

PROPOSITION SUJET DE THÈSE CONTRATS DOCTORAUX - 2021-2024

X Contrat doctoral fléché FR Agorantic

Directeur de thèse : MARREL Guillaume LBNC
Mail et Téléphone : guillaume.marrel@univ-avignon.fr 06 72 64 89 35

Co-directeur : LABATUT Vincent LIA
Co-encadrant : JOURLIN Pierre LIA

Titre en français : **Les élu·e·s dans Wikipédia**
Astroturfing et e-réputation entre peopolisation politique et contrôle participatif numérique du travail biographique

Titre en anglais : Elected representatives in Wikipedia : Astroturfing and e-reputation between political peopolisation and digital participatory control of biographical work

Résumé en 7 lignes : Le projet contribue à l'analyse de l'individualisation du champ politique et des usages politiques des médias numériques collaboratifs. Il examine la médiation numérique du travail de représentation biographique des élus plus ou moins professionnalisés en France, à travers les pages personnelles et politiques de l'encyclopédie libre et collaborative [Wikipédia](#). Il vise à identifier et analyser 1) les propriétés socio-politiques des élus du territoire qui disposent d'une biographie dans cette base de référence, ainsi que 2) les conditions particulières de production de ces récits, entre *astroturfing* controversé et contrôle contributif et 3) leurs contenus et leurs usages politiques et scientifiques en sociologie politique.

Mots clés: Personnel politique, Wikipédia, Biographies, e-réputation, Peopolisation, Traitement automatisé du langage naturel, Apprentissage profond

1 - Présentation du sujet

Le 25 juillet 2014, *The Guardian* titre : [US Congress banned from editing Wikipedia after staff caught trolling](#)¹. Quelques jours plus tard, le journaliste et développeur anglais Tom Scott crée une application open-source visant à suivre les modifications apportées à Wikipédia depuis le Parlement britannique afin de vérifier si les députés et les pairs s'adonnent à *l'astroturfing* (Kolivos and

¹ Le Congrès américain interdit de modifier Wikipédia après que son personnel a été pris en flagrant délit de trolling.

Kuperman, 2012; Laurens, 2015)² sur l'encyclopédie participative. Le 18 mai 2020, un twitoo français réplique l'outil et propose de révéler sur twitter "toutes les suppressions anonymes de contenu depuis les IPs de l'Assemblée nationale" sur le compte https://twitter.com/wiki_assemblee, après que "@Mediapart et @lesnums ont révélé que la députée LREM @LaetitiaAvia aurait vandalisé plusieurs fois Wikipédia à son avantage". Wikipédia n'est plus étrangère aux stratégies politiques de visibilité et d'arrangements hagiographiques des candidats, des élus et des leaders de la vie politique. La gestion d'e-réputation politique prend désormais grand soin des notices biographiques des acteurs politiques, presque autant que du storytelling sur les réseaux sociaux (Ratkiewicz et al., 2011). Mais des "vigies" citoyennes contrecarrent en retour ces pratiques de manipulations politiques.

Contexte et enjeux

De fait, Wikipédia est devenue la base de connaissances de référence qui sort presque systématiquement en tête du classement fourni par le moteur de recherche de Google (McMahon et al., 2017), principal outil d'orientation du Web. Elle est non seulement la première source des moteurs de recherche aujourd'hui, mais aussi l'une des principales bases biographiques de référence des personnalités publiques dont la notoriété justifie une notice individuelle et qui veillent désormais à leur e-réputation. Les notices biographiques individuelles constituent une part importante des entrées de l'encyclopédie. Elles participent du travail politique de mise en scène numérique de soi (Marrel and Labatut, 2018; Theviot, 2019).

Bénéficiant par définition d'une certaine popularité, le personnel politique occupe par hypothèse une place singulière dans Wikipédia, au côté des scientifiques, des sportifs et des artistes. La biographie s'est "démocratisée" à la fin de l'Ancien régime et au début du XIXe siècle, avec la multiplication des supports imprimés et l'essor des dictionnaires biographiques territoriaux, sectoriels, thématiques... Dans son "format papier", le "dictionnaire biographique" est en tant que tel un projet éditorial de nature politique dans les démocraties naissantes (Ferret, 2019) : il participe de l'identification et de la construction sociale des "nouvelles élites" de la modernité. Dans la démocratie représentative, faire campagne, être élu et réélu, c'est se faire connaître et reconnaître. L'apparition d'un récit biographique personnel témoigne de cette publicité et participe de la professionnalisation politique. Critique ou hagiographique, autobiographique ou confié à son entourage, le texte biographique et ses réécritures révèlent différents états de cette notoriété. Il est donc pertinent d'observer comment la biographie accompagne la carrière politique.

Or Wikipédia libère par hypothèse ce type de récit des contraintes éditoriales et des filtres académiques classiques, en amplifiant par hypothèse l'individualisation et la peopolisation du champ politique (Le Bart, 2013), en élargissant à la fois le spectre des personnalités susceptibles d'accéder à la postérité biographique, mais aussi en permettant un contrôle collectif et participatif du travail

² Selon Wikipédia, l'astroturfing ou la désinformation populaire planifiée ou orchestrée (traduit par similitantisme au Québec) désigne des techniques de propagande manuelles ou algorithmiques utilisées à des fins publicitaires ou politiques ou encore dans les campagnes de relations publiques, qui ont pour but de donner une fausse impression d'un comportement spontané ou d'une opinion populaire sur Internet. Cette tentative de manipulation fait référence à la pelouse artificielle de marque AstroTurf (en) utilisée dans les stades. Elle consiste à simuler un mouvement citoyen, venu de la base (« grassroots movement »).

biographique dans le cadre du projet éditorial collaboratif de l'encyclopédie. Le format numérique et ouvert de Wikipédia permet par ailleurs une analyse systématique de contenu des formes et transformations de la présence nominative des élus de France dans les notices de l'encyclopédie, depuis sa création en 2001. La plateforme archive également l'intégralité des enregistrements réalisés sur ses notices et autorise donc un suivi généalogique des controverses et des modifications auxquelles l'écriture biographique donne lieu. Cette pratique sensible fait d'ailleurs l'objet de plusieurs recommandations générales relatives aux [conventions de style](#), harmonisées dans le "[Projet biographie](#)" de Wikipédia, mais aussi aux enjeux spécifiques des "[biographies de personnes vivantes](#)" et de "[l'autobiographie](#)". L'enquête intègre également le projet "[Mettons nos élu·e·s sur Wikidata !](#)".

Objectifs

Le projet de thèse suggère d'explorer trois séries de questions :

1. **Le profil des élus identifiés dans Wikipédia** : Qui sont les élus qui disposent d'une notice Wikipédia, et qui sont ceux qui n'en disposent pas ? Quelle proportion des élus cela représente-t-il ? Quelles sont leurs propriétés sociales et politiques ? Peut-on repérer des variables socio-politiques statistiquement significatives dans la distribution de cette population ? Quelles positions occupaient-ils au moment de la création de leurs notices ?
2. **Le contenu des notices politiques sur Wikipédia** : Quelles sont les informations biographiques diffusées sur l'encyclopédie ? Peut-on identifier des standards dans les contenus ? À quel rythme les notices évoluent-elles et comment cette évolution peut-elle être liée ou non aux événements de la carrière de l'individu ? Ces élus qui disposent d'une notice sont-ils cités dans d'autres notices biographiques ou non biographiques ? Si oui lesquelles ? À quelle fréquence ? Selon quelle "tonalité" ? Les élus qui ne disposent pas de notice biographique sont-ils cités dans d'autres notices biographiques ou non ? Si oui lesquelles ? À quelle fréquence ? Comment extraire et structurer ces informations publiées en langage naturel dans des bases de données socio-biographiques exploitables en sociologie du personnel politique ? Comment représenter et cartographier ces informations ?
3. **L'investissement de Wikipédia par les entreprises politiques et les communautés de contributeurs** : Le projet contribue également à mieux connaître les usages de Wikipédia. Le formidable développement de cette encyclopédie fait l'objet depuis le début des années 2010 d'un réel intérêt scientifique en sciences de l'information principalement (Barbe, Merzeau, et Schafer 2015). La question de l'exploitation de données biographiques à partir de Wikipédia n'est pas nouvelle (Viseur 2015). Comment les corpus de biographies politiques se développent-ils dans la courte histoire de Wikipédia ? À quel moment, depuis 2001, Wikipédia devient-elle un espace de publication de biographies de personnalités politiques vivantes pertinent et recherché ? Par qui et pourquoi ? Qui sont les auteurs et les correcteurs des biographies des femmes et des hommes politiques ? Comment l'écriture biographique donne-t-elle parfois lieu à controverses ? Comment le travail hagiographique des entreprises

de marketing politique est-il contrecarré par des communautés de contributeurs où se croisent militants, opposants, citoyens ordinaires et profanes ? Comment caractériser et décrire ces controverses ? L'action des "communautés" de contributeurs permet-elle de produire une image "équilibrée" des personnalités politiques controversées ?

Méthodologie

Pour exploiter l'immensité de son contenu, nous proposons, pour le terrain français, de croiser Wikipédia, avec la Base Révisée des Élus de France ([BRÉF](#)), qui a été élaborée dans le cadre d'Agorantic et compile les données socio-biographiques élémentaires du Répertoire National des Élus (RNE) du ministère de l'Intérieur ainsi que d'autres sources. L'étude vise donc d'abord à recenser, comparer et interpréter les modalités d'apparition des patronymes des élus de France dans l'encyclopédie (après résolution des problèmes d'homonymie), par l'exploitation des métadonnées contenues dans les Infobox, des données semi-structurées (balises wiki) ou de données extraites automatiquement des narrations, grâce aux algorithmes de TALN (Traitement Automatisé du Langage Naturel).

Dans l'optique d'extraire automatiquement du texte plein, les entités nommées (séquences de mots faisant référence à des personnes, des lieux, des institutions, des dates, etc.) et les associations existant entre elles (nom de l'élu, date de naissance, lieu de naissance, mandat, date de début de mandat...), nous mettrons en application les techniques dites "prédictives" basées sur des architectures complexes de réseaux de neurones artificiels profonds, utilisant l'apprentissage par transfert et qui constituent aujourd'hui l'état-de-l'art de la discipline (Le et al., 2020). Nous étudierons également les performances comparées d'algorithmes fondés sur des grammaires formelles locales, tant du point de vue de la précision, que du coût énergétique pour effectuer les mêmes tâches.

2 - Profil du candidat :

La personne candidate devra disposer d'une solide formation de sciences sociales, en sociologie politique ou science de l'information et de la communication. Si une formation aux méthodes quantitatives constitue un atout, celle-ci n'est pas un prérequis tant que la personne fait montre d'un intérêt pour l'évolution des outils numériques de mesure et d'objectivation statistique ainsi que de dispositions au travail pluridisciplinaire. Elle sera invitée à participer activement aux travaux de la fédération de recherche Agorantic (FR 3621).

3 - Opportunités de mobilité à l'international du doctorant(e) : oui

Le projet défini pour le cas des élus de France dans le Wikipédia en langue française pourra avantageusement intégrer une dimension comparative, et donc une mobilité, avec le Québec francophone d'une part, mais aussi avec une aire culturelle anglophone, où les technologies du TALN sont par ailleurs davantage développées que pour le Français. Une comparaison avec des régimes politiques non pluralistes ou autoritaires pourrait également montrer de l'intérêt.

4 - Références bibliographiques :

- Ferret, O. (2019) Présentation générale. In Mercier-Faivre, A.-M. (Ed.), *Biographie & Politique : Vie Publique, Vie Privée, de l'Ancien Régime à La Restauration*. Lyon: Presses universitaires de Lyon.
- Kolivos, E., and Kuperman, A. (2012) Consumer law: Web of lies - legal implications of astroturfing. *Keeping Good Companies*.
- Laurens, S. (2015) Astroturfs et ONG de consommateurs téléguidées à Bruxelles. Quand le business se crée une légitimité « par en bas ». *Critique internationale* 67(2): 83–99.
- Le Bart, C. (2013) *L'ego-politique: essai sur l'individualisation du champ politique*. Paris: A. Colin.
- Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., Crabbe, B., Besacier, L., and Schwab, D. (2020) FlauBERT: Unsupervised Language Model Pre-training for French. In *LREC*. Marseille, France.
- Marrel, G., and Labatut, V. (2018) La visibilité politique en ligne de la maire de Paris. Contribution à la mesure de l'écho Web-médiatique d'Anne Hidalgo. In Alcantara, C., Charest, F., and Agostinelli, S. (Eds.), *Big Data et Visibilité En Ligne. Un Enjeu Pluridisciplinaire de l'économie Numérique* (Presses des Mines.). Paris.
- McMahon, C., Johnson, I., and Hecht, B. (2017) The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies. *Proceedings of the International AAAI Conference on Web and Social Media* 11(1).
- Ratkiewicz, J., Conover, M., Meiss, M., Goncalves, B., Flammini, A., and Menczer, F. (2011) Detecting and Tracking Political Abuse in Social Media. *Proceedings of the International AAAI Conference on Web and Social Media* 5(1).
- Theviot, A. (2019) Faire campagne sur Youtube : une nouvelle « grammaire » pour contrôler sa communication et influencer sur le cadrage médiatique ? *Politiques de communication* N° 13(2): 67–96.