

"Les élus dans Wikipédia : l'e-réputation politique entre promotion personnelle et contrôle participatif numérique du travail biographique"

Victor Planche – membre du LBNC (futur JPEG) et de la FR Agorantic
Sous la direction de Guillaume Marrel (LBNC) et de Pierre Jourlin (LIA)

À travers cette thèse je compte interroger la neutralité du contenu encyclopédique de Wikipédia, pensée comme acquise grâce aux règles strictes qui encadrent l'écriture sur la plateforme. Alors que de multiples enquêtes journalistiques mettent en lumière les tentatives de caviardage mises en œuvre par les membres du personnel politique sur leur propre notice biographique dans Wikipédia, il s'agira de voir si le contenu de l'encyclopédie se politise en interrogeant la manière dont elle peut servir ou non d'espace de promotion ou de disqualification du personnel politique. Je questionnerai ainsi la fabrique de l'image publique des candidats dans Wikipédia en étudiant les modalités de rédaction de leur notice biographique mais aussi leur contenu.

L'étude du travail biographique constitue donc l'objet de cette thèse. Mon étude vise à réactualiser une problématique ancienne supposant que la neutralité des récits sur le personnel politique est nécessairement ambiguë. Il est pensé que les politiques maîtrisent la manière dont ils apparaissent auprès du public à travers un travail de représentation constant dont témoignent leurs différents actes qui visent à façonner une identité séduisante pour l'électorat. De ce fait, ils auraient la capacité d'imposer leur propre récit à leur biographe. Pourtant, d'un autre côté, les journalistes qui écrivent ces récits essaient d'adopter une posture distante vis-à-vis du jeu politique pour éviter de représenter les acteurs politiques tels qu'ils le souhaiteraient.

Wikipédia apparaît comme un espace qui permet de réinterroger cette tension entre récit à prétention objective porté par des contributeurs dont l'ethos les rapproche des journalistes, et récit promotionnel porté par des candidats ou des militants. En effet, du fait de son aspect collaboratif, Wikipédia permet à ces deux catégories d'agents de contribuer en même temps aux mêmes notices biographiques. Ma problématique de thèse m'amène ainsi à m'interroger sur la manière dont l'image projetée des candidats dans Wikipédia devient un enjeu de lutte entre plusieurs acteurs qui cherchent à la manipuler promotionnellement ou à but informatif.

Pour répondre à cette problématique je m'appuierai sur un corpus composé des fiches Wikipédia des candidats aux élections législatives entre 2002 et 2022. La première année qui délimite mon corpus constitue l'année de création de la plateforme collaborative. Ce corpus permettra de raconter l'histoire de l'investissement électoraliste de la plateforme encyclopédique et les stratégies pour le réguler.

Grâce aux archives numériques (*dumps*) de Wikipédia disponibles en libre accès, il s'agira de détailler les multiples caractéristiques de ces notices biographiques telles que leur date de création ou encore le nombre de modifications qu'elles reçoivent. Il s'agira également d'analyser le contenu textuel de ces notices biographiques grâce à des méthodes informatiques de traitement automatique du langage naturel (TALN). Nous collecterons dans un premier temps automatiquement les propos que les wikipédiens signalent comme non-neutres. Dans un second temps, nous nous appuierons sur un algorithme de type *Large Language Model* (LLM) pré-entraîné sur un large corpus de texte, qui sera capable de prédire quand des formes de propos peuvent être considérées comme non-neutres au sens des wikipédiens. Cette analyse permettra alors de repérer les pages biographiques sur lesquelles se maintient un contenu qui est pensé comme promotionnel ou disqualifiant par les wikipédiens. Pour le reste, il faut préciser que

cette méthode est essentiellement inductive tant il paraît difficile de prédire les résultats qu'elle va fournir. L'analyse se fera donc principalement *ex post*.

Nous articulerons cette méthode d'analyse quantitative avec une approche qualitative. Il s'agira de restituer l'entièreté du cheminement rédactionnel dans un système d'écriture collaboratif. Du fait de la transparence totale de ses données, l'étude de Wikipédia permet l'élaboration d'une approche innovante pour étudier le travail biographique. Nous pourrions déplier l'entièreté de ce travail en restituant les propriétés des acteurs, les propos qu'ils rédigent et les discussions qui les amènent à encadrer le format d'écriture. Je mènerai ainsi des entretiens avec les candidats pour comprendre s'ils perçoivent leur fiche biographique comme un outil de communication. J'interrogerai également les wikipédiens sur leur manière de contrôler l'écriture biographique dans Wikipédia. Il s'agira en outre de repérer les controverses qui naissent dans les discussions entre contributeurs concernant les notices biographiques. Je chercherai alors à comprendre comment les contributeurs règlent ces controverses et invoquent des arguments pour indiquer comment doivent être écrites les pages des candidats. Enfin, en m'appuyant sur une observation fine du web, je mènerai une analyse de l'entièreté du processus de rédaction de plusieurs notices biographiques.

Bibliographie indicative :

AGARWAL Pushkal, REDI Miriam, SASTRY Nishanth, WOOD Edward et BLICK Andrew, 2020, « Wikipedia and Westminster: Quality and Dynamics of Wikipedia Pages about UK Politicians », New York, NY, USA, Association for Computing Machinery.

ALDRIN Philippe et HUBÉ Nicolas, 2022, *Introduction à la communication politique*, Louvain-la-Neuve, De Boeck Supérieur.

BOULLIER Dominique, 2015, « Les sciences sociales face aux traces du big data. Société, opinion ou vibrations ? », *Revue française de science politique*, vol. 65, n° 5-6, p. 805-828.

BOURDIEU Pierre, 1986, « L'illusion biographique », *Actes de la Recherche en Sciences Sociales*, vol. 62, n° 1, p. 69-72.

BOYADJIAN Julien, OLIVESI Aurélie et VELCIN Julien, 2017, « Le web politique au prisme de la science des données. Des croisements disciplinaires aux renouvellements épistémologiques », *Réseaux*, vol. 204, n° 4, p. 9-31.

BROWN Adam R., 2011, « Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage », *PS: Political Science & Politics*, vol. 44, n° 2, p. 339-343.

COLLOVALD Annie, 1999, *Jacques Chirac et le gaullisme: biographie d'un héritier à histoires*, Paris, Belin.

COLLOVALD Annie, 1988, « Identité(s) stratégique(s) », *Actes de la Recherche en Sciences Sociales*, vol. 73, n° 1, p. 29-40.

DAMAMME Dominique, 1994, « Grandes illusions et récits de vie », *Politix. Revue des sciences sociales du politique*, vol. 7, n° 27, p. 183-188.

ENLI Gunn Sara et SKOGERBØ Eli, 2013, « Personalized Campaigns in Party-Centred Politics », *Information, Communication & Society*, vol. 16, n° 5, p. 757-774.

FARKAS Johan et NEUMAYER Christina, 2020, « Disguised Propaganda from Digital to Social Media » dans Jeremy Hunsinger, Matthew M. Allen et Lisbeth Klastrup (eds.), *Second International Handbook of Internet Research*, Dordrecht, Springer Netherlands, p. 707-723.

GÖBEL Sascha et MUNZERT Simon, 2018, « Political Advertising on the Wikipedia Marketplace of Information », *Social Science Computer Review*, vol. 36, n° 2, p. 157-175.

GODMER Laurent et MARREL Guillaume, 2016, *La politique au quotidien: L'agenda et l'emploi du temps d'une femme politique*, Paris, ENS Éditions.

GODMER Laurent Robert, 2021, *Le travail électoral: ethnographie d'une campagne à Paris*, Paris, l'Harmattan.

GREENSTEIN Shane et ZHU Feng, 2012, « Is Wikipedia Biased? », *American Economic Review*, vol. 112, n° 3, p. 343-348.

GREFFET Fabienne, 2011, « Le web, espace de luttes partisans » dans *Continuerlalutte.com*, Paris, Presses de Sciences Po.

KACIAF Nicolas, 2013, *Les pages « Politique »: histoire du journalisme politique dans la presse française, 1945-2006*, Rennes, Presses universitaires de Rennes.

LE BART Christian, 2013, *L'ego-politique: essai sur l'individualisation du champ politique*, Paris, Armand Colin.

LE BART Christian, 2012, *La politique en librairie: les stratégies de publication des professionnels de la politique*, Paris, Armand Colin.

LEFÉBURE Pierre, 2020, « «C'est totalement ridicule»: stratégie et contenu d'une communication négative dissimulée lors de la campagne présidentielle française de 2017 », *Questions de communication*, n° 38, p. 145-172.

LEGAVRE Jean-Baptiste, 2011, « Entre conflit et coopération. Les journalistes et les communicants comme "associés-rivaux" », *Communication & langages*, vol. 169, n° 3, p. 105-123.

MARREL Guillaume, 2002, « Sociologie historique des carrières de cumul. Une expérience propographique » dans *Faire de la science politique. Pour une analyse socio-historique du politique*, Paris, Belin.

NEIHOUSER Marie, 2020, « Les collaborateurs parlementaires en charge de la communication numérique des députés », *Les Enjeux de l'information et de la communication*, vol. 21/2, n° 2, p. 51-65.

NEVEU Érik, 1992, « Le sceptre, les masques et la plume », *Mots. Les langages du politique*, vol. 32, n° 1, p. 7-27.

OLLIVIER-YANIV Caroline, 2010, « Discours politiques, propagande, communication, manipulation », *Mots. Les langages du politique*, n° 94, p. 31-37.

PANT Kartikey, DADU Tanvi et MAMIDI Radhika, 2020, « Towards Detection of Subjective Bias using Contextualized Word Embeddings », Taipei Taiwan, ACM.

THEVIOT Anaïs, 2018, *Faire campagne sur Internet*, Villeneuve d'Ascq, Presses universitaires du Septentrion.