la très forte activité des partisans de Marine Le Pen (FN) entre les deux tours comparativement aux autres communautés.

Ce graphique montre surtout que la fréquence relative de la plupart de ces classes de discours est liée à des événements médiatiques et institutionnels. On notera alors que les classes relatives à l'affaire de Pénélope Fillon (marquées par des cercles) commencent dès le jour de la parution du *Canard enchaîné* et perdurent pendant environ un mois et demi avant de diminuer d'intensité. Cela se traduit dans le graphique par la coloration des lignes marquées par un cercle à partir de la première ligne verticale. On peut aussi aisément repérer l'effet des débats télévisés et du premier tour de l'élection qui, dans pratiquement toutes les analyses, provoquent la surreprésentation des discours sur les commentaires des résultats, des classes d'insultes et des commentaires sur les candidats (marqués par des étoiles). Dans le graphique, toutes les classes marquées par des étoiles sont coloriées autour de débats télévisés ou entre les deux tours des élections. À l'inverse, les classes sur les programmes (marquées par des carrés) ont toutes tendance à être surreprésentées avant le premier tour, mais perdent en intensité après cet événement. Ceci se traduit dans le graphique par la quasi-absence de coloration des classes marquées par des carrés entre les deux tours. Il ne faut pas perdre de vue que ce graphique ne met en évidence que la surreprésentation des classes à certains moments : la plupart de ces types de discours sont présents tout le long de l'indexation, ce qui varie, c'est la place qu'ils occupent dans une journée de discussion.

Résumé des relations lexicales entre les classes

Graphique 14. Résumé des relations lexicales entre les classes



Source : auteurs.

De façon à résumer les relations de proximité et de distance entre les 56 classes issues des analyses lexicales, nous les avons soumises à une étude complémentaire. Nous avons calculé la distance lexicale entre chacune de ces classes en utilisant la distance de Labbé (Labbé et Monière, 2000 ; Labbé et Labbé, 2003). Cet indice permet d’apprécier dans quelle mesure deux textes sont proches ou distants du point de vue du lexique qu’ils contiennent. Il varie entre 0 et 1. La valeur 0 signale deux textes exactement identiques et la valeur 1 deux textes qui n’ont aucun mot en commun. Comme nous utilisons cet indice pour comparer des classes produites avec la méthode Reinert, nous limitons le calcul aux seules formes pleines (Ratinaud et Marchand, 2016). La matrice de distance qui compare tous les textes 2 à 2 a ensuite subi une classification avec la méthode de Ward pour obtenir une représentation arborée des distances inter-classes. Dans cette représentation, plus les nœuds liant deux classes sont proches, plus les classes présentent un lexique similaire.

Nous retrouvons sur ce graphique les principaux éléments d'interprétation que nous avons parcourus : les proximités entre classes que nous avons déterminées à partir d'une étude qualitative du lexique qui les caractérise se confirment à partir d'un indicateur quantitatif basé sur le lexique. Ce graphique permet également de produire une lecture synthétique/synoptique de l'ensemble des thématiques identifiées dans ce corpus. Nous noterons donc la présence de classes très critiques envers les médias, dans les analyses du FN et de LR, ainsi que des classes liées aux *live-tweets* pendant les débats télévisés. L'élément marquant de cette analyse est la grande distance que l'on perçoit entre les discours qui relèvent de la présentation des programmes et des agendas médiatiques et ceux qui renvoient aux discours de dénigrement et aux *live-tweets*. On peut faire ici l'hypothèse que « l'auteur collectif » des tweets de la première thématique n'est pas le même que « l'auteur collectif » des tweets de la seconde. Nous avons déjà pu conforter cette hypothèse en ce qui concerne le Front National. Une analyse complémentaire sur cette communauté montre qu'effectivement les tweets sur le programme sont significativement plus le fait de politiciens professionnels ou d'élus alors que les discours plus familiers sont davantage le fait de supporters anonymes (Benbouzid *et al.*, 2017).

CONCLUSION

Cette étude se fixait comme objectif de décrire et d'étudier les échanges d'informations intervenant sur Twitter pendant une campagne présidentielle. Inscrite dans le cadre théorique des représentations sociales, elle a consisté à investiguer les discours des communautés de soutien des cinq principaux candidats à partir de méthodes relevant de l'analyse des réseaux et de l'analyse statistique des données textuelles. La méthodologie utilisée ici reprend en fait la démarche classiquement utilisée dans l'étude des représentations sociales, mais l'applique à des données massives. Elle profite de ce fait des opportunités offertes par les réseaux socio-numériques en termes de quantité de données et de spontanéité des discours. Elle n'a pas pour autant vocation à se substituer aux enquêtes traditionnelles mobilisées dans ce champ. Au-delà du manque de représentativité de l'audience étudiée, les données de Twitter souffrent également d'un manque de qualité quand il s'agit de s'intéresser aux aspects textuels. Une part entière des informations qui circulent sous forme d'images, de vidéos ou d'URL échappe complètement à ce type d'analyse. Même si des approches se mettent en place pour aider l'investigation des

images circulant sur Twitter (Bottini et Julliard, 2017), elles relèvent pour l'instant d'approches qualitatives qui paraissent difficilement applicables à des échantillons de plusieurs millions de tweets. Les 280 caractères récoltés par tweet ne peuvent pas remplacer la richesse des entretiens semi-directifs, des enquêtes par questionnaires ou des protocoles expérimentaux. Il nous semble plutôt que l'application des anciennes méthodes sur ce type de corpus, ou l'utilisation des méthodes émergentes, ne sont que de nouvelles cordes tendues sur l'arc des chercheurs en sciences humaines et sociales. La facilité technique grandissante de ce type de recueil et le temps que doit lui consacrer le chercheur positionnent cette démarche comme particulièrement pertinente dans le cadre de protocoles qui relèvent de la triangulation méthodologique, c'est-à-dire de l'utilisation de différentes approches pour aborder un même objet.

L'analyse que nous proposons ici permet dans un premier temps de prendre conscience de l'organisation des différentes communautés qui ont animé la campagne sur Twitter. La stratégie de repérage automatique de ces communautés, renforcée par la comparaison avec une indexation manuelle, met en évidence une nouvelle fois l'organisation axiologique qui émerge de l'étude du graphe des retweets. La structuration des relations entre les comptes étudiés montre effectivement que ceux-ci se regroupent autour des cinq principaux candidats. Elle souligne par exemple le faible nombre de soutiens du candidat Emmanuel Macron en dehors des comptes satellites de son mouvement politique et une grande proximité avec de nombreux médias *mainstream*. La particularité du lien entre cette communauté et les principaux comptes médiatiques devra faire l'objet d'une investigation particulière. Cette méthode nous permet surtout de produire des corpus des tweets originaux qui rendent possible la comparaison des discours entre ces communautés.

Les résultats de l'analyse de ces corpus permettent de montrer comment les discours portant sur les programmes et sur les agendas médiatiques des cinq principaux candidats à l'élection se démarquent très nettement, dans presque toutes les communautés politiques, des autres formes de discours. On peut voir dans cette opposition une concrétisation de la tension entre les attitudes paradigmatiques (« parler de quelque chose ») et syntagmatiques (« parler à quelqu'un ») (Marchand, 2010).

La comparaison des classes relevant des mêmes thématiques entre les différentes communautés de soutien rend saillantes les différences de perception

et de représentation des événements, comme le Penélope Gate, ou d'objets, comme la personnalité d'Emmanuel Macron. Ces différences sont interprétées ici comme la conséquence d'un positionnement idéologique qui régule les représentations sociales d'une pluralité d'objets. Cependant, ces différenciations se font dans un champ culturel restreint qui dicte des règles et des thématiques obligées. Bien que les discours sur les programmes ne soient pas majoritaires, toutes les communautés présentent une ou plusieurs classes sur les propositions du candidat. Toutes les communautés présentent également des classes de discours liées aux commentaires des sondages et des résultats ou des lives-tweet d'émissions télévisées.

L'étude de la chronologie de surreprésentation de l'ensemble des thématiques identifiées montre assez clairement que les événements médiatiques et les dates des élections sont les principaux régulateurs de l'expression de ces discours. Plus généralement, notre étude confirme le lien désormais avéré entre les discours développés dans les médias sociaux et les logiques politiques et médiatiques dans le cadre d'événements de nature institutionnelle comme les élections. Il s'agit d'une tendance mise en évidence dans des études empiriques en France comme aux États-Unis : bien que soit observée la « géométrie variable » d'une communication transversale à plusieurs étages, en ce qui concerne les sujets les plus partagés et commentés, l'agenda observable sur Twitter reflète en grande partie un agenda médiatique plus large (Rieder et Smyrniaios, 2012 ; Conway *et al.*, 2015). Plutôt que de constituer un espace public numérique autonome, les débats qui ont lieu au sein de Twitter sont largement surdéterminés, à la fois par les représentations que les groupes construisent dans de multiples espaces sociaux, mais aussi par l'agenda imposé par les institutions politiques et les médias. De ce point de vue, Twitter constitue une interface – déformante – où se reflètent des préoccupations, des opinions, mais aussi des hiérarchies construites ailleurs (Rieder, 2012).

 RÉFÉRENCES

APOSTOLIDIS T. (2003), « Représentations sociales et triangulation : Enjeux théorique-méthodologiques », in J.-C. Abric (dir.), *Méthodes d'étude des représentations sociales*, Ramonville-Saint-Agne, Erès, pp. 13-36.

BASTIAN M., HEYMAN S., JACOMY M. (2009), « Gephi: An Open Source Software for Exploring and Manipulating Networks », *International AAAI Conference on Web and Social Media*, Third International AAAI Conference on Weblogs and Social Media.

BENBOUZID B., SMYRNAIOS N., FIGEAC J., FRAISIER O., L'HÔTE A., LOVELUCK B., PERRIER A., RATINAUD P., SALORD T. (2017), *Twitter vs. Facebook : réseaux et discours de l'extrême droite*. Présentation orale. décembre 2017. Participation du projet Listic au Sprint Datapol organisé par Sciences Po.

BLONDEL V. D., GUILLAUME J.-L., LAMBIOTTE R., LEFEBVRE E. (2008), « Fast unfolding of communities in large networks », *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, n° 10.

BORRA E., RIEDER B. (2014), « Programmed method: developing a toolset for capturing and analyzing tweets », *Aslib Journal of Information Management*, vol. 66, n° 3, pp. 262-278.

BOTTINI T., JULLIARD V. (2017), « Entre informatique et sémiotique. Les conditions techno-méthodologiques d'une analyse de controverse sur Twitter », *Réseaux*, n° 204, pp. 35-69.

BOURDIEU P. (1977), « La production de la croyance : contribution à une économie des biens symboliques », *Actes de la recherche en sciences sociales*, n° 13, pp. 3-43.

BRUNS A., MOON B., MÜNCH F., SADKOWSKY T. (2017), « The Australian Twittersphere in 2016: Mapping the Follower/Followee Network », *Social Media + Society*, n° 3(4).

CAILLAUD S., FLICK U. (2016), « Triangulation méthodologique ou comment penser son plan de recherche », in G. LO MONACO, S. DELOUVÉ, P. RATEAU (dir.), *Les représentations sociales. Théories, méthodes et applications*, Louvain-la-Neuve, De Boeck Supérieur, pp. 135-148.

CHAVALARIAS D., GAUMONT N., PANABI M. (2019), « Hostilité et prosélytisme des communautés politiques en ligne : ce que nous enseignent les réseaux sociaux », *Réseaux*, n° 214-215, à compléter.

COINTET J., PARASIE S. (2018), « Ce que le big data fait à l'analyse sociologique des textes : un panorama critique des recherches contemporaines », *Revue française de sociologie*, vol. 59, n° 3, pp. 533-557.

CONOVER M. D., GONÇALVES B., RATKIEWICZ J., FLAMMINIA A., MENCZER F. (2011), « Predicting the political alignment of Twitter users », in *Proceedings-2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom*, pp. 192-199.

CONWAY B. A., KENSKI K., WANG D., (2015), « The Rise of Twitter in the Political Campaign: Searching for Intermedia Agenda-Setting Effects in the Presidential Primary », *Journal of Computer-Mediated Communication*, vol. 20, n° 4, pp. 363-380.

DOISE W. (1985), « Les représentations sociales. Définition d'un concept », *Connexions*, n° 45, pp. 243-253.

DOISE W. (1989), « Attitudes et représentations sociales », in D. JODELET (dir.), *Les représentations sociales*, Paris, PUF, pp. 220-238.

FLAMENT C., ROUQUETTE M.-L. (2003), *Anatomie des idées ordinaires, comment étudier les représentations sociales*, Paris, Armand Colin.

FLAXMAN S., GOEL S., RAO J. M. (2016), « Filter Bubbles, Echo Chambers, and Online News Consumption », *Public Opinion Quarterly*, vol. 80, n° S1, pp. 298-320.

FRAISIER, O., CABANAC G., PITARCH Y., BESANÇON R., BOUGHANEM M. (2018), « #Élysée2017fr: The 2017 French Presidential Campaign on Twitter », in *International Conference on Weblogs and Social Media (ICWSM 2018)*, Stanford, California, États-Unis, 25/06/2018-28/06/2018, AAAI Press.

GAUMONT N., PANAHI M., CHAVALARIAS D. (2018), « Reconstruction of the socio-semantic dynamics of political activist Twitter networks-Method and application to the 2017 French presidential election », *PLOS ONE*, n° 13(9).

GOLDBECK J., HANSEN D. (2014), « A method for computing political preference among Twitter followers », *Social Networks*, vol. 36, pp. 1-184.

GONZÁLEZ-BAILÓN S., BORGE-HOLTHOEFER J., WANG N., RIVERO A. MORENO Y. (2012), « Assessing the Bias in Communication Networks Sampled from Twitter », (No. arXiv: 1212.1684).

GRANJON F., DENOUEL J. (2010), « Exposition de soi et reconnaissance de singularités subjectives sur les sites de réseaux sociaux », *Sociologie*, vol. 1, pp. 25-43.

GUIMELLI C., DESCHAMPS J.-C. (2000), « Effet de contexte sur la production d'associations verbales. Le cas des représentations sociales des gitans », *Les Cahiers internationaux de psychologie sociale*, vol. 47-48, pp. 44-54.

JACOMY M., VENTURINI T., HEYMANN S., BASTIAN M. (2014), « ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software », *PLOS ONE*, 9(6).

JODELET D. (1989), « Représentations sociales : un domaine en expansion », in D. JODELET (dir.), *Les représentations sociales*, Paris, PUF, pp. 47-78.

KATZ E., LIEBES T. (2007), « 'No More Peace!': How Disaster, Terror and War Have Upstaged Media Events », *International Journal of Communication*, n° 1, pp. 157-166.

LABBÉ C., LABBÉ D. (2003), « La distance intertextuelle », *Corpus*, n° 2, pp. 95-117.

LABBÉ D., MONIÈRE D. (2000), « La connexion intertextuelle. Application au discours gouvernemental québécois », in M. RAJMAN, J.-C. CHAPPELIER (dir.), *Actes des 5^{es} Journées internationales d'analyse statistique des données textuelles*, Lausanne, EPLF, pp. 85-94.

MANGE J. (2016), *Amorçage de rapports : intergroupes et conséquences*, Document de synthèse pour l'obtention d'HDR, Université de Caen.

MARCHAND P. (2010), « Attitudes towards language: new perspectives », *International Review of Social Psychology*, vol. 23, n° 3-4, pp. 125-153.

MARIS., VOLPATOC., PAPASTAMOUS., CHRYSOCHOOU X., PRODROMITIS G., PAVLOPOULOS V. (2017), « How Political Orientation and Vulnerability Shape Representations of the Economic Crisis in Greece and Italy », *International Review of Social Psychology*, vol. 30, n° 1, pp. 52-67.

MATUSZAK C. (2007), « L'environnement discursif des forums politiques : le cas des forums d'organisations politiques marginales », *Études de communication*, vol. 30, n° 1, pp. 117-131.

MOLINER P. (1994), « Les deux dimensions des représentations sociales », *Revue internationale de psychologie sociale*, n° 2, pp. 73-86.

MOSCOVICI S. (1961), *La psychanalyse, son image et son public*, Paris, PUF.

MUNTZ D. C. (2002), « Cross-cutting social networks: Testing democratic theory in practice », *American Political Science Review*, vol. 96, n° 1, pp. 111-126.

PARISER E. (2011), *The Filter Bubble: What The Internet Is Hiding From You*, London, Penguin.

RATEAU P. (2000), « Idéologie, représentation sociale et attitude : étude expérimentale de leur hiérarchie », *Revue internationale de psychologie sociale*, vol. 13, n° 1, p. 2957.

RATINAUD P. (2003), *Les professeurs et Internet. Contribution à la modélisation des pensées sociale et professionnelle par l'étude de la représentation professionnelle d'internet d'enseignants du secondaire*, Thèse de doctorat en Sciences de l'Éducation, Université de Toulouse-Le Mirail.

RATINAUD P. (2014), IRaMuTeQ : Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires (Version 0.7 alpha 2) [Windows, GNU/Linux, Mac OS X], <http://www.iramuteq.org>.

- RATINAUD P. (2016), « Étude d'un large corpus textuel avec IRaMuTeQ: Twitter et le hashtag #mariagepourtous », in G. LO MONACO, S. DELOUVÉ, P. RATEAU (dir.), *Les représentations sociales. Théories, méthodes et applications*, Louvain-la-Neuve, De Boeck Supérieur, pp. 107-112.
- RATINAUD P., MARCHAND P. (2012), « Application de la méthode ALCESTE à de "gros" corpus et stabilité des "mondes lexicaux" : analyse du "CableGate" avec IRaMuTeQ », in *Actes des 11^{es} Journées internationales d'analyse statistique des données textuelles* (JADT 2012), Liège, Belgique, pp. 835-844.
- RATINAUD P. MARCHAND P. (2015), « Des mondes lexicaux aux représentations sociales. Une première approche des thématiques dans les débats à l'Assemblée nationale (1998-2014) », *Mots. Les Langages du politique*, n° 108, pp. 57-77.
- RATINAUD P., MARCHAND P. (2016), « Quelques méthodes pour l'étude des relations entre classifications lexicales de corpus hétérogènes : application aux débats à l'assemblée nationale et aux sites web de partis politiques », in D. MAYAFFRE, C. POUDAT, L. VANNI, V. MAGRI, P. FOLLETTE (dir.), *Statistical Analysis of Textual Data*, Nice, France. pp. 193-202.
- REINERT M. (1983), « Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte », *Les Cahiers de l'analyse des données*, vol. 8, n° 2, pp. 187-198.
- REINERT M. (1990), « ALCESTE. Une méthodologie d'analyse des données textuelles et une application : Aurélia de Gérard de Nerval », *Bulletin de méthodologie sociologique*, n° 26, pp. 24-54.
- RIEDER B. (2012), « The refraction chamber: Twitter as sphere and network », *First Monday*, vol. 17, n° 10.
- RIEDER B., SMYRNAIOS N. (2012), « Pluralisme et infomédiation sociale de l'actualité : le cas de Twitter », *Réseaux*, n° 175, pp. 105-139.
- ROUQUETTE M.-L. (1973), « La pensée sociale », in S. MOSCOVICI (dir.), *Introduction à la psychologie sociale*, Paris, Larousse, pp. 298-327.
- ROUQUETTE M.-L. (1996), « Représentations et idéologie », in J.-L. BEAUVOIS, J.-C. DESCHAMPS (dir.), *Des attitudes aux attributions, sur la construction de la réalité sociale*, Grenoble, Presses universitaires de Grenoble, pp. 163-173.
- ROUQUETTE M.-L. (1998), *La communication sociale*, Paris, Dunod.
- ROUQUETTE M.-L. (2009), *La pensée sociale. Perspectives fondamentales et recherches appliquées*, Toulouse, Erès.
- SALÈS-WUILLEMIN E. (2011), « Contrat de communication et représentations sociales », in P. CASTEL, E. SALÈS-WUILLEMIN, M.-F. LACASSAGNE (dir.), *Psychologie sociale, communication et langage*, Bruxelles, De Boeck, pp. 87-106.

SCHERER K. R., GILES H. (eds.) (1979), *Social markers in speech*. Cambridge, Cambridge University Press/Éditions de la Maison des Sciences de l'Homme.

SEVERO M., LAMARCHE-PERRIN R. (2018), « L'analyse des opinions politiques sur Twitter », *Revue française de sociologie*, vol. 59, n° 3, pp. 507-532.

SMYRNAIOS N., RATINAUD P. (2014), « Comment articuler analyse des réseaux et des discours sur Twitter », *Tic et société*, vol. 7, n° 2, pp. 120-147.

SMYRNAIOS N., RATINAUD P. (2017), « The Charlie Hebdo Attacks on Twitter: A Comparative Analysis of a Political Controversy in English and French », *Social Media + Society*, n° 3.

TAFANI E. (1997), *Attitude et approche structurale des représentations sociales : de l'ancrage psychologique à l'ancrage sociologique*, Thèse de doctorat en psychologie sociale, Université de Provence.

VAN DIJK T. A. (1998), *Ideology, a multidisciplinary approach*, London, Sage.

VENTURINI T., LATOUR B. (2010), « The Social Fabric: Digital Traces and Qualitative Methods », in *Proceedings of Futur en Seine 2009*, pp. 87-101.

WELLER K., BRUNS, A., BURGESS J., MAHRT M. PUSCHMANN C. (eds.) (2013), *Twitter and Society*, New York, Peter Lang.