



Le Web comme miroir du travail politique quotidien ?

Reconstituer l'écho médiatique en ligne des événements d'un agenda d' élu

Guillaume MARREL, MCF de Science politique, Agor@ntic, LBNC-Cherpa, UAPV
guillaume.marrel@univ-avignon.fr

Vincent LABATUT, MCF d'informatique, Agor@ntic, LIA, UAPV
vincent.labatut@univ-avignon.fr

Marc EL BÈZE, PR d'informatique, Agor@ntic, LIA, UAPV
marc.elbeze@univ-avignon.fr

Le Web peut-il être appréhendé comme un espace de mise en scène du travail politique ordinaire ? Que restitue-t-il des événements de l'emploi du temps d'un élu ? Quelles traces le travail politique quotidien d'un maire laisse-t-il sur la toile, en dehors des campagnes électorales ? On se propose dans cette contribution de présenter les premiers résultats d'une enquête pluridisciplinaire en cours sur l'extraction et l'exploitation des données du Web concernant l'agenda personnel d'un leader politique local. La démarche reste expérimentale, monographique et configurationnelle.

Cette recherche se développe dans le cadre d'une collaboration de recherche pluridisciplinaire engagée à l'Université d'Avignon entre sociologie du travail politique et fouille de données en science informatique, autour du projet de recherche [Tr@nspolo](#)¹ dans le cadre de la FR [Agor@ntic](#). Elle part du postulat que l'activité quotidienne d'un élu est plus ou moins reportée et commentée en ligne par son propre service de communication, les professionnels de la politique et de l'information qui l'entourent, mais aussi de plus en plus par des « citoyens ordinaires », sur les pages institutionnelles, les blogs, les murs Facebook, Twitter et autres réseaux socio-numériques. Le Web est ainsi saisi idéalement comme un miroir pouvant refléter l'activité politique et en réalité, du fait de certaines réfractions, comme un prisme déformant. La maîtrise de cette déformation médiatique devient sans doute un enjeu stratégique de premier ordre avec l'Internet dont la dimension conversationnelle (Cardon, 2010) change vraisemblablement le rapport des leaders aux instruments de contrôle de leur image, à tel point que l'e-réputation fait aujourd'hui l'objet de conseils spécialisés (Frochot and Molinaro, 2010). L'objectif de la recherche est donc de parvenir à objectiver ces déformations pour ensuite les interpréter et, à terme, dégager les variables de l'écho web-médiatique du travail politique dans les années 2010.

1 [Tr@nspolo](#) : Transparence politique et organisationnelle dans les sociétés numériques.

La communication en ligne d'un maire urbain en cours de mandat

Jusqu'à la fin du XXe siècle, c'est principalement à travers la presse, la radio et la télévision que le pouvoir se donne à voir et se met en scène (Balandier, 2006). Les médias de masse se sont alors imposés comme les relais incontournables entre le personnel politique et les citoyens dans la consolidation des régimes démocratiques, contribuant à une redéfinition des règles du jeu du débat public, de la conquête et de l'exercice du pouvoir (Bailey, 1971), autour des principes généraux de publicité et de transparence.

Le développement de l'Internet public et du Web au début des années 1990 vient recomposer les logiques et le sens de l'information et de la communication, mais aussi l'analyse de leurs effets et l'interprétation de leur signification². Conçu initialement pour faciliter les échanges interindividuels et non pour adresser des messages à une masse de récepteurs, Internet marque l'avènement d'un média d'une nouvelle nature, à bien des égards incomparable aux médias de masses du XXème siècle (Cardon, 2010, p. 9). S'y confondent désormais la sphère de la conversation interpersonnelle ordinaire des interactions privées, et l'espace professionnalisé de l'information publique, des industries culturelles et de la politique. Réseau militaire puis universitaire initialement, le Web ouvre dans mes années 1990 l'espace public aux individus et ce faisant, il recompose les formes mêmes de la démocratie moderne (présupposé égalitaire, coopération, auto-organisation, participation, consensus...). Au fil des échanges profanes, « la société démocratique sortirait de l'orbite de la politique représentative » (Cardon, 2010, p. 99). Dans cette « démocratie électronique », les publics citoyens s'émanciperaient du paternalisme infantilisant des professionnels et des experts. Et cet espace public ouvert, décentralisé et renouvelé (Coleman and Blumler, 2009) fait alors l'objet de tentatives plus ou moins efficaces de reconquêtes de la part des institutions classiques : médias de masse, industries culturelles, entreprises et partis politiques. En cela, l'Internet change déjà globalement les règles du jeu politique plus ou moins stabilisées au siècle dernier.

Rapidement pensé comme un projet d'émancipation politique, Internet est désormais réinvesti par l'ensemble des acteurs centraux et institutionnels de la vie politique des démocraties représentatives. Mais la nature conversationnelle des formes politiques du Web limite *a priori* l'investissement partisan institutionnel (Blondeau-Coulet and Allard, 2007). De fait, l'Internet politique est longtemps resté un Internet militant (Granjon, 2001), reposant principalement sur les sites des groupes activistes périphériques, les blogs et pages Facebook des personnalités politiques, les sites de campagne des candidats où sont valorisés l'individualisation de l'expression et l'échange conversationnel.

Cette soit-disant « Web-incompatibilité » des partis politiques, trop souvent réduits au centralisme et à l'immobilisme sclérosé, est néanmoins aujourd'hui contestée par les faits et dans les travaux de science politique (Gibson and Ward, 2009; Greffet, 2012). Tous les partis sont aujourd'hui visibles et actifs en ligne. Les net-campagnes se systématisent dès 2002 en France et la « cyberisation » des organisations partisans semble s'accélérer à chaque nouveau scrutin. La

2 Les sciences sociales s'interrogent très rapidement sur les ressorts de la « société de l'information » et les effets des médias sur les comportements des publics, spectateurs plus ou moins passifs et variablement sensibles aux messages diffusés par les outils de propagande et de publicité. Mais elles questionnent également à d'autres niveaux les transformations médiatiques des règles et des formes de la compétition politique (personnalisation et désacralisation du pouvoir, politique-spectacle, réduction du discours politique à la « petite phrase », peopolisation...), de l'action publique (démocratie de l'opinion, construction des problèmes publics, théorie de l'agenda gouvernemental...) et du travail politique (dépendance accrue des acteurs politiques aux communicants professionnels et à leurs outils – coaching, relooking, media training – et aux commentateurs – sondeurs, journalistes et politologues).

plupart des travaux disponibles se focalisent néanmoins sur la manière dont l'Internet affecte les acteurs de la compétition électorale et concentrent donc l'attention sur les périodes de campagne et les acteurs collectifs qui les animent, au détriment des temps politiques ordinaires (Nicot, 2012). De fait moins nombreuses sont les enquêtes consacrées à l'usage du Web par les élus durant leur mandat et la manière dont les formes renouvelées de communication en ligne affectent le travail politique de représentation « hors campagne » (Marques et al., 2014; Norton, 2007)³. En France, de telles recherches ne sont pas nombreuses (Chibois, 2014), alors que les pratiques des élus et de leurs entourages contribuent à la restructuration « par le bas » des règles du jeu politique.

Or la communication politique en ligne dépasse largement le cadre et le moment d'intensité particulier auquel correspond la campagne électorale. Une fois élus, le député ou le maire semblent le plus souvent maintenir désormais une importante activité de communication numérique en ligne sur leur activité de représentation, via les sites institutionnels de leurs collectivités, leurs blogs, leurs pages Facebook et leurs comptes Twitter. Les travaux portant sur les années 2000 ont tendance à montrer qu'à l'Assemblée nationale « les sites ou blogs des députés n'ont qu'un rôle limité dans les stratégies électorales, que ce soit comme instrument de communication partisane ou comme outil de propagande et de mobilisation » (Nicot, 2012, p. 105) et que localement « internet n'est guère une voie de communication privilégiée par les acteurs partisans, bien que d'usage croissant » (Bué, 2011, p. 229). On émet néanmoins l'hypothèse selon laquelle cette activité de valorisation et de légitimation par la mise en récit et la mise en scène de l'action politique, occupe une place croissante dans les collectifs de travail politique, voire dans l'emploi du temps de l'élu lui même lorsque celui-ci investit personnellement et réellement l'échange conversationnel.

Comment le travail politique hors campagne est-il impacté par l'usage des outils d'information numériques ? Qui diffuse, quand, par quels outils et avec quels relais ? Les stratégies de communication de ce type d'élu sont-elles affectées par les nouveaux vecteurs de diffusion de l'information ? Les entreprises politiques à la tête des exécutifs municipaux maîtrisent-elles leur image numérique ? Peut-on identifier des modèles de diffusion de l'information concernant les événements de l'activité du maire sur le Web, les acteurs qui y prennent part, la rapidité et la structure de la propagation d'une information traitée de manière positive ou négative ? Enfin, les formes de la diffusion de l'information et les *feedbacks* qu'en restitue le Web rétroagissent-elles sur la manière de communiquer sur l'agenda du maire, voire sur la manière de « faire de la politique » et donc sur l'emploi du temps effectif de l'élu ?

L'écho web-médiatique de l'agenda personnel de l'élu

L'enquête s'inscrit dans le cadre d'un projet de recherche plus large engagé sur les configurations de travail politique local (collectivisation, apprentissages, confusions politico-administratives, effets de trajectoires...) (Demazière and Le Lidec, 2014) et l'examen de l'agenda et de l'emploi du temps des membres d'exécutifs locaux participant directement à la production de l'action publique (Godmer and Marrel, 2014a; Lefebvre, 2014). L'agenda personnel est saisi ici comme le témoignage objectivé d'une orchestration des temps, des déplacements et des rencontres (Veitl, 2001). Il est le reflet et la trace d'une cristallisation biographique située, d'une aptitude à gérer un espace-temps politique, à contrôler le sens pratique du métier d'élu et à investir

3 Il est certes artificiel de distinguer le « temps de campagne » et le « temps du gouvernement » dans le continuum indifférencié du travail politique indistinctement « électoraliste » et « administratif » (Marrel and Payre, 2006). Il s'agit ici de cibler par convention les périodes d'activité politiques les moins affectées par l'inévitable anticipation des échéances électorales, que ce soit pour le renouvellement du mandat en cours, ou bien dans la perspective de la conquête d'un autre trophée politique cumulable ou conduisant à une « reconversion ».

durablement un territoire qui donne accès au leadership (Marrel and Payre, 2006). Son caractère confidentiel rend difficile l'exploitation de cette source quasi inédite en science politique, les conditions d'accès au terrain étant particulières (proximité, confiance, instrumentation...). La dimension numérique des PDA (Personal Digital Assistants) généralisés depuis le début de la décennie 2010 transforme cependant les conditions d'exploitation (exportation et traitement statistique de masse sans saisie manuelle à partir d'un agenda papier). Mais cette numérisation affecte surtout l'organisation même du travail politique au sein des équipes politiques (actualisation en temps réel des agendas partagés en réseau fermé, accélération et saturation de l'emploi du temps) et enfin, dans une autre dimension, les stratégies de communication politique (publication des agendas en ligne). L'analyse des agendas politiques et des emplois du temps d'élus articule ainsi trois dimensions :

- L'analyse des logiques de la production de l'agenda (arbitrages collectifs sur la ventilation de l'espace-temps de l'élu dans ses activités de gestion publique et de représentation politique).
- L'analyse des contenus (distribution et volumétrie des rôles, tâches, espaces et interlocuteurs... qui constituent effectivement le quotidien de l'élu, accélération...)
- L'analyse du jeu médiatique autour de l'action politique et du « faire savoir » (travail de mise en scène de l'action politique, occupation du Web, allumage de contre-feux, contrôle de l'image...)

L'enquête dont il est ici question concerne principalement la troisième dimension. Les chargés de communication des politiques et les médias classiques communiquaient relativement peu sur l'emploi du temps des élus et sur leurs agendas. L'idéal de transparence (J Meijer et al., 2012) conduit désormais certains acteurs à jouer le jeu de la communication sur leur activité, souvent pour mieux dissimuler ce qui restera quoi qu'il arrive de l'ordre du secret d'alcôve. Le média Web vient potentiellement recomposer ce jeu autour de la mise en scène du travail politique, les stratégies de communication, la gestion de la transparence, le rapport au secret (Florini, 1998), jusqu'à l'organisation même de l'agenda possiblement affecté de manière rétroactive par d'éventuels calculs sur l'augmentation du risque médiatique (ne plus rencontrer tel type d'interlocuteur susceptible de ternir l'image de l'élu, ou de perturber la mise en œuvre d'une politique...).

Que se joue-t-il donc alors entre l'entreprise politique en action et la nébuleuse communicationnelle du Web ? Qu'est-ce que le Web laisse au final transparaître de l'activité concrète de ces personnalités publiques particulières que sont les élus en cours de mandat ? Peut-on aller jusqu'à générer à partir des innombrables messages en circulation sur la toile, une image de l'emploi du temps d'une personnalité politique ? Et, tout du moins, que nous apprend la comparaison entre cet écho médiatique et le contenu du PDA de l'élu ? Plus encore, le Web peut-il être saisi comme un espace de dévoilement de certaines dimensions du travail politique ? Parce que l'injonction à la transparence qu'il véhicule pousserait les entrepreneurs politiques à publier davantage d'éléments de leurs agendas, y compris de manière partielle ou déformée. Parce qu'également, le Web conversationnel serait susceptible de diffuser des informations non officiellement communiquées, voire indiscretes sur les rendez-vous, les réunions, les déplacements... de l'élu dans le cadre de son mandat.

Ajustements de la démarche d'investigation

Le projet avait initialement pour objectif d'utiliser une base de données d'événements d'un agenda électronique personnel récent comme référence dans l'analyse de contenu de ce que le Web et les réseaux socio-numériques pouvaient restituer de l'emploi du temps du maire de la ville

d'Avignon, dans le Sud de la France. Le choix du maire d'Avignon renvoie à la fois aux aspects pratiques liés à la proximité du terrain d'investigation expérimental et monographique avec le lieu de travail des chercheurs, mais aussi aux relations d'interconnaissance de l'équipe avec l'entrepreneur politique Cécile Helle élue en mars 2014 à la tête de la municipalité. Le projet envisageait de saisir le Web comme un vaste corpus de données à fouiller au moyen de méthodes de détection et de reconstitution d'événements. Les enjeux de recherche en informatique concernaient essentiellement la fouille de données et le traitement automatique du langage naturel (TALN, ou NLP en anglais) et pouvaient se décliner en sept étapes :

- La première étape consistait à extraire des données issues du PDA les informations pertinentes relativement à notre objectif, soit les événements eux-mêmes, mais aussi les différentes expressions utilisées pour les désigner.
- La deuxième étape portait sur l'identification de contenus accessibles via le Web et relatifs aux événements identifiés. L'éventail des méthodes applicables est ici relativement large : on peut envisager dans un premier temps de se concentrer sur des sites sélectionnés manuellement, puis d'utiliser des heuristiques ou une première analyse rapide des sources disponibles, afin de sélectionner automatiquement les sites les plus intéressants. Techniquement, le traitement associé à cette étape impliquait de définir des outils capables de naviguer sur le Web et d'en extraire l'information ciblée (*crawlers*).
- La troisième étape consistait à mettre en relation les événements identifiés côté agenda et côté Web. Le point technique consistait ici à définir des mesures de similarité afin de pouvoir comparer les événements. Il s'agissait de déterminer quelles expressions textuelles pouvaient raisonnablement être considérées comme faisant référence aux mêmes événements.
- Une fois la mise en relation entre les événements détectés dans le PDA et sur le Web effectuée, la quatrième étape consistait à récupérer le contexte de ces derniers dans le but d'obtenir le texte brut contenant les commentaires relatifs aux événements. Ce texte devrait alors être traité via des méthodes de NLP, afin d'extraire la nature de l'opinion exprimée.
- La cinquième et dernière étape concernait l'interface homme-machine (IHM). Il s'agissait de présenter les résultats obtenus de façon à la fois intelligible et pratique, à l'utilisateur final du logiciel, notamment de manière à permettre la comparaison entre agenda reconstitué et agenda de référence. Il pouvait être intéressant de lui offrir des fonctionnalités de navigation, permettant de raffiner la recherche initiale, éventuellement par une succession de recherches de plus en plus précises et poussées.
- Formellement, le dispositif pourrait également présenter des indicateurs statistiques : nombre d'occurrences de chaque événement repéré dans le Web, distribution des événements en fonction des espaces numériques où ils sont commentés (sites institutionnels, blogs, réseaux sociaux...), polarité des commentaires (plutôt positif / plutôt négatif), etc.
- La recherche portait enfin sur les enjeux de fiabilité de l'outil de fouille : comment interpréter les résultats ? Le non repérage d'écho médiatique en ligne d'un événement est-il lié à l'absence effective de commentaire sur le Web, ou bien à un défaut de programmation ?

Pour le politiste, il s'agissait d'expérimenter l'exploration d'un nouveau corpus de données en rapportant le travail politique aux actes de communications en ligne qu'il est susceptible de susciter. Mais la dimension pluridisciplinaire de la recherche implique également une division des tâches spécialisées. Celle-ci suppose d'abord une réflexion sur les conditions d'un travail pluridisciplinaire équilibré et non instrumental. Les informaticiens trouvent dans cette étude de cas l'occasion d'éprouver la robustesse et la portabilité de leurs instruments, mais aussi des défis inédits, comme en matière de reconstitution d'événements. La pluridisciplinarité donne en retour accès au politiste à des dispositifs techniques de recherche ou de construction de données spécialisés et formalisés

qu'exige la nature singulière du Web, comme terrain de recherche. Le contrôle élémentaire des conditions de production des données suppose un minimum de compréhension des dispositifs mis en place, de leur fonctionnement, et des inévitables biais qu'ils incorporent.

Les difficultés d'accès à l'agenda personnel du maire ont conduit à resserrer l'expérience sur le seul écho web-médiatique disponible en ligne des événements de l'emploi du temps du maire. Il s'agit donc de fouiller le Web le plus largement possible pour y recenser l'ensemble des informations, communications et commentaires relatifs aux diverses interactions constitutives du travail politique de l'édile, afin de reconstituer l'agenda-miroir. La contribution restitue le test manuel effectué sur quatre semaines de travail édilitaire et les difficultés de la construction d'un outil de fouille automatisée.

Le test de fouille manuelle : *Google* - « Cécile Helle » [30/03/2015-26/04/2015]

Le test de fouille manuel des échos de l'emploi du temps du maire d'Avignon Cécile Helle a été réalisé sur une période de quatre semaines consécutives du 30 mars au 16 avril 2015. Cette séquence a été identifiée comme une période d'activité politique ordinaire, sans événement politique national ou local particulier et en dehors de vacances scolaires. Elle est aussi distante d'au moins un mois avec la récupération des données réalisée le 7 juin 2015, afin de travailler sur un état du Web et un retour des résultats du moteur de recherche par hypothèse stabilisés par la clôture théorique de l'activité de commentaire des événements passés.

La collecte des résultats a été effectuée à l'aide du moteur de recherche *Google* par la saisie du patronyme complet de l'élue pour élargir les retours et mesurer l'ampleur du « bruit » : <cécile helle> doublé de l'utilisation de la fonction « Custom range » permettant de personnaliser le tri à partir de bornes chronologiques, triant vraisemblablement à partir des dates de publication. Par défaut, l'algorithme a classé les résultats par pertinence décroissante.

La démarche manuelle peut se décomposer en six étapes :

- 1) la saisie des mots clefs dans la fenêtre du moteur de recherche et l'exécution de la requête.
- 2) la copie des résultats dans un tableur : lien, adresse, résumés *Google*...
- 3) le traitement formel des résultats (alignement, séparation en champs distincts ; dates, lien, contenus...).
- 4) l'analyse manuelle des liens un par un, par ouverture de la page web, lecture visuelle et recherche dans le texte sur le mot « helle ».
- 5) la catégorisation inductive des résultats dans 18 items
- 6) l'analyse

Les 18 items sont répartis comme suit :

1	Rang	n°
2	Titre de la page	Lien actif Google
3	Adresse	Adresse web
4	Date identifiée par Google	Date de mise à jour indiquée par Google
5	Résumé <i>Google</i>	Contenu du résumé fourni par Google
6	Actif	Lien effectivement actif : o/n

7	CH	Contenu concernant l'élue : o/n
8	Type support	Catégorie de support web (article de presse, billet de blog, pdf...)
9	Source	Acteur à l'origine de la publication (association, journaliste, institution, élu...)
10	Type contenu	Nature du contenu (interaction, publication, sollicitation, information...)
11	Type d'interaction	Nature de l'interaction (
12	Thème interaction	Thématique de l'objet (agriculture, commerce...)
13	Evénement	Description de l'événement
14	Date interaction	Date de l'événement
15	Heure interaction	Heure de l'événement
16	Pertinence date	Pertinence de la date de l'événement dans la période étudiée : o/n
17	Lieu interaction	Lieu de l'événement
18	Commentaires	-

Les champs 6 à 18 font l'objet d'une saisie manuelle. Les champs 8, 9, 10, 11 et 12 renvoient à une catégorisation construite de manière inductive, par regroupements ou distinctions progressifs des modalités découvertes sur les pages consultées. Les types de support (8) se répartissent principalement en sept catégories :

Article de presse
Article de site
Article de Blog
Pdf : document
Post facebook
Post twitter
Vidéo Dailymotion

Les types de sources (9) pertinentes pour l'étude se répartissent en neuf catégories :

Association
Elu local
Elu national
Entreprise
Institution publique
Institution municipale
Journaliste
Militant
Parti

Les types de contenu (10) sont réduits à quatre modalités :

Information impliquant le maire
Interaction
Publication
Sollicitation

Parmi eux les types d'interaction (11) se déclinent ainsi en onze catégories reprenant plus ou moins précisément les types d'interaction repérés dans l'analyse de l'emploi du temps d'une vice-présidente

de région entre 2010 et 2012 (Godmer et Marrel 2014)⁴.

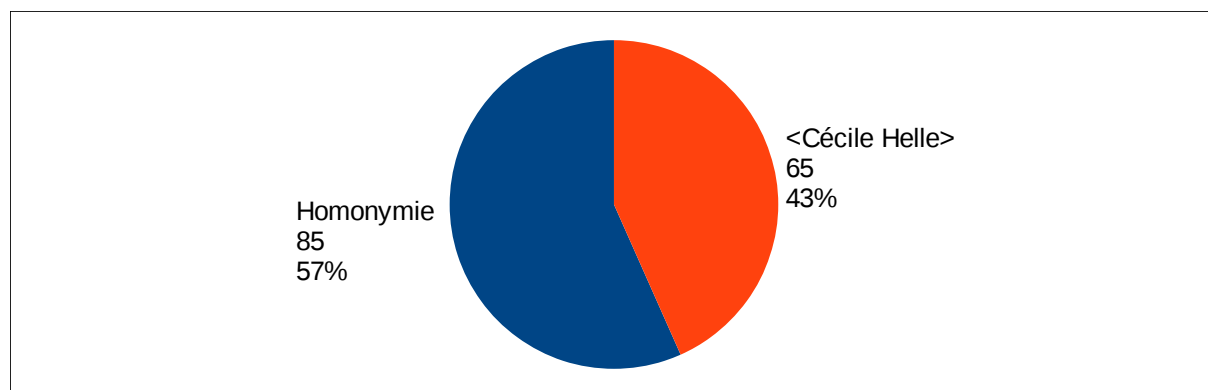
Cérémonie
Communiqué
Conseil
Dîner
Discours
Interview
Manifestation sportive
Rencontre
Réunion
Signature
Visite

La dernière catégorie (12) décline les champs d'action publique et politique auxquels l'édile est susceptible de participer.

150 liens, 18 interactions : un agenda-miroir lacunaire

La recherche ainsi calibrée fournit 150 résultats sur 15 pages. Seul 6 liens récupérés sont inactifs au moment de l'extraction. Un autre est en accès payant par abonnement (presse locale). Les 143 autres sont accessibles en ligne. L'inactivité des liens ou leur accès payant n'empêche pas la production d'un résumé Google, qui permet alors une analyse superficielle. Sur l'ensemble des 150 liens une importante minorité de 43,3 % (soit 65) concernent effectivement plus ou moins directement la maire d'Avignon.

Pertinence des 150 liens fournis par Google



Les 85 autres liens sont le plus souvent des « erreurs » compilées par homonymie sur le nom de l'élue : autres personnes dans le monde portant le nom Helle et pages de langues germaniques principalement utilisant le nom commun « helle » qui signifie « pierre » en Danois ou « clair » en Allemand, ou comportant la chaîne de caractère <helle>, comme dans « Michelle » ou « La Rochelle ». Cinq liens ne contiennent aucune mention du mot « helle », mais une partie seulement

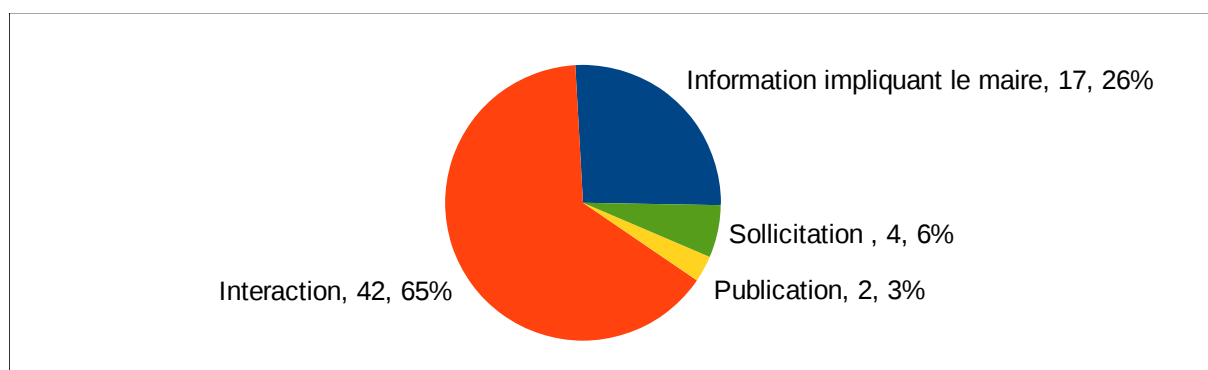
4 À savoir par ordre décroissant d'importance en volume horaire pour une période donnée : Réunion / Déplacement / Rendez-vous / Cérémonie / Repas de travail / Manifestation publique / Débat public / Contact médias / Spectacle / Visite / Action collective / Bureau.

associé à « cécile », sans qu'il y ait nécessairement continuité.

Le classement par pertinence du moteur Google est relativement fiable puisque le premier lien incohérent n'arrive qu'en 61^{ème} position et que les cinq derniers liens relatifs à l'élue se positionnent en 62, 63, 66, 75 et 93^{ème} positions sur 150. Cette expérience montre qu'il est utile d'explorer systématiquement jusqu'à la 6^{ème} page de résultats rapportés par Google.

Parmi les 65 pages concernant plus ou moins directement Cécile Helle, plus des deux-tiers (42) relatent plus ou moins directement une interaction de l'élue renvoyant à son emploi du temps. Une petite minorité des liens font référence à des publications officielles signées du maire (marchés public) et à des événements au cours desquels le maire est sollicité (articles concernant des rassemblements de citoyens demandant une audience). Le quart restant renvoie à des informations dans lesquelles le nom du maire apparaît sans toutefois qu'il ne soit fait référence à une action précise et datée renvoyant à un événement constitutif de l'emploi du temps de l'élue.

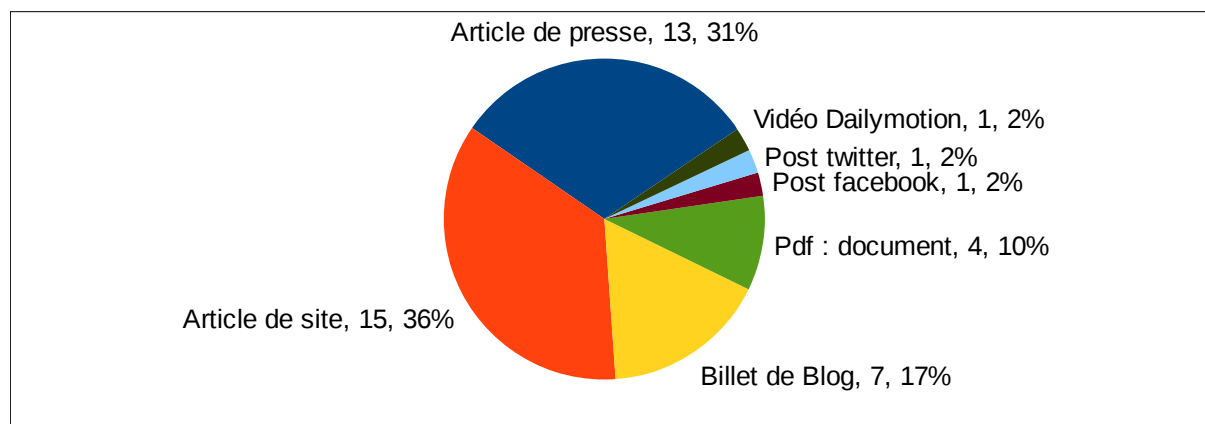
Répartition par types de contenus des 65 pages concernant l'élue



Les 42 liens fournis par Google permettant de renseigner un événement de l'agenda du maire renvoient pour l'essentiel à des supports classiques de type article de site web d'organisations (36%), de site de presse (31%) et de blog (17%). Le moteur renvoie encore pour 10 % des liens à des documents PDF produits par des organisations. On recense enfin trois sources marginales dans les résultats de Google : les deux réseaux socio-numériques Twitter et Facebook et la plate-forme vidéo Dailymotion. L'utilisation du moteur de recherche Google est ici clairement en cause. Ces espaces d'information appartiennent à des sociétés directement concurrentes des filiales Google+ et YouTube, ce qui d'après Google induit des difficultés techniques rendant leur indexation trop coûteuse pour être réalisée⁵. En pratique, l'existence des pages hébergées par ces services est bien répertoriée, mais pas leur contenu, ce qui empêche toute recherche un tant soit peu poussée. Malgré les performances de Google et son statut de « référence standard » en matière de recherche d'information en ligne, l'expérimentation montre qu'il faudra combiner les résultats de plusieurs moteurs et comparer les retours de *Google* avec les deux autres « leaders » du marché : l'un commercial : *Microsoft Bing* ; l'autre issu du « libre » : *DuckDuckGo*.

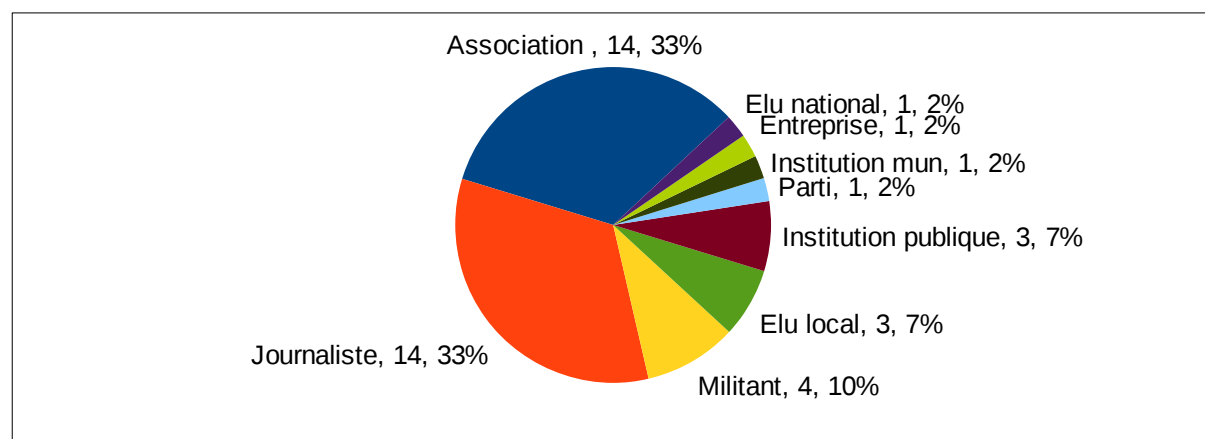
5 <http://searchenginewatch.com/sew/news/2325343/matt-cutts-facebook-twitter-social-signals-not-part-of-google-search-ranking-algorithms>

Types de supports web recensés parmi les 42 liens relatifs à l'agenda du maire



L'analyse montre encore que ces 42 liens renvoient pour les deux-tiers à des textes ou des documents produits par des acteurs associatifs non politiques (33%) et des journalistes (33%). L'action du maire est par ailleurs évoquée ou commentée par des militants politiques (10%), des élus locaux (7%), des institutions publiques (7%) ou encore de manière plus marginale par des élus nationaux, des responsables partisans, les institutions municipales ou des entreprises.

Types de sources (auteurs) recensés parmi les 42 liens relatifs à l'agenda du maire

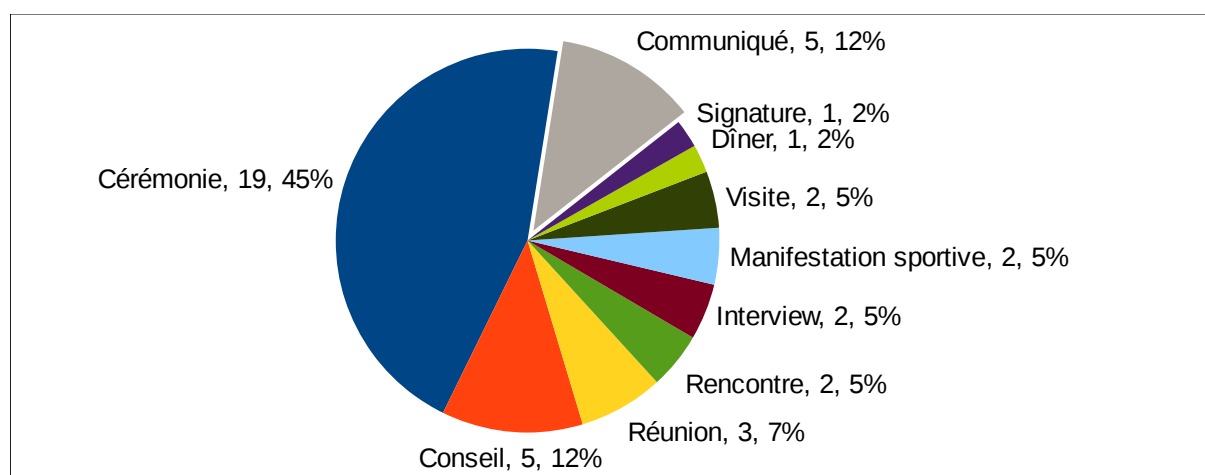


Alors que l'emploi du temps de la vice-présidente régionale étudié antérieurement, à partir de son agenda personnel, était principalement occupé, outre les temps de déplacement, en rendez-vous et en réunions de travail, l'écho web-médiatique de celui du maire est quant à lui composé pour près de la moitié de « cérémonies politiques » (45%). Le web fouillé selon notre protocole donne d'abord à voir un maire qui inaugure dans la ville divers événements comme un salon des loisirs mécaniques, des rencontres ESS et « territoires solidaires », un jardin partagé, un marché paysan et la réouverture d'une enseigne de grande distribution discount. Une élue qui assure également le discours d'ouverture d'un programme régional de découverte des talents et du colloque du réseau viticole des 80 ans de l'INAO (Institut national de l'origine et de la qualité), mais aussi le lancement d'un AAP sur l'habitat participatif. Une maire qui participe enfin à une conférence organisée par l'ambassade de Chine et qui assiste aux cérémonies nationales du souvenir et de la déportation.

L'image reproduite ensuite est celle de l'élue siégeant au conseil municipal ou communautaire (12%), lorsque l'actualité politique donne à ces assemblées régulières une importance singulière comme à la mairie les 8 et 24 avril où fut soldé le conflit entre le maire et ses adjoints EELV à propos de l'abatage de platanes, ou à l'occasion de la discussion médiatisée du Contrat de Plan Etat-Région à l'agglomération le 13 avril.

De manière attendue, le web compile des actes d'information et de communication qui portent principalement sur la partie publique, visible ou médiatisée du travail politique du maire. On y trouve également des traces de rencontres avec des personnels grévistes de l'hôpital, des interviews datant de 2012 et 2013, le classement d'une course à pied locale à laquelle le maire a participé, des visites de salons ou d'acteurs économiques locaux, la signature d'une convention sur le BTP, un dîner gastronomique au profit de l'AVAC (Aider à vivre avec le cancer).

Type d'interactions recensées parmi les 42 liens relatifs à l'agenda du maire



La fouille restitue enfin des actes de communication écrits datés mais ne donnant pas nécessairement lieu à l'occupation d'un temps de travail délimité identifiable, mentionnés ici comme des « communiqués », au nombre de cinq sur la période, concernant un grève du ramassage des ordures ménagères, l'annonce d'une solution d'aménagement routier proposée dans un courrier officiel à un ministre et un appel à mobilisation des abstentionnistes pour le second tour des élections départementales.

Ces 42 liens dessinent au final un agenda lacunaire et un emploi du temps très partiel. Le sondage effectué sur une période de quatre semaines délimitée dans le paramétrage personnel du moteur de recherche Google rapporte des liens qui évoquent des interactions hors périodes. Quinze liens renvoient à treize événements distincts antérieurs au 30 mars 2015, début de l'observation, parce que l'activité d'information et de commentaire en ligne s'étale naturellement sur une période jusqu'à 10 jours postérieure à la phase observée. Une minorité de huit événements plus anciens sont restitués par Google, remontant à plus de trois, six et douze mois et pour les plus lointains à trois ans (2012). Deux autres liens renvoient également à deux événements postérieurs à la fin de l'observation, le 26 avril 2015, parce que, à l'inverse, des messages mis en ligne durant la phase d'observation peuvent annoncer des événements à venir. C'est le cas ici de l'annonce du lancement d'un AAP de la mairie sur l'habitat participatif prévu le 20 mai et de l'inauguration par le maire d'un jardin partagé le 6 juin suivant.

Entre les deux, il reste 25 liens qui documentent en réalité 18 événements différents répartis sur les quatre semaines. Les interactions documentées ou commentées en ligne sont réparties sur 14 journées de la période, soit exactement la moitié des 28 jours observés, avec deux tendances :

- une sensible augmentation du nombre d'événements par semaine sur les séquences les plus récentes, comme si le moteur avait plus de facilité à retrouver des interactions moins anciennes. Après examen, le travail opéré par Google dans et sur le temps reste obscur. L'observation plus précise de la date de publication des pages web recensées par Google montre que dans 14 des 25 liens, l'interaction est commentée le jour même. Quatre liens annoncent un événement de deux à quatre jours avant qu'ils ne se produise. Quatre autres pages sont publiées le lendemain et encore deux quatre jours après.
- une faible concentration des événements reconstitués par journée : on compte deux interactions médiatisées pour seulement trois des 14 jours documentés. Le moteur renvoie le plus souvent un événement par jour.

Répartition sur quatre semaines des 18 interactions de l'emploi du temps de l'élue reconstituées à partir du web

S1 n=4	L30	M31	M1	J2	V3	S4	D5
				E1	E3		
				E2	E4		
S2 n=2	L6	M7	M8	J9	V10	S11	D12
		E6	E7				
S3 n=5	L13	M14	M15	J16	V17	S18	D19
	E8	E9	E10	E11			E12
S4 n=6	L20	M21	M22	J23	V24	S25	D26
			E13	E14	E15	E17	E18
					E16		

- Interactions mentionnées dans un seul lien
- Interactions mentionnées dans deux liens
- Interactions mentionnées dans trois liens

Le test permet encore de rapporter chaque interaction reconstituée à l'importance qu'elle aura prise en ligne, en comptant le nombre de liens qui renvoient au même événement. La majorité des interactions ne sont mentionnées qu'une seule fois dans les résultats de Google. Mais pour six d'entre elles on relève deux liens différents, attestant d'un plus vif intérêt de l'espace médiatique pour des événements à enjeux ou bien impliquant différents acteurs : une cérémonie relative à la culture, une visite dans un salon du tourisme, un communiqué sur la grève des éboueurs, une réunion sur l'aménagement du territoire et un conseil municipal mouvementé. Un seul événement est présent à trois reprises dans les résultats de la recherche du fait de la popularité des acteurs

impliqués, mais aussi des réactions politiques qu'il suscite : la réouverture dans un quartier périphérique de la ville d'un magasin *Leader Price* par le très médiatique Jean-Pierre Coffe, en présence du maire, qui ne manque pas de susciter les sarcasmes de l'animateur du blog qui « a pour but d'informer sur les actions entreprises par les élu(e)s du Front de Gauche à la mairie d'Avignon », et qui titre « Jean-Pierre Coffe et Cécile Helle, VRP de Leader Price⁶ ».

Mais au final, l'agenda reconstitué est loin de correspondre à l'agenda théorique comportant cinq à six plages horaires réservées par jour en moyenne, par comparaison avec l'étude réalisée sur celui d'une vice-présidente de région (Godmer and Marrel, 2014a). Le test manuel de fouille de l'écho web-médiatique de l'agenda ne rapporterait que 10 % des interactions planifiées dans le PDA de l'élue. Plusieurs hypothèses doivent alors être envisagées : une première totalement improbable supposerait que l'agenda réel du maire d'Avignon demeure assez léger.

Une seconde hypothèse renvoie à l'idée que l'information en circulation sur le Web concernant les interactions de l'emploi du temps de l'élue resterait limitée, avec ici trois sous-hypothèses complémentaires :

1) L'activité en ligne de mise en scène du travail politique quotidien par l'entrepreneur politique est réduite, parce que l'élue ne le souhaite pas ou n'en saisit pas l'importance et ne dispose donc pas d'un service spécifique et de plate-formes de diffusion actives (blog, Facebook, Twitter). De fait, peu active, la page publique Facebook de l'élue ne publie aucun poste entre le 11 février et le 9 mai 2015. De même, au moment de l'enquête, les deux derniers tweets de l'élue datent du 11 janvier et du 11 février 2015. Son blog de campagne n'est plus actif depuis le 30 mars 2014 et n'a pas été remplacé par un nouvel espace de communication équivalent depuis l'élection.

2) l'activité locale de commentaire journalistique publiée en ligne reste modérée, parce que la structure du champ médiatique local est à l'échelle des enjeux et des publics de la vie politique d'une petite ville. Cette hypothèse doit être relativisée au vu du recensement des articles de presse incluant <Cécile Helle> sur deux sites significatifs : *ledauphine.com* qui publie 45 papiers citant l'élue entre le 30 mars et le 26 avril 2015, alors que Google n'en identifie qu'un seul ; *laprovence.com* qui en publie 19, contre 1 seul dans Google et *citylocalnews.com*>Avignon qui en publie 13, encore une fois contre un 1 dans Google.

3) l'activisme cybermilitant reste mesuré sur ce territoire du Sud-Vaucluse limitant les relais d'informations et commentaires sur ce que fait le maire dans la blogosphère, Facebook et Twitter. Le recensement des acteurs associatifs, militants et « citoyens » présents de manière significative en ligne sur le territoire reste à faire⁷.

6 <http://fdgavignon.fr/inauguration-de-leader-price-pont-des-deux-eaux/>

7 Le protocole d'enquête en cours intègre une société de web marketing locale (semaweb.com) spécialisée dans la veille et la gestion d'e-reputation, susceptible de nous aider au repérage des cyber-leaders d'opinion locaux.

42 liens - 18 événements pertinents dans la période [30/03-26/04] - 13 événements antérieurs – 2 événements postérieurs

Nbr liens	Rang GG	Adresse	Date GG	Actif	Type support	Source	Type interaction	Thème	Nbr Evt	Événement	Date	Heure	Lieu
1	53	www.arhImpacacorse.com/pages.../ped.../plan-du-site/	18/04/15		Article de site	Association	Interview	Habitat	1		2012		Avignon
2	27	www.lalogitude.com/	23/04/15		Article de site	Association	Interview	Viellisse	2		30/11/13		Avignon
3	5	www.dailymotion.com/video/x2lpfzu	03/04/15	Non	Vidéo Dailymotion	Journaliste	Cérémonie	Élections	3	Victoire du 30 mars 2014	30/03/14		Avignon
4	32	www.cbeventoux.com/images/pdf/RencontreESS.pdf	11/04/15		Pdf : document	Institution publique	Cérémonie	ESS	4	Inauguration Rencontres ESS	18/10/14	9h	Avignon
5	21	www.grandavignon.fr/.../projet-courtine-confluence/	07/04/15		Article de site	Intstitution municipale	Réunion	Aménagement	5	Quartier Courtine	19/11/14		Avignon
6	13	https://www.facebook.com/.../893217850709590:0	02/04/15		Post facebook	Militant	Cérémonie	Commerces	6	Annonce réouverture du Leaderprice	19/12/14	Matinée	Avignon : Pont-des-deux-Eaux
7	54	www.townglance.fr/city/avignon	15/04/15		Article de presse	Journaliste	Conseil	Aménagement	7	Conseil communautaire	20/12/14		Avignon
8	23	www.calameo.com/books/0010381831e40420d1f63	31/03/15	Non	Article de presse	Entreprise	Visite	Entrepreneurs	8	Visite patronat local	?		?
9	37	www.lamarseillaise.fr/.../37992-la-direction-recule-les-...	14/04/15		Article de presse	Journaliste	Rencontre	Hôpital	9	Grève à l'hôpital	?/03/2015		Avignon
10	31	www.sud-chu-caen.fr/.../le-centre-hospitalier-d-avigno...	31/03/15		Blog	Militant				Rencontre des grévistes à l'hôpital	?/03/2015		Avignon
11	11	cluster005.ovh.net/~marseilleh/.../home/homeamf2015/	08/04/15	Non	Article de site	Association	Cérémonie	Loisirs motorisés	10	Inauguration festival	20/03/15		Avignon : Parc des expositions
12	6	irfss-paccac.croix-rouge.fr > Actualité	31/03/15		Article de site	Institution publique	Cérémonie	Formation	11	Remise de diplôme			Avignon IFSI
13	62	https://plus.google.com/107957699539151397621	30/03/15		Blog	Militant	Communiqué	Élections	12	Appel à la mobilisation 2d tour	22/03/15		Avignon Mairie
14	58	https://twitter.com/fareiodanoff	31/03/15		Post twitter	Elu local							
15	24	www.humanite.fr/festival-davignon-2015-olivier-py-p...	30/03/15		Article de presse	Journaliste	Cérémonie	Culture	13	Présentation Festival d'Avignon	27/03/15		Avignon
16	7	www.journalzibeline.fr/.../la-region-fait-connaître-ses-...	02/04/15		Article de presse	Journaliste	Cérémonie	Culture	1	Présentation Tal'ents Région #3	02/04/15		Avignon Mairie
17	36	www.presseagence.fr/.../avignon-talents-region-jeudi-...	02/04/15		Article de site	Association							
18	50	www.education-ambchine.org/.../info115092.htm	02/04/15		Article de site	Institution publique	Cérémonie	International	2	Rencontre chine			
19	12	www.ditex.fr/salon/programme/	01/04/15		Article de site	Association	Visite	Tourisme	3	Annonce visite salon Ditex	03/04/15	10h30	Avignon : Parc des expositions
20	22	www.quotidiendutourisme.com/.../salon-ditex-c-est-pa...	01/04/15		Article de presse	Journaliste							
21	25	www.escandihado84.com/marche-des-producteurs-a-...	04/04/15		Article de site	Association	Cérémonie	Agriculture	4	Inauguration Marché des producteurs			Montfavet
22	3	www.ledauphine.com/.../cecile-helle-lance-un-appel-a-...	08/04/15	€	Article de presse	Journaliste	Communiqué	Urbanisme	5	Grève éboueurs	07/04/15		Avignon Mairie
23	14	www.francebleu.fr/.../les-eboueurs-du-grand-avignon-...	07/04/15		Article de presse	Journaliste							
24	28	www.laprovence.com/.../abattage-des-platanes-poursui...	09/04/15		Article de presse	Journaliste	Conseil	Conseil municipal	6	Exclusion vert	08/04/15	Soirée	Avignon Mairie

25	20	michel-vauzelle.fr/contrat-de-plan-etat-region-dans-le-...	13/04/15		Blog	Elu local	Réunion	Développement territorial	7	Réunion de concertation du CPER	13/04/15	9h30-12h	Avignon : Préfecture
26	44	www.pertuisien.fr/flash.php?flash=68267	13/04/15		Article de site	Association							
27	41	info-jobs.fr/et-pourquoi-pas-letranger/	10/04/15		Blog	Militant	Cérémonie	Jeunes	8	Ouverture Rencontre Territoires Solidaires	14/04/15	10h00	Avignon Mairie
28	17	www.tpbm-presse.com/l-office-btp-et-la-ville-d-avign...	15/04/15		Article de presse	Journaliste	Signature	BTP	9	Signature convention Office BTP	15/04/15	?	Avignon Mairie
29	26	www.bio-bretagne-ibb.fr/.../INAO-Colloque-Program...	16/04/15		Pdf : document	Association	Cérémonie	Agriculture	10	Accueil colloque INAO	16/04/15	9h	Avignon : Palais
30	15	https://lavignonnaise.wordpress.com/.../julienpanzier/	22/04/15		Article de site	Association	Manifestation sportive	Loisirs sportifs	11	Participation à la course L'avignonnaise (classement)	19/04/15	Matinée	Avignon
31	56	www.nikrome.com/pdf/Avignonnaise2015.pdf	19/04/15		Pdf : document	Association							
32	1	www.citylocalnews.com/Avignon/Dans l'actu	22/04/15		Article de presse	Journaliste	Cérémonie	Commerces	12	Inauguration Leader-price	22/04/15	17h	Avignon : Pont-des-deux-Eaux
33	18	www.leaderprice.fr/.../OUVERTURE%20AVIGNON.p...	22/04/15		Pdf : document	Entreprise							
34	29	fdgavignon.fr/inauguration-de-leader-price-pont-des-...	23/04/15		Blog	Parti							
35	57	www.segolène.fr/nav-newsarchives	19/04/15		Blog	Elu national	Communiqué	Transport	13	Courrier au ministre / LEO	23/04/15	?	Avignon Mairie
36	46	www.alvinet.com/similaires/cognac.../25106128	11/04/15		Article de presse	Journaliste	Conseil	Agriculture	14	Conflit Velleron	24/04/15	Soirée	Avignon Mairie
37	47	www.alvinet.com/similaires/torchon-brule.../2466983...	10/04/15		Article de presse	Journaliste							
38	39	creavignon.over-blog.fr/.../discours-de-jean-pierre-cer...	24/04/15		Blog	Elu local	Conseil	Conseil municipal	15	Discours vert exclu			
39	30	aideravivreavec.fr/calendrier/	25/04/15		Article de site	Association	Dîner	Santé	17	Dîner AVAC	25/04/15	Soirée	Capelongue
40	60	article.wn.com/.../Homage-aux-victimes-et-rescape...	26/04/15		Article de site	Association	Cérémonie	Mémoire	18	Cérémonie nationale du souvenir de la déportation	26/04/15	Matinée	Avignon Mairie
41	51	habitatparticipatif-paca.net/eh84	19/04/15		Article de site	Association	Cérémonie	Habitat	1	Announce lancement AAP HP	20/05/15		?
42	63	jardins-partages.org/spip.php?page=sedna...	21/04/15		Article de site	Association	Cérémonie	Environnement	2	Inauguration jardin partagé	06/06/15		Avignon

Troisième hypothèse générale, le test de fouille est lacunaire, le choix du moteur de recherche et des mots clefs utilisés est limitatif et les résultats « passent à côté » d'informations sur l'emploi du temps pourtant bien disponibles en ligne. Cette hypothèse est la plus probable, compte tenu du test comparatif effectué sur les trois sites de presse. Dans l'étude de l'écho web-médiatique du travail politique, comment objectiver les résultats, lorsqu'on ne sait pas si ce qu'on obtient est le reflet fiable des informations et commentaires nécessairement partiels publiés en ligne sur les interactions de l'élue, ou bien une image déformée par la boîte noire de l'algorithme et les mots clefs utilisés pour la recherche ? De fait le test manuel réalisé sur le seul moteur Google montre de grandes lacunes. Outre l'instabilité des résultats dans le temps et selon la durée de la période interrogée, la recherche est nécessairement personnalisée par le profil Google de l'utilisateur, à partir d'un compte personnel ou de l'adresse IP utilisée, introduisant un biais lié à la manière dont l'algorithme structure les résultats de manière à les adapter à l'historique de l'utilisateur. La faible présence à l'état de « quasi échantillon » des articles de presse dans les résultats, doit également être interrogée, de même qu'il serait nécessaire de compléter la fouille à partir d'outils adaptés à Facebook et Twitter.

Le montage d'un outil automatisé de détection des interactions de l'élue

À terme, l'objectif de l'outil proposé est d'automatiser autant que possible le traitement manuel décrit précédemment. La version présentée ci-dessous doit cependant être considérée comme une étude de faisabilité, sujette à un certain nombre de limitations, détaillées plus loin. Le traitement implémenté se décompose en plusieurs étapes, adapté du projet initial :

1) Choix du moteur

La première étape consiste à invoquer un *moteur de recherche Web*. Il s'agit d'un service capable de traiter une requête et de renvoyer un certain nombre d'adresses correspondant à des pages Web jugées pertinentes. Dans cette première version, nous avons sélectionné le moteur de Google⁸, car il est celui qui indexe le plus de pages, et constitue un standard *de facto*. Une première limitation apparaît ici, car le moteur accessible programmatiquement est une version bridée de celui utilisé par un utilisateur humain. Plus précisément, il renvoie beaucoup moins de résultats (100 au plus) et, faute d'option « custom rang », ne fournit pas certaines informations importantes comme la date à laquelle la page a été indexée pour la dernière fois. Cette limitation pourrait être résolue en utilisant un ou plusieurs moteurs de recherche alternatifs, tels que Microsoft Bing⁹ ou DuckDuckGo¹⁰. Ces divers moteurs pourraient être combinés pour améliorer la recherche, mais une étude comparative serait également pertinente.

2) Filtrage par format

La deuxième étape consiste à filtrer les pages Web renvoyées par le moteur de recherche, de manière à évacuer celles qui ne sont pas exploitables. Dans cette première version, nous ignorons par exemple les documents qui ne sont pas des pages Web : PDF, fichiers texte, images, vidéos, etc. La limitation porte ici sur le fait qu'un utilisateur humain serait capable d'exploiter ces données, et il y a donc une perte d'information potentielle. Celle-ci pourrait être résolue de façon relativement simple pour les différents formats encodant du texte (PDF, MS Word, texte brut, etc.). De la même façon qu'un utilisateur humain a besoin d'un logiciel approprié pour lire un document MS Word, un

8 <https://www.google.com>

9 <http://www.bing.com/>

10 <https://duckduckgo.com/>

logiciel a besoin d'une bibliothèque logicielle appropriée pour accéder au contenu d'un tel fichier. En revanche, le verrou concernant les données multimédia (images, vidéos, audio) ne semble pas pouvoir être cassé à court ni à moyen terme. Le développement de technologies telles que le Web sémantique, la reconnaissance de visage, le résumé textuel d'images et la transcription automatique de flux audio n'est pas encore assez avancé pour le permettre.

3) Récupération des articles

Les pages restant après le filtrage sont ensuite téléchargées grâce aux adresses Web renvoyées par le moteur de recherche lors de la première étape. À cette occasion, il est possible que certaines adresses ne répondent pas et que les pages correspondantes soient donc inaccessibles. Cependant, le même problème se poserait pour un utilisateur humain. Les pages Web téléchargées sont enregistrées localement afin de pouvoir être traitées rapidement, et de permettre un accès rapide dans le futur. Cependant, les pages récupérées ne sont pas constituées uniquement de contenu pertinent : on y trouve également des liens vers d'autres sites, des menus, des mentions légales, de la décoration, etc. Il est donc nécessaire d'identifier quelle est la partie de la page sur laquelle l'extraction sémantique à suivre devra se concentrer. Nous appellerons cette partie l'*article* dans le reste de notre description. Nous avons développé un algorithme relativement élégant permettant de traiter ce problème de façon générique. Notre méthode peut, de plus, être manuellement adaptée à des sites ciblés pour leur intérêt, si nécessaire. Supposons par exemple que l'on veuille tout particulièrement traiter des articles du journal Le Monde. On devra alors identifier manuellement, dans le code source HTML d'une page Web du monde, la position du contenu pertinent (l'article visé), puis adapter notre algorithme de manière à cibler cette position. L'opération ne doit être effectuée qu'une seule fois, et permet par la suite de traiter efficacement toute page issue de ce site Web-là. Bien sûr, notre logiciel est susceptible de commettre des erreurs. C'est notamment le cas pour des sites codés *manuellement*, i.e. qui ne sont ni maintenus par des professionnels (journaux, entreprises, institutions, etc.), ni hébergés sur des plate-formes spécialisées (WordPress, Blogger, etc.). Ces sites ne respectent pas toujours les bonnes pratiques et les standards du Web, et peuvent tromper notre logiciel là où un humain serait plus robuste. Cependant, il constituent une minorité des sources qui nous intéressent.

4) Détection des entités nommées

L'étape suivante consiste à détecter dans l'article les mentions d'« entités nommées ». Au sens strict, une entité nommée est un concept du monde réel caractérisée par un nom propre : personne, entreprise, lieu, etc. Le but de l'opération est d'identifier les références à ces entités dans les articles traités. Le traitement repose sur des outils issus du TALN (Traitement Automatique du Langage Naturel). La version actuelle du logiciel permet de détecter les personnes, lieux et organisations (entreprises, institutions, associations...), ainsi que les dates (qui ne sont pas des entités nommées au sens strict). La principale limitation ici vient du peu d'outils existant pour le traitement du français (là où les anglophones sont bien pourvus). En effet, idéalement, la détection d'entités nommées devrait être complétée de plusieurs traitements pour extraire toute l'information disponible. Tout d'abord, il serait nécessaire de détecter les *situations anaphoriques*, i.e. les cas où une expression textuelle qui n'est pas un nom propre fait référence à une mention d'entité apparue auparavant dans le texte. La tâche de *résolution de coréférences* permet de déterminer à quelle mention cette expression fait référence. Par exemple, considérons le texte suivant : "François Hollande a déclaré à la presse ne pas en savoir plus sur cette affaire. Le président s'est ensuite envolé vers Paris pour participer au conseil des ministres." On voudrait déterminer automatiquement que non seulement "Le président" est une anaphore, mais qu'en plus elle fait référence à "François Hollande". Enfin, la désambiguïsation de mentions consiste à déterminer à

quelle entité exactement un groupe de mentions fait référence. Par exemple, on veut déterminer si l'expression "Michael Jordan" fait référence au fameux basketteur¹¹ ou au pionnier de l'apprentissage automatique¹². La résolution de ces limitations est envisageable à court ou moyen terme, car plusieurs méthodes existent déjà pour l'anglais, et les travaux pour les adapter au français (ou développer des solutions spécifiques à cette langue) sont en cours.

5) Filtrage par contenu

Une fois les entités détectées, elles sont utilisées pour effectuer un second filtrage. Dans l'hypothèse où la requête concernait une personne, nous allons tout d'abord vérifier que l'article considéré contient bien une mention explicite lui faisant référence. Si un intervalle temporel a été spécifié pour la recherche, il est également possible d'écarter les pages qui ne contiennent pas de date explicite située dans cet intervalle. Il s'agit ici d'écarter le bruit renvoyé par le moteur de recherche, i.e. les pages qu'il avait jugées pertinentes, mais qui ne le sont pas en réalité (du point de vue statistique, il s'agit de *faux positifs*). Notre outil permet également de définir d'autres critères pour ce filtrage, par exemple l'obligation de (ne pas) contenir certaines expressions textuelles jugées discriminantes.

6) Extraction des événements

On fait ici l'hypothèse simplificatrice qu'un article décrit un événement. La sémantique associée à cet événement est décrite par l'ensemble des entités détectées dans l'article, ainsi que par le titre de celui-ci. Les différents types d'entités correspondent aux dimensions utilisées pour caractériser l'événement : *où* (lieux où l'événement s'est déroulé), *quand* (dates de l'événement) et *qui* (personnes et organisation ayant participé à l'événement). Bien entendu, un article peut en réalité décrire plusieurs événements. Notre hypothèse simplificatrice est dictée par la limitation relative à la détection des entités nommées, que nous avons mentionnée précédemment. Le fait d'être dans l'incapacité de traiter les anaphores nous empêche de descendre à un niveau de granularité plus fin tel que le paragraphe ou la phrase.

7) Perspectives

Mises à part les différentes limitations relevées dans notre description, de nombreuses pistes existent pour le développement de notre outil. La plus intéressante est certainement de définir des méthodes de comparaison d'événements, afin de déterminer automatiquement quels articles décrivent les mêmes événements. Ceci permettrait notamment de mesurer la présence médiatique en ligne d'un événement donné, voire l'évolution de cette présence dans le temps. Cependant, cette fonctionnalité dépend à la fois de la bonne détection des événements, et de l'existence d'un mécanisme de désambiguïsation de mentions (qui plus est efficace). Ce développement est donc assujéti à la levée de certains verrous liés au TALN.

Une autre piste à creuser est celle de l'exploitation des médias sociaux. En effet, les services tels que Facebook ou Twitter sont très utilisés par la population qui nous intéresse, et constituent donc des cibles privilégiées. Cependant, les contenus publiés sur ces plate-formes ne sont pas indexés par Google, ce qui signifie qu'ils ne sont pas inclus dans les résultats des recherches menées via ce moteur. De plus, même si c'était le cas, ces plate-formes rendent extrêmement ardu l'accès programmatique aux données qu'elles hébergent, essentiellement pour des raisons commerciales (d'où leur non-indexation par Google). Il faut cependant noter que Twitter met à disposition des

11 https://fr.wikipedia.org/wiki/Michael_Jordan

12 [https://fr.wikipedia.org/wiki/Michael_Jordan_\(informaticien\)](https://fr.wikipedia.org/wiki/Michael_Jordan_(informaticien))

chercheurs des sauvegardes régulières de ses tweets. Il est donc possible de les exploiter *a posteriori*, avec un délai minimal de quelques mois. De plus, ces données sont volontairement incomplètes (elles ne constituent pas l'ensemble des tweets postés sur la période considérée).

Comparaison des résultats de fouille manuelle et automatique

Long a programmer et à paramétrer, l'outil automatique permet un traitement de l'information très rapide d'une dizaine de minutes maximum (principalement lié au temps d'accès au réseau et d'extraction des données en ligne), alors que la démarche manuelle a nécessité une vingtaine d'heure de travail.

Mais d'une manière générale, l'outil automatique est encore en l'état peu performant. Sur la même période de quatre semaines du 30 mars au 26 avril 2015, l'outil automatique provisoire rapporte 70 résultats avec les mêmes mots clefs, contre 150 pour la fouille manuelle, soit moins de la moitié. Huit liens sont filtrés par le moteur parce que jugés non exploitables en l'état actuel de l'outil, par exemple parce qu'il s'agit d'un document que la machine est encore incapable de déchiffrer. L'examen visuel de contrôle montre que deux documents PDF concernent bien l'élue, le premier citant Cécile Helle parmi les membres du CA de l'ANRU, l'autre citant le maire dans le compte rendu d'activité d'une école d'architecture venue enquêter à Avignon.

Parmi les liens lisibles automatiquement, 31 sont jugés comme non pertinents parce que ne concernant pas Cécile Helle. L'ouverture des liens montre pourtant que sept d'entre eux concernent bien Cécile Helle, sans qu'il s'agisse nécessairement d'un événement. Les 31 autres liens sont classés comme valides. L'outil y détecte les entités nommées parmi lesquelles <Cécile Helle> apparaît le plus souvent. Pourtant, là aussi plusieurs erreurs ont été détectées. Sur les sept pages dans lesquelles l'outil ne recense pas le nom de l'élue, trois citent pourtant bien nommément le maire. En revanche quatre liens pourtant validés renvoient soit à des homonymes soit à une partie seulement de la chaîne de caractères ayant servi pour la recherche (Michelle).

A quatre reprises encore l'outil affirme ne pas détecter d'événement parce qu'aucune date ou référence temporelle n'a été trouvée, ce que le contrôle visuel valide. Au final, l'outil détecte 23 liens considérés comme des événements contre 42 manuellement. Mais la détection automatique de l'inscription temporelle pose encore problème. L'utilisation programmée de Google par un robot interdit l'utilisation des bornes chronologiques pour le tri des résultats dans l'indexation quotidienne des pages du Web que le moteur réalise. L'opération de filtrage est donc effectuée par le robot *a posteriori* à partir de l'ensemble des retours provoqués par la recherche sur <Cécile Helle>, sur la base des dates repérées dans le texte. Ces dates sont parfois incohérentes comme dans 4 des 23 cas vérifiés, quand l'outil considère comme une date l'expression d'une grandeur et indique par exemple [24/4/2015-?/?/9800] pour un article de *citylocalnews.com* sur le conseil municipal du 24 avril.

Toutefois, malgré les fragilités constatées et l'imprécision des résultats produits par l'outil automatique de détection, notamment du point de vue des bornes chronologiques, on observe que pas moins de 13 interactions détectées parmi les 20 retours non ambigus correspondent à l'un des 18 événements reconstitués manuellement sur la même période. Dans des conditions presque équivalentes, le robot est donc encore loin d'être aussi performant que la détection visuelle humaine, mais son score n'est pas tellement plus mauvais que celui de l'homme, limité pour sa part par les choix « aveugles » (?) de l'algorithme de Google.

Résultats de l'outil automatique validés comme événements de l'emploi du temps de l'élue (20)

Rang	Titre page	CH	Date	Lieu	Entité nommée Cécile Helle	Recoupe manuelle
5	Discours de Jean-Pierre Cervantes pour le Conseil Municipal du 24 Avril 2015 - Le blog de creavignon	Oui	??/2013-??/5/2014	Avignon, Mme Helle, Paris	Cécile Helle,...	Oui
6	Une France qui renie son histoire ? - Fenêtre sur Edito	Oui	24/4/2011-??/2015	France, Italie, Paris	Cécile Helle,...	
13	Concertation CPER dans le Vaucluse	Oui	30/5/2015-??/2050	Avignon, Carpentras, Vaucluse	Cécile Helle,...	Oui
16	Société Française des Roses - Visite de jardin par la Société : Altera Rosa à Avignon, exposition de roses et rosiers	Oui	??/1180-??/2016	Austin, Avignon, Bully, France...	? mais vérifié	
17	« Journal de bord d'une élue en pays FN » par Lilian Alemagna et Elsa Di Meo, une occasion manquée Toutelaculture « Journal de bord d'une élue en pays FN » par Lilian Alemagna et Elsa Di Meo, une occasion manquée	Oui	25/12/1905-26/4/2015	Algérie, France, Fréjus, Lucéram...	Cécile Helle,...	
27	Actualite Avignon/ Jean-Pierre Cervantès convoqué par le maire ce soir	Oui	8/4/2015-8/4/2015	Avignon	Cécile Helle,...	Oui
28	Dans l'actu Avignon/ Le Collectif de l'eau veut rencontrer Cécile Helle	Oui	??/2017-20/4/2020	Avignon, Pujaut, Sauveterre	Cécile Helle,...	Oui
29	Dans l'actu Avignon / Jean-Pierre Coffe et Cécile Helle, VRP de Leader Price	Oui	22/4/2015	Avignon	Cécile Helle,...	Oui
30	Dans l'actu Avignon/Suivez le conseil municipal en direct	Oui	24/4/2015-??/9800	Carpentras, Caumont...	Cécile Helle,...	Oui
34	Vaucluse	Oui	2/2/2011	Avignon, Cergy, France 2	Cécile Helle,...	
36	Projet Courtine-Confluence Communauté d'Agglomération du Grand Avignon : site officiel	Oui	??/20-??/2030	Avignon, Europe, Vaucluse, man	Cécile Helle,...	Oui
38	IMBE - Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale - Cécile Miramont	Oui	??/1004-??/2017			
40	Société: Opération de découverte des cultures urbaines Tal'ents Région à Avignon le 11 avril	Oui	12/7/2014-24/7/2014	Avignon, Miramas, Méditerranée...	Cécile Helle,...	Oui
44	Abattage des platanes : un adjoint au maire se rebiffe La Provence	Oui	??/2015	Montpellier	Henri IV	Oui
47	Au sommaire du iRavi/i n°128, avril 2015	Oui	25/12/2015	Avignon, Marseille, Noailles, Tunis	Cécile Helle,...	
52	Museumexperts : appels d'offres	Oui	??/1516-5/5/2015	Avignon, Mayot	Cécile Helle,...	Oui
56	Ditex : C'est parti pour deux jours intenses à Avignon - Salon sur Le Quotidien du Tourisme	Oui	??/2014-23/6/2015	Alberta, Avignon, Bali Emirates...	Cécile Helle,...	Oui
60	Le Centre Hospitalier d'Avignon souhaite supprimer 123 postes et 60 lits - Le blog de sud-chu-caen	Oui	??/2014-23/3/2015		Cécile Helle,...	Oui
61	A Avignon, un tramway semé d'embûches	Oui	8/4/1072-1/4/2015	Avignon, Vaucluse	Cécile Helle,...	
62	L'Office BTP et la Ville d'Avignon signent une convention	Oui	15/4/1073-15/4/2015	Avignon, Vaucluse	Cécile Helle,...	Oui

Résultats de l'outil automatique récupérés comme événements de l'emploi du temps de l'élue parmi les PDF et les liens jugés non pertinents

Rang	Titre page	CH	Date	Lieu	Entité nommée Cécile Helle	Recoupe manuelle
2	Avignon Motors Festival 2015	Récupéré				Oui
8	Habitat Participatif PACA - eh84	Récupéré				Oui
9	Félicitations aux Secrétaires médicales et médico-sociales diplômées - Avignon - Institut de Formation Croix-Rouge PACA & Corse	Récupéré				
32	Programme - DITEX Avignon 2015	Récupéré				Oui
46	Patrick Bloche - tous les articles sur Patrick Bloche - Le Point	Récupéré				
54	MARSEILLE / Région Provence-Alpes-Côte d'Azur - Agenda du 30 mars au 5 avril 2015 - La lettre économique et politique de PACA	Récupéré				
58	Habitat Rencontres Tsiganes	Récupéré				
20	*	Récupéré				
21	*	Récupéré				

Résultats de l'outil automatique validés comme événements de l'emploi du temps de l'élue : dates incohérentes

Rang	Titre page	CH	Date	Lieu	Entité nommée Cécile Helle	Recoupe manuelle
3	Avignon: Veolia chasse ses fuites Coordination eau bien commun	Oui	?/?/90-?/?/2020	Avignon, Barcelonette, Italie...	Cécile Helle,...	?
18	La Provence 15 avril : Avignon : gestion de l'eau : une guerre de trente ans contre veolia	Oui	?/?/90-?/?/2020	Arcachon, Autun, Avignon...	Cécile Helle,...	?
24	Les actualités	Oui	?/?/20-?/?/6000	Avignon, Durance, Vaucluse	Cécile Helle,...	?

Résultats de l'outil automatique concernant l'élue mais ambigus sur le statut d'événement de l'emploi du temps de l'élue

Rang	Titre page	CH	Date	Lieu	Entité nommée Cécile Helle	Recoupe manuelle
33	Les éboueurs du Grand Avignon de nouveau en grève ce jeudi / France Bleu	Oui	Aucun évènement détecté			?
43	Centre hospitalier d'Avignon : la direction recule, les personnels respirent - Journal La Marseillaise	Oui	Aucun évènement détecté			?
45	Vaucluse Avignon : les écologistes vont bloquer l'abattage des platanes	Oui	Aucun évènement détecté			?
55	MARSEILLE / Région Provence-Alpes-Côte d'Azur - Agenda Presse du 6 au 12 avril 2015 - La lettre économique et politique de PACA	Oui	Aucun évènement détecté			?

Conclusions

Le projet d'un examen abouti des usages du web et des réseaux socio-numériques dans la mise en scène du travail politique quotidien est loin d'être achevé. L'hypothèse d'un ajustement stratégique des pratiques concrètes de représentation sous la pression des entourages politiques en charge de l'image numérique des élus reste à démontrer. Elle supposerait de pouvoir suivre sur une période significativement longue de une à deux années, la composition d'un agenda personnel d'élu et l'écho web-médiatique de celui-ci. Cela implique d'abord d'accéder à cette source rare qu'est l'agenda électronique planifiant l'emploi du temps du maire, par nature secret et réservé à l'intimité de l'entourage proche de l'élue, qui le partage en réseaux protégé. La négociation de terrain en cours n'a de ce point de vue pas encore abouti et a retardé, modifié et limité le protocole d'enquête initialement envisagé. Dans l'attente d'un accès effectif à la base de référence de l'emploi du temps « réel »¹³ du maire, l'équipe s'est surtout consacrée au montage et à l'évaluation d'un outil de fouille et de détection d'événements opérationnel. Cette contribution en est le reflet mais elle dévoile aussi l'ampleur des ajustements nécessaires à la seule reconstitution de l'agenda de l'élue sur l'unique base des messages circulant en ligne.

Le suivi longitudinal des pratiques de gestion de l'e-reputation du maire implique ensuite de parvenir à reconstruire l'image miroir la plus complète possible de toutes les traces en ligne des interactions concrètes de l'élue. La difficulté principale repose ici sur l'absence de référence fiable quant au volume et la nature des réunions, des rendez-vous, des cérémonies et autres discours impliquant l'élue qui « passent en ligne », sous une forme ou sous un autre. Si bien que l'évaluation de la fiabilité des dispositifs de fouille de ces échos d'activité reste difficile, que ces dispositifs soient manuels ou automatisés.

Du point de vue du test de contrôle manuel, comme de l'automatisation de l'extraction des événements de l'emploi du temps d'une personnalité publique sur le Web et les réseaux socio-numériques, l'expérience montre qu'il faut :

1. Systématiser l'usage comparatif et compilatoire de plusieurs moteurs de recherche en complément de Google, dont la situation monopolistique engendre des biais d'indexation, malgré ses performances (*Microsoft Bing* et *DuckDuckGo*).
2. Fouiller ou programmer spécifiquement les sites de la presse locale mal indexés dans les résultats, en dépit de l'importance numérique des publications pertinentes.
3. Fouiller ou programmer à partir des termes anaphoriques (co-référence, désambiguïsation) afin d'augmenter la captation des publications qui ne citent pas nommément Cécile Helle, qu'elle qu'en soit la raison.
4. Explorer de manière complémentaire les possibilités de la recherche sur les réseaux sociaux, notamment Twitter et Facebook, où devaient pouvoir être saisie une information moins « institutionnelle », disons davantage « conversationnelle » et peut-être plus contrastée sur le travail politique du maire.
5. Poursuivre l'amélioration du design et de l'ergonomie de l'outil automatique.

13 Même rigoureusement tenu, l'agenda de planification des tâches n'est jamais fidèle à l'occupation réelle du temps de l'élue (Godmer and Marrel, 2014b).

D'une manière plus générale, l'enquête nécessiterait :

1. D'élargir les tests à l'emploi du temps d'un maire dont la « présence en ligne » est plus significative : de ce point de vue, l'enquête amorcée sur les posts d'Anne Hidalgo sur sa page Facebook peut orienter l'enquête comparative vers la mairie de Paris (Marrel and Rieffers, 2015).
2. De penser l'articulation complémentaire de cette recherche pluridisciplinaire en ligne avec les démarches hors-ligne déjà entreprises en science politique, que ce soit l'observation des réunions d'agenda des équipes politiques (Lefebvre, 2014), l'analyse de contenu quantitative et qualitative par entretiens de reconstitution d'emploi du temps et par observations directes en suivi d'élu (Godmer and Marrel, 2014b), ou encore l'étude ethnographique des usages du Web et du numérique comme instrument de communication ou de mobilisation politique au sein des équipes. L'étude proposée ici ne peut faire l'économie d'une analyse contextuelle et configurationnelle du *cyberespace* local, des usages du « web politique » des différents acteurs du territoire et des conditions des appropriations partisans différenciées de cette ressources de communication politique, à la manière de ce qu'a initié Nicolas Bué sur Calais au début des années 2000 par exemple (Bué, 2011).
3. D'interroger les limites éthiques de l'utilisation du dispositif informatique susceptibles d'émerger de cette recherche appliquée, visant à produire un prototype de reconstitution d'emploi du temps de personnalité publique. La question éthique se pose généralement dans le domaine des « méthodes du web » en termes de respect des traces de la vie privée d'une personnalité publique et d'obtention du consentement requis (ou non) à leur exploitation (Latzko-Toth and Proulx, 2013). L'outil implique ici un risque de transgression des fragiles frontières entre données publiques et données privées sur le Web, le robot étant à terme théoriquement capable d'objectiver la diffusion d'informations de nature confidentielle. En l'absence d'agenda « réel » de contrôle, il existe également un risque de reproduction par l'outil des rumeurs sur des interactions fictives inventées pour salir la réputation d'une personnalité publique. Par ailleurs, les acteurs politiques et professionnels intéressés par ce type d'instrument sont probablement nombreux, tant l'incertitude réputationnelle peut potentiellement augmenter pour eux. On peut alors s'interroger sur les débouchés de cette recherche en terme de « valorisation industrielle ». Scientifiquement, nous avons intérêt à suivre le développement de tels outils, voire à l'anticiper, y compris pour mieux comprendre les effets de rétroaction de la communication politique sur le contenu même de l'activité politique.

Références :

- Bailey, F. G. (1971) *Les règles du jeu politique: étude anthropologique*. (Copans, J., Trans.). Paris: Presses universitaires de France.
- Balandier, G. (2006) *Le pouvoir sur scènes*. Paris: Fayard.
- Blondeau-Coulet, O., and Allard, L. (2007) *Devenir média: l'activisme sur Internet, entre défection et expérimentation*. Paris, France: Éd. Amsterdam, impr. 2007.
- Bué, N. (2011) Le web partisan dans une ville moyenne. Une ressource d'usage limité. In Greffet, F. (Ed.), *Continuerlalutte.com*. Presses de Sciences Po.
- Cardon, D. (2010) *La démocratie Internet: promesses et limites*. Paris, France: Seuil, DL 2010.
- Chibois, J. (2014) Twitter et les relations de séduction entre députés et journalistes. La salle des Quatre Colonnes à l'ère des sociabilités numériques. *Réseaux* 188(6): 201–228.
- Coleman, S., and Blumler, J. G. (2009) *The Internet and democratic citizenship: theory, practice and policy*. Cambridge, Royaume-Uni, Etats-Unis.
- Demazière, D., and Le Lidec, P. (Eds.) (2014) *Les mondes du travail politique: les élus et leurs entourages*. Rennes: Presses universitaires de Rennes.
- Florini, A. (1998) The end of secrecy. *Foreign Policy* (111): 50–63.
- Frochot, D., and Molinaro, F. (2010) *E-réputation : suivre, soigner, défendre l'image de l'élu local sur le Net*. (Territoriales éditions., Vol. Numéro 230). Voiron.
- Gibson, R. K., and Ward, S. J. (2009) Parties in the Digital Age. A Review Article. *Representation* 45(1): 87–100.
- Godmer, L., and Marrel, G. (2014a) La production de l'agenda. Comment se fabrique l'emploi du temps d'une vice-présidente de conseil régional. In Demazière, D. and Le Lidec, P. (Eds.), *Les Mondes Du Travail Politique. Les Élus et Leurs Entourages* (PUR.). Rennes.
- Godmer, L., and Marrel, G. (2014b) Que font vraiment les professionnels de la politique ? L'agenda électronique et l'emploi du temps d'une élue régionale. In Mazeaud, A. (Ed.), *Pratiques de La Représentation Politique* (PUR.). Rennes.
- Granjon, F. (2001) *L'Internet militant: mouvement social et usages des réseaux télématiques*. Rennes, France: Éd. Apogée.

- Greffet, F. (Ed.) (2012) *Continuerlalutte.com*. Paris, France: Presses de Sciences Po.
- J Meijer, A., Curtin, D., and Hillebrandt, M. (2012) La gouvernance ouverte: relier visibilité et moyens d'expression. *Revue internationale des sciences administratives* 78(1): 13–32.
- Latzko-Toth, G., and Proulx, S. (2013) Enjeux éthiques de la recherche sur le Web. In Barats, C. (Ed.), *Manuel D'analyse Du Web En Sciences Humaines et Sociales*. Paris: Armand Colin.
- Lefebvre, R. (2014) Les élus comme entrepreneurs de temps : les agendas des cumulants. In Demazière, D. and Le Lidec, P. (Eds.), *Les Mondes Du Travail Politique. Les Élus et Leurs Entourages* (PUR.). Rennes.
- Marques, F. P. J. A., Aquino, J. A. de, and Miola, E. (2014) Congressmen in the age of social network sites: Brazilian representatives and Twitter use. *First Monday* 19(5).
- Marrel, G., and Payre, R. (2006) Temporalités électorales et temporalités décisionnelles. Du rapport au temps des élus à une sociologie des leaderships spatio-temporels. *Pôle Sud* 25(2): 71–88.
- Marrel, G., and Rieffers, A. (2015, February 7) *Les posts du maire de Paris en temps ordinaires. Des stratégies de visibilité du travail politique sur facebook ?*. Presented at the CoSpoF, ST 26 : Comment Internet change (ou pas) les règles du jeu politique Lausanne.
- Nicot, A.-L. (2012) Les sites internet des députés, terrains annexes de lutte partisane. In *Continuerlalutte.com* (Presses de Sciences Po.). Paris.
- Norton, P. (2007) Four Models of Political Representation: British MPs and the Use of ICT. *The Journal of Legislative Studies* 13(3): 354–369.
- Veitl, P. (2001) Une technique de gestion d'un espace-temps local : Maurice Pic, élu de la Drôme. In Girault, J. (Ed.), *L'implantation Du Socialisme En France Au XX^e Siècle. Partis, Réseaux, Mobilisation*. Paris: Publications de la Sorbonne.