

**APPEL A PROJETS RECHERCHE 2021**

**FÉDÉRATION DE RECHERCHE AGORANTIC  
«CULTURE, PATRIMOINES, SOCIÉTÉS NUMÉRIQUES »**

<b>Titre</b>	<b>Les usages politiques de Wikipédia</b>
<b>Acronyme</b>	<b>WikiPOL</b>
<b>Nom du/des porteur(s)</b>	Guillaume Marrel
<b>Laboratoires associés</b>	LBNC : Noémie Févrat, Christèle Lagier, Guillaume Marrel, Jeanne Vermeirsche LIA : Malek Hajjem, Tania Jiménez, Pierre Jourlin, Vincent Labatut, Eric Sanjuan
<b>Budget demandé</b>	<b>7 381 €</b>
<b>Résumé</b> <b>Max. 1 000 caractères espaces compris</b>	WikiPol analyse la présence des élus et des discours politiques dans le Wikipédia de l'espace francophone. Le projet explore les usages politiques et la neutralité de Wikipédia, à partir de l'analyse du croisement de plusieurs corpus : une Base de données des élus de France avec les notices biographiques du personnel politique enregistrées dans l'encyclopédie, d'une part, et un jeu d'articles et de communiqués de presse liés aux discours nationalistes et populistes en France et au Québec, avec les pages francophones relatives aux discours et aux idées politiques, d'autre part. Le projet associe politistes et informaticiens dans l'expérimentation d'outils innovants de traitement du langage naturel.

**1. Contexte, positionnement, objectif(s)/problématique et cohérence par rapport à la thématique « Culture, Patrimoines, Sociétés Numériques »**

Le projet WikiPol rassemble des enjeux de recherche et des chercheurs de sociologie politique et d'informatique dans le cadre de l'axe Culture, Patrimoines, Sociétés numériques, autour des différents types de pages politiques de l'encyclopédie numérique contributive Wikipédia. La place acquise par Wikipédia dans le paysage médiatique 2.0 est incontestable, tout comme son rôle dans la diffusion des savoirs, de l'information et des représentations. Le rôle de ses contributeurs et modérateurs est de grande importance, ainsi que leur impartialité. La recherche pluridisciplinaire propose ici d'éclairer les usages politiques de cette encyclopédie participative et ouverte, dont le formidable développement fait l'objet depuis le début des années 2010 d'un réel intérêt scientifique (Barbe, Merzeau, et Schafer 2015).

Cet espace de référence désormais central dans les retours des moteurs de recherche fait l'objet d'un réel investissement de la part des acteurs politiques, qui cherchent d'une part à y être référencés afin d'améliorer leur notoriété et, d'autre part, à diffuser des idées, des représentations et des discours à des fins partisans. L'objectif scientifique du projet est d'expérimenter différents outils du TALN sur la base des connaissances de référence de Wikipédia, dans le champ politique. WikiPol mobilise ainsi deux opérations de

recherche et un corpus de données récemment développées dans le cadre de la FR Agorantic :

- le projet WikiRNE, financé en 2020, qui se base sur l'utilisation du web sémantique, en particulier de Wikidata afin d'identifier les élus du Répertoire National des Élus (RNE) du Ministère de l'Intérieur qui disposent d'une notice biographique dans Wikipédia ;
- la Base Révisée des Élus de France (BRÉF) hébergée par la fédération qui consiste, à partir du RNE et de divers autres sources compilées, en un corpus de données socio-biographiques et politiques sur l'ensemble du personnel politique, élus locaux et nationaux ;
- et une dimension du projet de recherche doctoral de Jeanne Vermeirsche (FR Agorantic-LBNC) qui propose d'interroger l'existence d'un Wikipédia politique. Il s'agit d'y repérer la porosité entre discours politiques et médiatiques autour des thèmes du nationalisme et du populisme, dans une comparaison France-Québec.

### Les élus dans Wikipédia

Le projet WikiRNE a permis en 2020 d'initier l'examen de la présence des élus de France dans la base de connaissances universelles de référence que constitue Wikipédia. L'enquête concernait les mutations numériques de l'accompagnement biographique des carrières politiques. L'objectif était de considérer Wikipédia comme l'un des "nouveaux" espaces de référence permettant de construire, et donc mesurer, la notoriété d'une personnalité publique, dans la continuité des dictionnaires biographiques compilant les récits synthétiques des trajectoires exemplaires des élites culturelles, politiques et économiques depuis le début du XIXe siècle. Il s'agissait de rapprocher (matching) la liste de tous les élus de France des données sources (RNE) avec celles de Wikidata. Pour ce faire, nous avons utilisé les données nominatives des infobox structurées enregistrées dans la base wikidata, de manière à vérifier et à qualifier la présence en ligne de ces personnalités publiques à la notoriété plus ou moins étendue selon le type d'élection et de mandat. L'objectif est ici de poursuivre cette recherche à partir de la Base Révisée des Elu•e•s de France (BRÉF). Cette base des données sur les personnes ayant occupé au moins un mandat sous la Ve République en France a été initiée dans le cadre du travail doctoral de Noémie Févrat (FR Agorantic-LBNC). La BRÉF est destinée à être accessible en open science à terme. Elle est construite à partir de la source principale du Répertoire National des Elus (RNE), complétée des bases de données de l'Assemblée Nationale, du Sénat et du Parlement européen, des archives nationales et du Ministère de l'Intérieur (bureau des élections), ainsi que de sources issues de l'espace médiatique (<https://www.politiquemania.com>). Cette base a vocation à être étendue ultérieurement, en exploitant plus complètement ces sources secondaires, et à plus long terme en intégrant de nouvelles bases de données ainsi que des données non-structurées (texte en langage naturel) dont l'exploitation pourrait être facilitée par le Web-sémantique.

L'objectif est d'abord de consolider l'analyse des personnels politiques disposant d'une notice en ligne, en tenant compte du projet "[Mettons nos élu•e•s sur Wikidata](#) !", qui vise à rapprocher le RNE et Wikidata depuis l'ouverture Opendata des données du RNE en janvier 2019. Mais il s'agit aussi d'explorer les parties rédigées en langage naturel et donc non structurées des données de Wikipédia relatives au personnel politique français. En mobilisant les outils du TALN, la finalité est d'exploiter les biographies des personnalités

politiques présentes dans l'encyclopédie, puis de fouiller l'ensemble des autres textes liés au champ politique, de façon à récupérer des informations de nature à compléter la base de données socio-biographiques et politiques de la BRÉF. L'enquête vise également l'exploration des dynamiques de production des récits biographiques des acteurs politiques dans l'espace d'écriture collaborative de Wikipédia et l'examen des stratégies individuelles et collectives de nature hagiographique, déjouant et contournant les instruments d'auto-contrôle et de neutralisation des contributions anonymes de Wikipédia, destinées "à faire référence".

### **Les discours populistes dans Wikipédia**

La recherche porte ici sur la confrontation entre l'impératif de neutralité de Wikipédia et sa forte réactivité aux discours politiques et médiatiques. Wikipédia ouvre un espace de concurrence au langage politique légitime, voire de contre-pouvoir dans l'espace public numérique (Cardon, 2019). Au vu de l'écrasante fréquentation de ses pages, Wikipédia constitue à l'évidence un espace de médiation que l'on peut aujourd'hui difficilement ignorer pour comprendre la construction du rapport des citoyen.ne.s ordinaires au politique. Il reste pourtant paradoxalement assez peu étudié en science politique.

Ce projet propose d'interroger les recouvrements entre discours politiques, discours médiatiques et Wikipédia à l'heure du développement des formations populistes. La comparaison France/Québec nous semble exemplaire des métamorphoses sémantiques opérées par ces formations sur la thématique du nationalisme. Très actives dans l'espace numérique, les formations populistes irriguent en effet les termes du débat médiatique et politique en valorisant, imposant, détournant ou redéfinissant le "nationalisme". Quels échos en trouve-t-on alors dans Wikipédia dont la neutralité est revendiquée ? Projet sans cesse en mouvement, qui s'améliore au fil du temps (Ertzscheid, 2008), Wikipédia s'est progressivement doté de modes de fonctionnement propres relevant d'une logique épistémique exigeante (Cardon, Levrel, 2009). Les normes relatives à la pluralité et la diversité des contributions et points de vue garantiraient l'objectivation des contenus édités (Barbe, Merzeau, Schafer, 2015). Nous proposons ici de mettre à l'épreuve cette «neutralité de point de vue» en projetant les textualités des discours médiatiques et politiques recueillis au sein de deux corpus constitués automatiquement. En cherchant à saisir les emprunts et démarcations qui peuvent s'observer entre Wikipédia et ces derniers, nous proposons de mesurer en quoi Wikipédia est sensible ou non aux manières dont s'écrit l'actualité politique et médiatique en ligne.

## **2. Méthodologie et résultats attendus**

Sur le premier axe de recherche, il s'agit d'abord de croiser la base nominative BRÉF avec Wikipédia pour extraire, lorsqu'elles sont disponibles, toutes les informations biographiques concernant les élus. Un.e élu.e peut disposer d'une notice biographique personnelle. Son nom peut être mentionné dans le cadre d'une autre notice biographique, thématique ou institutionnelle. Dans certains cas, ces informations sont déjà structurées sous la forme d'infobox. Certaines sont même déjà extraites sous forme de triplets RDF

(Resource Description Framework, voir Web Sémantique et dbpedia), donc facilement exploitables sous forme relationnelle. Pour les autres informations qui ne sont pas du tout structurées, il faut recourir à des techniques de traitement automatique du langage naturel. Des outils d'annotation syntaxique automatique, de désambiguïsation et d'extraction d'entités nommées sont utilisés au LIA et seront mis à profit pour réaliser cette tâche. Plus généralement, la recherche prolonge l'exploitation des données biographiques à partir de Wikipédia (Viseur 2015). Comment les corpus de biographies politiques se développent-ils dans la courte histoire de Wikipédia ? Quand Wikipédia devient-il 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 notices ? A quels types de controverses l'écriture biographique politique donne-t-elle lieu ? Peut-on caractériser et décrire ces controverses ? Que nous apprennent ces wiki-controverses sur l'écriture biographique politique et la gestion de l'e-réputation des personnalités publiques aujourd'hui ?

Concernant le second axe de recherche, nous proposons une approche comparative des discours médiatiques, politiques et profanes qui utilisent Wikipédia comme référentiel pour la recherche d'information. La méthode de contextualisation permet de mesurer objectivement la distance relative entre un discours et un ensemble de concepts. Nous appliquons cette méthodologie au cas des usages du nationalisme dans les discours médiatiques et politiques en France et au Québec. Deux corpus sont automatiquement constitués : le premier rassemble des articles archivés d'une sélection de médias français et québécois, à la fois « mainstream » et plus alternatifs, disposant d'une interface web ; le second est nourri des discours politiques des grands partis (communiqués de presse politiques et données parlementaires) dans les deux espaces nationaux. Ces deux corpus seront projetés dans l'espace participatif du Wikipédia à partir d'une extraction (dump) du Wikipédia français à différentes temporalités. Pour ce faire nous utilisons des outils de la recherche d'information, tels que le LDA (Deveaud et al. 2014). Ce travail pourrait révéler des références implicites susceptibles de différencier mais aussi de rapprocher les discours. En comparant les projections sur plusieurs versions du Wikipédia, ce projet vise à répondre à plusieurs questions sur la matérialisation d'un Wikipédia politique : Peut-on objectiver la porosité entre discours politiques, discours médiatiques en ligne et les pages Wikipédia dans la diffusion des représentations politiques actuelles ? Wikipédia reste-t-il neutre ? Ne fait-il que refléter les débats politiques ou adopte-t-il une position particulière sur certaines thématiques ?

### **3. Dimension interdisciplinaire (champs disciplinaires associés)**

Les différents projets de recherche fédérés ici autour de Wikipédia sont tous de nature pluridisciplinaire et associent au LBNC et au LIA, science politique, sciences du langage et traitement du langage naturel en informatique. Le projet rassemble plusieurs équipes de recherche autour d'enjeux croisant d'une part la question de l'objectivation des investissements politiques dont fait l'objet Wikipédia et des inévitables variations de sa neutralité en science politique et, d'autre part, les défis de la recherche d'information, de la contextualisation et de la désambiguïsation, dans le champs du TALN en informatique.

#### 4. Partenariats extérieurs envisagés

Le projet WikiPol vise une double valorisation, en termes de communications et de publications dans les conférences et les revues disciplinaires et interdisciplinaires concernées, et en termes de diffusion scientifique lors des événements grand-publics.

Le dispositif fédératif qu'il s'agit ici de financer entre le LBNC et le LIA à vocation en cas de réussite à mobiliser d'autres acteurs académiques et non-académiques, autour des usages politiques de Wikipédia, au plan national et international.

Budget (€)*		
	Brève description	Montant
<b>Stages**</b>	M2 informatique 6 mois	3 480
<b>Post-doc</b>	IGE informatique 1.5 mois ETP (INM388)	3 901
<b>Budget total</b>		7 381
<b>Budget demandé à Agorantic</b>		<b>7 381</b>

\* Veuillez modifier les catégories de dépenses si besoin – ajoutez/supprimez des lignes à votre convenance

\*\* Gratification de stage obligatoire au-delà de 2 mois – prévoir environ 580€ par mois