La visibilité politique en ligne de la maire de Paris : Contribution à la mesure de l'écho Web-médiatique d'Anne Hidalgo

Guillaume Marrel, Vincent Labatut

▶ **To cite this version:**

Guillaume Marrel, Vincent Labatut. La visibilité politique en ligne de la maire de Paris : Contribution à la mesure de l'écho Web-médiatique d'Anne Hidalgo. Christophe Alcantara; Francine Charest; Serge Agostinelli. Big Data et visibilité en ligne : Un enjeu pluridisciplinaire de l'économie numérique, Presses des Mines, pp.271-286, 2018. hal-03173422

HAL Id: hal-03173422

https://univ-avignon.hal.science/hal-03173422

Submitted on 18 Mar 2021

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.

La visibilité politique en ligne de la maire de Paris

Contribution à la mesure de l'écho web-médiatique d'Anne Hidalgo

Guillaume Marrel et Vincent Labatut

Dépendants de l'opinion, tenus de se légitimer, les acteurs politiques n'échappent pas à la nécessité de produire, mettre en scène et contrôler leur image dans les médias de masse depuis la fin du XIXe siècle (Balandier, 2006), et aujourd'hui en ligne et sur les réseaux socio-numériques (Davis et al., 2017). Ils s'emparent des dispositifs numériques permettant d'entretenir leur légitimité à représenter et à gouverner et tentent de maîtriser cette image dans un nouvel espace public annoncé comme « ouvert, décentralisé et renouvelé » (Coleman and Blumler, 2009) de la « démocratie électronique » (Cardon, 2010). Avec l'essor des médias Web, l'activité quotidienne des élus et des gouvernants est de plus en plus reportée et commentée en ligne par leurs propres services de communication, les professionnels de la politique et de l'information qui les entourent, mais également, et de plus en plus par hypothèse, par des « citoyens ordinaires », sur les pages institutionnelles ou personnelles, les blogs, certains forums, les murs Facebook, Twitter et autres réseaux socio-numériques. Notre enquête vise à saisir l'ensemble des messages, discours et représentations qui circulent en ligne, concernant l'activité d'un·e élu·e sur une période donnée.

Cet objectif est partagé par les acteurs politiques eux-mêmes (Théviot, 2016) et les sociétés de services de veille et d'e-marketing auxquelles ils ont recours pour la gestion de leur e-réputation (Alcantara, 2015). Ceux-ci disposent d'outils et d'indicateurs plus ou moins élaborés, retraitant principalement les contenus des médias sociaux. Notre démarche est différente. Il s'agit pour nous de circonscrire la part de l'écho Web-médiatique d'une personnalité publique, que les outils de la recherche informatique, la structuration des médias, des supports, des messages, des formats et de l'accessibilité des données permettent aujourd'hui de reconstituer. Quels sont aujourd'hui les instruments dont nous disposons pour représenter ce qui est médiatisé en ligne, de l'action des élus de la démocratie représentative, dans l'espace public numérique ? Comment fixer de manière objective le reflet Web-médiatique par nature instable et incomparable – parce qu'individualisé – de leur travail politique ? Quels types d'événements transparaissent en ligne, avec quelle fréquence et selon quelles modalités ?

L'Internet véhicule aujourd'hui la quasi-totalité des messages relatifs à la vie politique nationale et territoriale, quels que soient leurs natures et formats originaux, écrits ou audio-visuels, confidentiels ou publics, contrôlés ou involontaires, unidirectionnels ou interactifs, informatifs ou polémiques, etc. Cet effet de « concentration numérique des contenus » et les progrès de la *fouille de données* laissent croire qu'il serait désormais possible de saisir, représenter et analyser l'ensemble de ces flux. C'est ce que notre enquête se propose de vérifier en croisant démarche numérique de fouille d'information et analyse manuelle qualitative. Nous proposons une approche empirique visant à tester la reconstruction de ce que les publics sont, d'une manière générale, susceptibles de recevoir comme information sur l'ensemble des supports aujourd'hui disponibles en ligne, concernant une période d'activité d'une personnalité politique importante.

L'enquête porte sur une semaine de l'activité de la maire de Paris, du 6 au 12 mars 2017. La période renvoie à une « phase d'activité politique normale » (Godmer and Marrel, 2015). Le choix d'Anne Hidalgo s'explique par sa forte visibilité nationale, son activité de publication en ligne, et le suivi étroit dont font l'objet les messages, le comportement et les décisions de cette dirigeante de l'exécutif de la plus grande collectivité urbaine de France. La démarche consiste à utiliser de façon programmatique les principaux moteurs de recherche du Web et des réseaux socio-numériques pour traiter des requêtes simples associant le patronyme de l'élue et la période. Les résultats filtrés permettent d'extraire des « événements », qui sont alors comparés au contenu de l'agenda

électronique professionnel de l'édile, récupéré avec son consentement et planifiant son emploi du temps effectif.

1. À quoi ressemble l'écho Web-médiatique de l'activité d'un·e élu·e ?

Le Web peut être saisi comme un grand miroir pouvant refléter l'activité politique, mais aussi un prisme déformant du fait de certaines réfractions. La maîtrise de cette déformation médiatique devient un enjeu stratégique de premier ordre avec l'Internet dont la dimension conversationnelle change le rapport des leaders aux instruments de contrôle de leur image (Roginsky and Perrier, 2014), à tel point que l'e-reputation politique fait aujourd'hui l'objet de conseils spécialisés. Les entrepreneurs politiques et leurs équipes sont désormais très sensibles à la nécessité de connaître et maîtriser leur identité et leur image numérique, en cours de mandat et plus encore en période de campagne (Giasson et al., 2017).

L'écho Web-médiatique de l'activité politique peut être défini comme l'ensemble des messages véhiculés par l'Internet à propos de l'action d'une personnalité politique sur une période donnée. Il est une déformation partielle de la réalité du travail effectué, qui peut être, par hypothèse, schématisé ainsi :

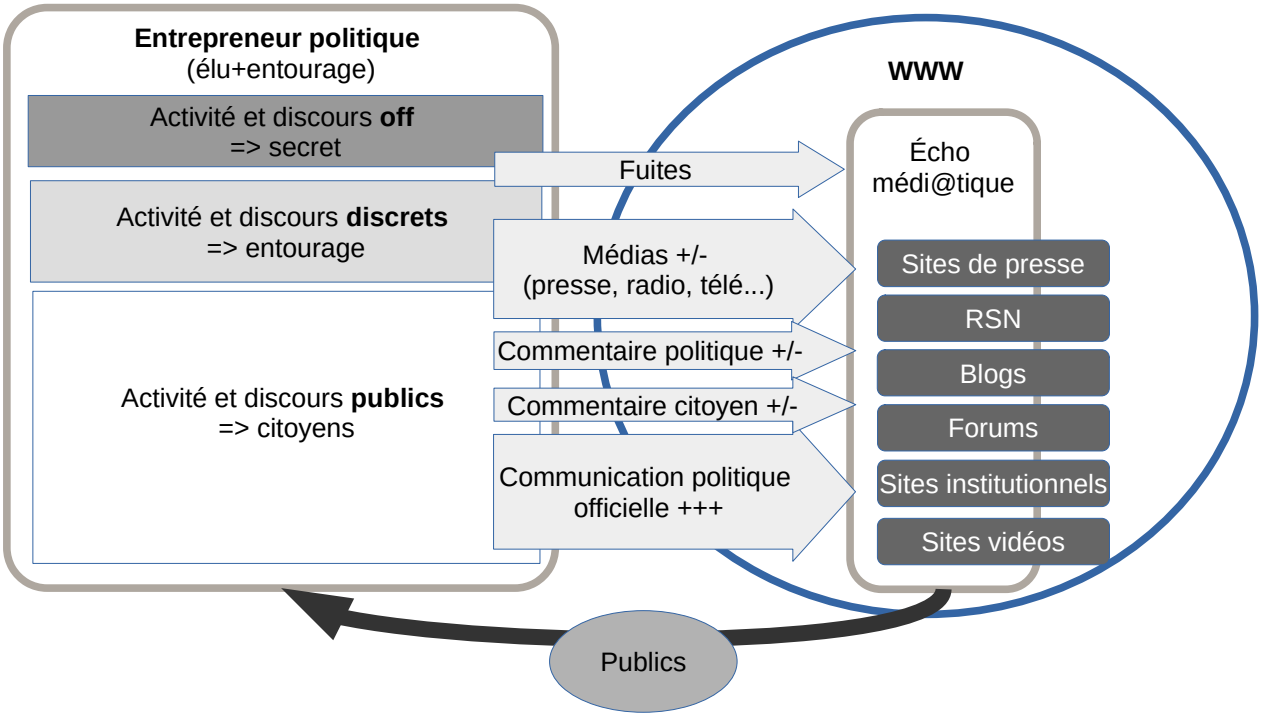

Figure 1. Représentation de l'écho Web-médiatique de l'activité d'un entrepreneur politique

L'action politique est, *a priori*, décomposée en trois catégories renvoyant à des niveaux de publicité différents. Une grande partie de ces activités et discours sont susceptibles d'être rendus publics par les services de communication de l'élu, les médias institutionnels, divers acteurs experts ou partisans, voire des citoyens ordinaires, militants ou leaders d'opinion. L'ensemble de ces actions de médiatisation de l'activité et des paroles de la maire nourrissent par hypothèse six catégories de médias et de format sur le Web : les sites et blogs officiels des institutions politiques et des collectivités publiques où circule la communication institutionnelle ; les sites et les pages Web des médias de presse écrite, audio-visuelle ou Web ; les blogs personnels de journalistes, d'experts, de personnalités politiques ou de citoyens plus ou moins ordinaires ; les médias sociaux et en particulier *Facebook* et *Twitter* ; des forums plus ou moins spécialisés ; les sites d'hébergement de

vidéos, comme *YouTube*. L'ensemble des textes et des images publiés sur ces supports par tous les acteurs pré-cités, concernant l'action de la maire et quels qu'en soient la véracité ou la polarité, constituent ce qu'il convient de nommer l'*écho Web-médiatique de l'activité politique*. La prise en compte des effets possibles des messages et de leur distorsion médiatique, sur les publics administrés et électeurs, peut être représentée par une boucle de rétroaction reposant sur l'hypothèse d'un ajustement des comportements politiques et des stratégies de communication de l'équipe au pouvoir. C'est ce corpus théorique qu'il s'agit ici d'identifier, d'extraire et de compiler, à l'aide d'un dispositif de fouille et d'extraction automatique afin de saisir les logiques de sa production et les possibilités de son éventuel contrôle.

2. Dispositifs de fouille : l'inaccessibilité des médias sociaux

Notre dispositif de fouille et d'extraction automatique est conçu pour traiter une grande partie de l'information circulant sur Internet, grâce aux moteurs de recherche indexant son contenu, et sur les médias sociaux, via l'accès direct qu'ils proposent. Il se décompose en 9 étapes de traitement décrites ci-dessous.

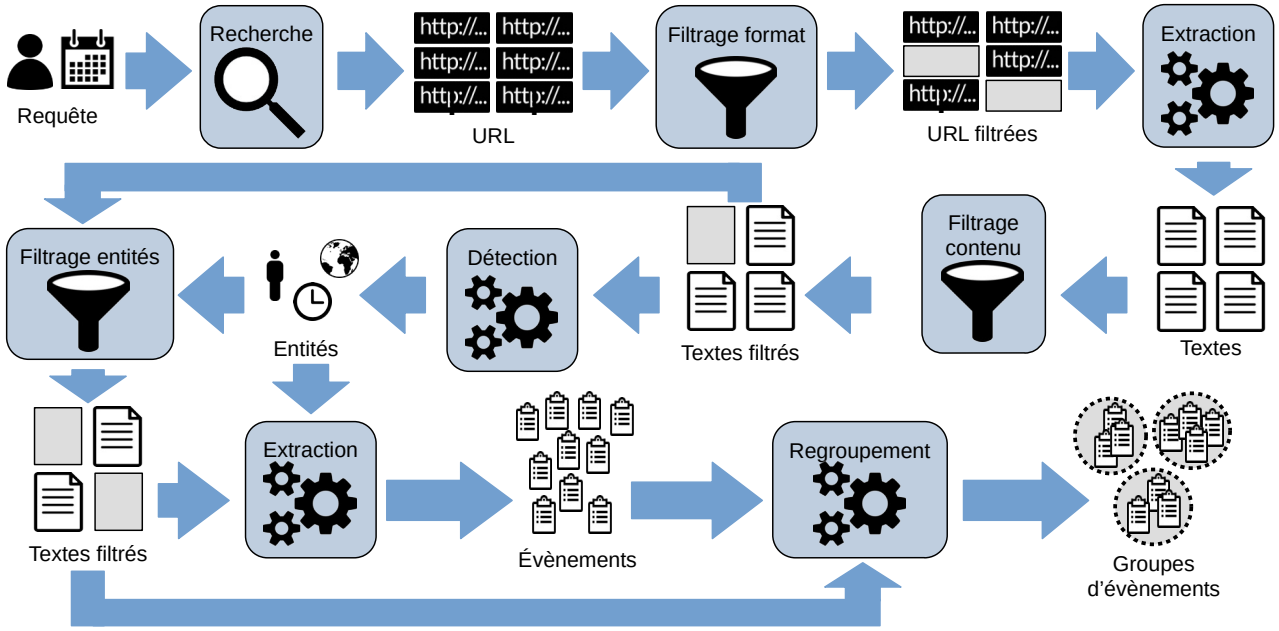

Figure 2. Représentation du traitement effectué sur les pages Web et les médias sociaux

Les 9 étapes de fouille et d'extraction automatique

Fouille du Web. L'utilisateur spécifie le nom de la personne ciblée et la période temporelle visée. Une recherche Web est effectuée afin d'obtenir les URL des documents supposés pertinents. Nous sollicitons les 4 principaux moteurs proposant un accès programmatique : Bing, Google, Qwant et Yandex. Ils proposent des fonctionnalités différentes, notamment sur le ciblage de la recherche, et certains imposent des limitations d'usage. L'utilisation de plusieurs moteurs nous permet de comparer leurs résultats et de bénéficier de leur potentielle complémentarité, mais nécessite de réaliser un certain nombre d'aménagements pour uniformiser leurs résultats.

Filtre par format. Les URL renvoyées par les moteurs sont filtrées de manière à évacuer celles qui ne sont pas exploitables par notre outil dans sa version actuelle. On ne garde que les documents textuels, laissant de côté images, sons et vidéos. Un agent humain serait capable d'exploiter ces données, et il y a donc une perte d'information potentielle. Cependant, la rareté des documents multimédia renvoyés ne justifie pas la mise en place des lourdes méthodes d'analyse nécessaires à leur exploitation.

Récupération des documents. Les pages Web correspondant aux URL restantes sont téléchargées. Puis, un extracteur développé par nos soins est chargé d'identifier et exporter uniquement leur texte pertinent, en ignorant les menus, mentions légales, etc. Nous utilisons des extracteurs spécifiques pour les sites de presse les plus courants,

tandis qu'un extracteur générique se charge de traiter le reste des pages. Nous nettoyons ensuite le texte récupéré, afin d'en supprimer certains caractères spéciaux problématiques.

Fouille des médias sociaux. Les médias sociaux ne sont pas indexés par les moteurs de recherche classiques. Il faut effectuer des recherches spécifiques sur ceux-ci, en se soumettant à leurs propres contraintes techniques. Cependant, l'information fournie est directement exploitable, ne nécessite pas d'effectuer les étapes de filtrage et d'extraction textuelle décrites précédemment. Faute de pouvoir librement accéder à Twitter, nous nous concentrons sur Facebook, qui impose de cibler des pages spécifiques. Nous traitons celle de la personne visée par l'utilisateur, et récupérons ses posts pour la période traitée, ainsi que les commentaires afférents et les posts publiés par leurs propres auteurs sur la même période. Nous appliquons le même traitement aux principaux opposants et soutiens de la personne ciblée.

Filtre par contenu. Nous réalisons un filtrage sur les contenus (Web et médias sociaux) dans le but d'écarter les textes non-pertinents. Notre outil identifie automatiquement la langue des textes traités, et ne garde que ceux rédigés en français. De plus, nous ne conservons que les textes contenant explicitement le patronyme de la personne ciblée (outre ceux dont elle est l'auteur).

Détection des entités nommées. Nous détectons dans chaque article les mentions d'*entités nommées* : noms de personnes, lieux, organisations, fonctions, productions, réunions, et dates. Notre outil repose sur une approche à base d'ensembles (Atdağ and Labatut, 2013).

Filtre par entité. Nous conservons seulement les documents contenant au moins une date comprise dans l'intervalle spécifié pour la recherche. On peut aussi les filtrer en exprimant des contraintes relatives aux autres types d'entités nommées détectées.

Extraction des évènements. Nous faisons l'hypothèse simplificatrice qu'un article décrit un seul événement, ce qui se vérifie le plus souvent sur ce corpus. La sémantique associée à cet événement est décrite par l'ensemble des entités détectées dans le texte : *où* (lieux où l'événement s'est déroulé), *quand* (ses dates) et *qui* (acteurs y ayant participé). Les autres types d'entités nommées sont optionnels et servent seulement à préciser l'événement.

Regroupement des documents. Un même événement est fréquemment décrit par plusieurs documents. Pour identifier ces cas de figure, nous regroupons les documents par similarité. Nous utilisons l'approche de recherche d'information classique : score *tf-idf* et distance cosinus (Manning et al., 2008), associés à une classification standard. Les documents formant un même groupe sont supposés décrire le même événement.

La mise en œuvre met en évidence une première dimension de l'écho web-médiatique : l'accessibilité très variable entre les pages Web et les réseaux socio-numériques, limitant immédiatement la recherche exploratoire. S'il est possible de recenser librement la plus grande part des publications institutionnelles, de la presse et des médias audio-visuels, ou encore des blogs et des forums, la fouille thématique et chronologique des médias sociaux n'est plus possible, au moment de l'enquête. La commercialisation du gigantesque patrimoine d'informations dont Twitter et Facebook sont désormais propriétaires interdit l'exploration globale du corpus des publications et commentaires hébergés relatives au mot-clef « Anne Hidalgo ». La reconstitution de l'écho médias-sociaux est pour ainsi dire impossible. L'accès programmatique est beaucoup moins stable que pour les moteurs Web. Il est en outre très limité car prévu pour le développement d'applications pour smartphones. Sur Facebook, la recherche est possible seulement parmi les posts publiés par une personne donnée sur une période donnée. Nous avons tenté de contourner ce biais en identifiant la page la maire, ainsi que celles de six de ses principaux soutiens et opposants, membres du conseil municipal, répartis sur l'échiquier politique, disposant d'un minimum de notoriété, et utilisant ce média[1]. Le recensement est cependant décevant. Sur la période, seules trois des sept personnes testées postent, et elles postent peu : la maire cinq fois, Goasguen cinq fois et Nadjovski une seule fois. Les 11 documents compilés ne donnent pas une vision représentative de l'écho de l'activité politique de la maire sur les médias sociaux et cela nous conduit à renoncer en l'état à leur exploitation. L'exploitation de Facebook pose un autre souci technique relatif à la distinction entre *compte utilisateur* (relié à une personne physique) et *page Facebook* (institution, personne publique...). Un compte utilisateur donné peut contrôler plusieurs pages. L'accès programmatique n'autorise à récupérer que les posts publics publiés sur une page, pas ceux d'un compte utilisateur.

1 J-L. Missika (PS), I. Brossat (PCF), C. Najdovski (EELV), N. Kosciusko-Morizet (Prog), C. Goasguen (LR), B. Kuster (LR).

Or, certaines personnalités politiques utilisent leur compte personnel, et non pas une page : un outil automatique ne peut donc accéder à leurs posts. Cette limitation empêche aussi d'explorer profondément les réactions du public aux posts publiés par les personnes étudiées.

Notre dispositif est davantage opérationnel sur le reste du Web. Une première exploration a été conduite le 20 juillet 2017 sur la période du 6 au 12 mars. La recherche a été réitérée le 12 septembre afin d'évaluer la stabilité des résultats. Elle montre d'abord une certaine instabilité du comportement des moteurs dans le temps, puisque 27 % des pages pertinentes identifiées au premier passage disparaissent des résultats au second. Elle souligne ensuite que les moteurs de recherche utilisés ne proposent qu'une infime partie des publications de la presse locale, pourtant indexée. Afin de corriger cette cécité, nous avons donc effectué une autre série de recherches, en ciblant uniquement le site Web du principal journal de la PQR : *Le Parisien*. Nous avons obtenu par ce biais 47 nouvelles références. Nous considérons donc d'abord les résultats obtenus sur la semaine du lundi 6 au dimanche 12 mars, combinant la recherche neutre (i.e. sur le Web entier) et la recherche ciblant spécifiquement *Le Parisien*. Finalement, nous récupérons 817 URL, dont 284 se révèlent être, après filtrage, des pages exploitables.

	Web	leparisien.fr	Total
Page indisponible	63	45	108
Langue étrangère	117		117
Mot-clef manquant	256	45	301
Format inexploitable	7		7
URL exploitables	**237**	**47**	**284**
Total	680	137	817

Table 1. Distribution des URL filtrées

3. Fragilité et pertinence des résultats

L'observation des résultats du filtrage amène une première remarque concernant la grande hétérogénéité des résultats des 4 moteurs utilisés. Yandex et Google renvoient respectivement moins de 9% et 20% des pages subsistant après le filtrage, alors que Bing et Qwant atteignent respectivement plus de 96% et 66%. Il est à noter que, de façon générale, ces deux derniers renvoient bien plus de résultats, car leur interface programmatique ne permet pas de spécifier explicitement la période visée par la recherche. Pour contourner cette limitation, nous incluons explicitement une date dans les requêtes qui leur sont soumises, et traitons donc autant de requêtes qu'il y a de jours dans la période ciblée.

Cette hétérogénéité, nous a poussé à réaliser une évaluation manuelle du filtrage de l'outil. Pour cela, chaque page subsistant à l'issue du filtrage a été ouverte, examinée visuellement et catégorisée en quatre étapes : 1) élimination d'une minorité d'URL indisponibles ou dont l'accès est payant (1,8%) ; 2) de celles renvoyant à des pages qui traitent d'homonymes ou simplement hors-sujet (4,6%) ; 3) des quelques doublons (3,2%), i.e. les pages associées à des URL différentes mais ayant le même contenu textuel, provenant essentiellement de la recherche ciblant *Le Parisien* ; et 4) un dernier filtrage plus important sur les dates de publication et la période du déroulement de l'événement dont il est question dans le texte ouvert par l'URL. L'analyse visuelle élimine ici une part importante des pages pourtant sélectionnées par notre outil, puisque 85 liens (soit 30 %) concernant bien l'activité d'Anne Hidalgo sont néanmoins considérés comme « hors-période » et donc inexploitables. Le corpus final pertinent se réduit finalement à 172 URL renvoyant à des pages Web évoquant bel et bien l'activité de l'élue entre le 6 et le 12 mars 2017.

	Filtres manuels	Corpus filtré = 284
Indisponible	5	112
Homonyme / erreur	13	
Doublons	9	
Hors périodes	85	
Corpus final		**172**

Table 2. Résultat du traitement manuel des pages issues du filtrage automatique

Les moteurs de recherche qui constituent l'étape initiale de notre processus renvoient donc de nombreux résultats non-pertinents, commettant trois types d'erreurs. La première est le *hors-sujet* : l'article ne traite ni de la personne ciblée, ni de thèmes connexes. La plupart de ces articles ne contiennent pas explicitement le patronyme recherché, et sont donc écartés lors du filtrage par contenu. Les homonymes qui parviennent à passer ce filtre sont très rares. Un outil de désambiguïsation d'entités nommées permettrait de résoudre ce problème. Tous les moteurs étant configurés pour ne renvoyer que des documents écrits en français et publiés sur des sites français, nous considérons également comme hors-sujet les pages rédigées dans une autre langue. Notre étape de filtrage à base de contenu est capable de les détecter efficacement.

Le deuxième type d'erreur est le *hors-sujet relatif* : ces pages traitent d'un sujet proche de la personne visée, mais sans jamais la mentionner directement. Il s'agit généralement de pages contenant des liens vers des documents pertinents, mais dont le contenu pertinent a été mal délimité par notre extracteur.

Enfin, les pages *hors-période* traitent bien de la personne visée par la recherche, mais pas pour la période visée. Le problème peut s'expliquer en partie par le fait que certains moteurs de recherche ne permettent pas de spécifier formellement la période visée. Il est difficile de filtrer automatiquement les articles selon un critère temporel, car ils ne contiennent souvent pas explicitement de date. L'information la plus intéressante semble être la date de publication, qui fait partie des méta-données souvent associées à l'article. Cependant, cette date n'est pas toujours disponible, et son extraction requiert la plupart du temps un traitement spécifique au site concerné.

Un second problème majeur de la recherche Web est l'extraction du contenu textuel pertinent à partir du code HTML de la page. Cette tâche, qui paraît triviale à un être humain, est difficile à automatiser en raison de la complexité et de la diversité que permet le code HTML. Deux types d'erreurs sont possibles, qui ne sont pas mutuellement exclusives. La première est de ne récupérer qu'une partie du contenu textuel pertinent, ce qui dégrade la suite du traitement. Cette situation survient notamment quand l'accès à l'article est payant. Le problème ne se pose pas pour les extracteurs « taillés sur mesure » pour des sites spécifiques, mais seulement pour l'extracteur générique. En effet, celui-ci utilise un ensemble de règles prédéfinies pour identifier le texte pertinent, mais certaines pages ne s'y conforment pas. La seconde erreur est de récupérer du contenu non-pertinent. Le problème est plus sérieux qu'une récupération incomplète, car le texte non-pertinent peut rendre complètement inexploitable une page initialement pertinente. Or, la plupart des pages contiennent ce type de bruit (menus, publicités, liens externes...) mentionnant souvent des entités nommées. Garder ce texte lors de l'extraction entraîne plus tard la détection d'un grand nombre d'entités qui ne sont généralement pas mentionnées dans l'article proprement dit, brouillant l'identification des événements décrits.

En définitive, l'outil se révèle pour l'heure incapable de traiter les formats non-textuels du Web, les contenus des médias sociaux et il rencontre certaines difficultés dans le filtrage et l'extraction des données pertinentes. Ces limites laissent donc un doute sur sa capacité à recenser

l'*intégralité* des messages qui circulent en ligne sur le travail politique de l'élue. Cependant, comme nous allons le voir, les résultats obtenus permettent néanmoins de décrire la manière dont l'activité édilitaire est publicisée.

4. Une semaine de travail de la maire de Paris sur le Web

L'analyse manuelle des 172 documents filtrés permet d'identifier cinq types de supports et quatre types d'auteurs différents. Les sites de presse écrite représentent plus de 71% des résultats valides. Avec les sites de télévisions et de radios, les supports Web des médias institutionnels représentent plus de 80% des sources recensées par les moteurs de recherche. Les blogs ne comptent que pour 18% et les forums pour moins d'1%. Aucun événement ne remonte via le site institutionnel de la ville ou du parti.

Supports (5)	**%**
Site de presse	71,5%
Blogs	18%
Site de TV	6,4%
Site de radio	3,5%
Forum	0,6%

Auteurs (4)	**%**
Journaliste	95,4%
Militant	2,3%
Entreprise	1,2%
Individu	1,2%

Table 3. Types de supports et d'auteurs identifiés parmi les 172 filtrées

Les auteurs des messages compilés dans l'écho Web-médiatique reconstitué sont donc principalement des journalistes (95%). Les médias sociaux n'ayant pu être complètement traités, les militants, les acteurs économiques et les individus isolés restent très marginaux dans la diffusion des messages recensés concernant la maire de Paris.

En termes de contenu, le regroupement automatique (clustering) effectué sur les 172 pages du corpus final nous permet d'identifier 55 thèmes. L'analyse manuelle permet quant à elle d'identifier 39 objets distincts, renvoyant à des événements, des secteurs d'activités et des types d'interaction plus ou moins différents. Le recouvrement entre les deux démarches est de 0,87[2], ce qui dénote une très grande proximité entre les deux partitions.

Les 39 objets peuvent être répartis en 18 « secteurs d'action publique ou politique ». Leur examen montre un écho web-médiatique principalement centré sur l'actualité relative d'abord au sexisme (7 événements), à l'occasion de la journée de la femme du 8 mars, ensuite à la compétition électorale (5 événements), en cette période d'ajustement des alliances politiques autour des candidats à l'élection présidentielle de mai 2017, et enfin à l'urbanisme, au transport, au sport et au social (3 événements chacun), dans le cadre notamment des compétences d'action publique de la maire. Cinq autres objets renvoient à deux secteurs singuliers classés à part : « la corruption » et « l'injure », lorsque la lecture des textes dévoile des mises en causes nominales de l'élue pour des malversations supposées, voir des injures à caractère sexuel, sur des forums notamment.

Secteurs (18)	***n***
Sexisme	7
Compétition électorale	5
Social	3

Interactions (12)	***n***
Cérémonie	9
Annonce de politique publique	7
Déclaration partisane	6

2 La similarité entre les deux partitions est mesurée par l'Information Mutuelle Normalisée (NMI) : 0 signifie qu'elles sont indépendantes, et 1 qu'elles sont exactement identiques.

Sport	3
Transports	3
Urbanisme	3
Climat	2
Culture	1
Économie	1
Fiscalité	1
International	1
Justice	1
Mémoire	1
Santé	1
Sécurité	1
Université	1
Corruption	3
Injure	2
Total	**39**

Engagement partisan	3
Communiqué	2
Discours de politique publique	2
Condamnation	1
Interview	1
Publication	1
Objet mise en cause	4
Objet injure	2
Objet trajectoire	1
Total	**39**

Table 3. Secteurs d'action publique ou politique et du type d'interaction caractérisant les 39 objets identifiés parmi les pages filtrées manuellement

Une seconde classification des 39 objets peut être proposée en fonction du type d'interaction mobilisant l'élue face à ses interlocuteurs. De manière classique, les médias semblent principalement publiciser sur cette période les cérémonies, les annonces de politiques publiques et les déclarations partisanes. Pour 7 cas, l'élue est également l'objet d'une mise en cause directe, d'une injure ou d'un rappel de sa trajectoire biographique.

L'étude de la fréquence des publications relatives à un même objet, permet de préciser la nature de l'écho web-médiatique du travail de l'élue sur la période. Les objets donnent d'abord lieu à une couverture médiatique théorique de 4,3 publications par événement. Celle-ci n'est cependant pas régulièrement répartie et d'importantes distorsions apparaissent entre les 5 objets donnant lieu à au moins dix publications distinctes (et jusqu'à 33), 6 autres publiés entre trois et huit fois, 6 encore publiés deux fois et les 22 derniers qui ne produisent d'une seule publication. En définitive, les 5 événements les plus diffusés, commentés et relayés concernent :

1. ***La vie politique nationale*** : la prise de position partisane de la maire dans la campagne présidentielle pour le candidat socialiste B. Hamon, qui donne lieu à 33 publications différentes, du fait notamment des relais médiatiques locaux et nationaux de cet engagement à caractère national dans la séquence politique disposant de la plus grande couverture médiatique ;
2. ***L'urbanisme parisien*** : l'annonce d'une série de mesures de propreté dans le cadre des politiques municipales, qui occasionne 27 publications, pour un enjeu très localisé, mais populaire ;
3. ***La culture nationale*** : le soutien de la maire au vaste projet d'investissement de Bernard Arnaud et LVMH autour du Musée des arts et des traditions populaires, qui produit 20 publications en ligne, pour un événement de niveau national très relayé dans le monde de la culture ;

4. ***La justice*** : l'annonce de la condamnation en appelle d'Anne Hidalgo pour diffamation à l'encontre de l'architecte J-F. Cabestan, qui donne lieu à 17 publications, reflètant l'importance médiatique de la personne de la maire et des poursuites judiciaires qui pourraient affecter son image ;
5. ***Les transports et la sécurité routière*** : l'appui de la maire à une « campagne mondiale » de sensibilisation associant de nombreuses personnalités médiatiques, largement diffusée et donnant lieu ici à 10 publications.

Viennent ensuite quatre événements avec au moins cinq médiatisations : le déplacement de la maire avec le Premier ministre M. Valls à Barcelone pour un match Barça-PSG, deux déclarations critiques à l'égard du programme du candidat E. Macron, et une petite polémique sur la présence de N. Sarkozy à Barcelone. Si l'on rassemble l'ensemble des publications relatives au ralliement d'A. Hidalgo à la candidature de B. Hamon et ses suites immédiates en termes de déclarations dans le courant de cette semaine, c'est clairement l'action partisane et électorale nationale qui domine l'écho Web-médiatique et l'image de l'élue parisienne entre le 6 et le 12 mars dernier.

L'examen des cinq premiers faits du classement des quatre moteurs de recherche Web montre que les événements qui font le « buzz » remontent partout en première ou deuxième position sur la semaine, parfois plusieurs fois dans le Top-5. Seule l'inauguration du projet culturel LVMH n'y figure pas. Plus étonnant : deux événements n'ayant donné lieu qu'à une seule publication, donc peu relayées figurent en 3ème et 4ème rangs dans nos moteurs : l'un sulfureux, puisqu'il s'agit de l'évocation d'une injure prononcée par un maire de l'agglomération parisienne, l'autre engagé et fortement lié à la journée mondiale des femmes du 8 mars, concernant la publication d'une tribune féministe.

<table>
<tr><th>Top5</th><th>Google</th><th>Bing</th><th>Qwant</th><th>Yandex</th></tr>
<tr><td>1</td><td colspan="4">10 mesures de propreté à Paris</td></tr>
<tr><td>2</td><td colspan="3">Déclaration de soutien Hamon</td><td>Condamnée en appel</td></tr>
<tr><td>3</td><td>10 mesures de propreté à Paris</td><td colspan="2">Injuriée par un maire</td><td>Tribune féministe</td></tr>
<tr><td>4</td><td>Tribune féministe</td><td colspan="2">Match Barça-PSG</td><td rowspan="2">10 mesures de propreté à Paris</td></tr>
<tr><td>5</td><td colspan="3">Condamnée en appel</td></tr>
</table>

Table 4. Les 5 premiers événements recensés par chacun des 4 moteurs de recherche Web

5. Agenda et écho Web-médiatique : une analyse comparative des contenus

L'agenda électronique de la semaine de travail politique observé a été transmis par le chef de cabinet de la maire, dans un format texte, après élimination des éléments confidentiels. Nous avons recomposé 30 événements planifiés dans l'emploi du temps de la maire à comparer avec les 39 objets de l'écho Web-médiatique. On observe immédiatement que seules un tiers (11) des actions programmées par le maire figurent sous une forme médiatisée sur le Web. Il s'agit des événements délibérément construits sur la période par l'équipe politique et les journalistes et observateurs, comme des rendez-vous médiatiques à publiciser : déclaration lors de la matinale d'*Europe1*, interview, conférence de presse, cérémonies liées à la culture, à la Journée de la femme, aux Jeux Olympiques ou aux Accords de Paris sur le climat. Le déplacement pour un match du PSG à Barcelone est d'autant plus commenté qu'Anne Hidalgo accompagne le Premier ministre et que l'événement donne lieu à un incident diplomatique avec l'ancien Président Sarkozy. En revanche,

19 engagements de l'élue ne laissent apparemment aucune trace en ligne. Il s'agit la plupart du temps de rendez-vous en tête à tête, de conseils restreints ou de réunions d'équipe par nature discrets, ou de thématiques peu ou pas médiatisés et dont il n'est sans doute pas étonnant de ne retrouver aucun écho sur le Web. Plus étonnant est le silence de l'outil sur la visite officielle du maire d'Erevan le mardi 7 mars et celle du président Arménien le lendemain, qui occupe le maire, son équipe et une partie du conseil de 11h à 13h30. Il s'agit en partie d'un défaut de compilation lors de la dernière extraction, lié à l'instabilité des retours des moteurs dans le temps. Un rapide passage en revue des 43 URL proposées par les moteurs le 20 juillet, mais disparus dans les résultats du 12 septembre, montre en effet que deux publications évoquent bien cette visite officielle, l'une sur le site de l'Ambassade d'Arménie en France, l'autre sur un blog francophone de la diaspora arménienne. Reste que cet événement ne trouve que peu d'écho dans les médias.

L'enquête montre par ailleurs que l'écho Web-médiatique dépasse assez largement l'action réellement engagée sur la période. Entre le 6 et le 12 mars, la maire de Paris est citée ou interpellée sur pas moins de 13 objets dont l'origine est antérieure au travail politique de la semaine en cours. Il s'agit soit de « vieilles histoires » d'engagements ou relatives à des affaires datant de plusieurs mois comme les injures du maire du Plessi-Robinson, la cérémonie à Colombey ou l'ouverture d'un centre de réfugié, soit encore d'annonces ou de prises de positions remontant à quelques jours comme la déclaration de la maire sur la diminution des ondes électromagnétiques du 3 mars ou encore le soutien qu'elle apporte vers le 5 mars au candidat B. Hamon face à B. Delanoë engagé auprès d'E. Macron. Ces éléments invitent à objectiver la temporalité des événements politiques que matérialise leur durée de vie médiatique.

Dans l'écho Web-médiatique, le plus intéressant concerne enfin les 15 publications concernant vraisemblablement la période observée mais ne correspondant à aucune base dans l'agenda professionnel communiqué par l'équipe politique de la maire : quatre le mardi 7 mars (sur la critique du programme de Macron, la publication d'une tribune féministe ou encore la réaction de la maire à l'annonce de l'incarcération du maire de Dakar, par ailleurs secrétaire général de l'AIMF, qu'elle préside) ; cinq le mercredi 8 mars (sur deux messages féministes, son soutien à Hamon encore une fois et une critique du candidat Fillon) ; deux le jeudi 9 mars (dont sans surprise sa condamnation en appel pour diffamation qui fait le buzz avec 27 publications le jour même) ; trois le vendredi 10 mars (sur son « divorce avec Delanoë » soutenant Macron) ; et enfin un le dimanche 12 mars (sur l'annonce très relayée des dix mesures pour la propreté de Paris). Sur cette dernière le silence de l'agenda est suspect. De fait, l'annonce est faite dans les colonnes du JDD et donc sans doute quelques jours avant pour l'agenda, peut-être à l'occasion d'une l'interview téléphonique programmée le 9 mars et qui ne correspondait encore à rien de précis dans l'écho-médiatique.

Enfin, trois publications ne renvoient à rien dans l'agenda de la semaine puisqu'elles font références à des annonces passées pour des projets futurs : une nouvelle hausse des amendes de stationnement et de nouveaux aménagements urbains pour le rond-point des Champs-Élysées pour 2018 et un projet de voiture volante (*Sea Bubble*) pour 2020.

Ces distorsions observées dans l'analyse souligne certaines limites de l'ouillage disponible en l'état pour compiler toute l'information circulant en ligne sur l'activité d'une personnalité publique. Elles révèlent surtout certaines dynamiques propres au temps médiatique, ses retards, son élasticité. Seuls deux événements du top5 des faits les plus médiatisés sont dans l'agenda : la conférence de presse au Musée des Arts populaires et celle sur la campagne de sécurité routière. Les trois autres objets les plus diffusés relèvent moins d'interaction politique programmables dans l'emploi du temps que de déclarations, discours ou prises de position : 1) une position partisane adoptée quelques jours avant l'observation et ses conséquences évoquées et commentées tout au long de la semaine ; 2) une condamnation en appel et 3) une annonce de dix mesures de propreté.

**

Nonobstant l'incertitude relative des résultats renvoyés par les moteurs de recherche, le filtrage et l'absence des éléments diffusés sur les médias sociaux, l'enquête invite à ajuster certaines hypothèses sur la structure de la médiatisation politique en ligne. En particulier, la diversité des publications attendues, de leurs sources et de leurs supports, qui n'est pas réellement présente dans les résultats recensés et appelle de nouvelles recherches. Si les correctifs sont envisageables pour l'exploration du Web (fouille de l'image et du son, spécification des extracteurs ou recours à l'apprentissage artificiel, filtrage pondéré…), la fouille libre des réseaux socio-numériques s'avère nettement plus compromise, en dehors d'hypothétiques accords de recherche avec les sociétés commerciales qui les exploitent. L'amélioration de l'analyse du contenu textuel passera quant à elle par l'intégration d'outils linguistiques plus poussés, pour la détection et la résolution de coréférences, mais aussi l'analyse de sentiment, destinés à l'identification de la polarité des textes.

Références

Alcantara, C. (Ed.) (2015) *E-réputation: regards croisés sur une notion émergente*. Issy-les-Moulineaux, France: Gualino-Lextenso éd., impr. 2015.

Atdağ, S., and Labatut, V. (2013) A comparison of named entity recognition tools applied to biographical texts. In *2nd International Conference on Systems and Computer Science*. Presented at the 2nd International Conference on Systems and Computer Science.

Balandier, G. (2006) *Le pouvoir sur scènes*. Paris: Fayard.

Cardon, D. (2010) *La démocratie Internet: promesses et limites*. Paris, France: Seuil, DL 2010.

Coleman, S., and Blumler, J. G. (2009) *The Internet and democratic citizenship: theory, practice and policy*. Cambridge, Royaume-Uni, Etats-Unis.

Davis, R., Holtz-Bacha, C., and Just, M. R. (Eds.) (2017) *Twitter and elections around the world: campaigning in 140 characters or less*. London, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord: Routledge,Taylor & Francis Group, 2017.

Giasson, T., Marland, A., and Esselment, A. L. (Eds.) (2017) *Permanent campaigning in Canada*. Vancouver.

Godmer, L., and Marrel, G. (2015) *La politique au quotidien. L'agenda et l'emploi du temps d'une femme politique*. Lyon: ENS Editions.

Manning, C. D., Raghavan, P., and Schütze, H. (2008) *Introduction to Information Retrieval*. (Cambridge University Press.). Cambridge University Press.

Roginsky, S., and Perrier, V. J. (2014) La fabrique de la communication des parlementaires européens : « Tweet ton député » et les « ateliers du député 2.0 ». *Politiques de communication* 3(2): 85–124.

Théviot, A. (2016) Les data : nouveau trésor des partis politiques ? : Croyances, constitutions et usages comparés des données numériques au Parti Socialiste et à l'Union pour un Mouvement Populaire. *Politiques de communication* 6(1): 137–166.