



HAL
open science

Pour une cartographie romanesque de Paris au XIXe siècle. Proposition méthodologique

Noémie Boeglin, Michel Depeyre, Thierry Joliveau, Yves-François Le Lay

► **To cite this version:**

Noémie Boeglin, Michel Depeyre, Thierry Joliveau, Yves-François Le Lay. Pour une cartographie romanesque de Paris au XIXe siècle. Proposition méthodologique. Conférence Spatial Analysis and GEomatics, Dec 2016, Nice, France. hal-01619600

HAL Id: hal-01619600

<https://hal.science/hal-01619600>

Submitted on 10 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pour une cartographie romanesque de Paris au XIXe siècle. Proposition méthodologique

Noémie Boeglin¹, Michel Depeyre¹, Thierry Joliveau¹, Yves-François Le Lay²

1. Laboratoire Environnement-Ville-Société, Université Jean Monnet de Saint-Etienne

6 rue Basse-des-Rives, F42000 Saint-Etienne, France

noemie.boeglin@univ-st-etienne.fr, michel.depeyre@univ-st-etienne.fr,

thierry.joliveau@univ-st-etienne.fr

2. Laboratoire Environnement-Ville-Société, Ecole Normale Supérieure de Lyon

15 parvis René Descartes, F69007 Lyon, France

yves-francois.le-lay@ens-lyon.fr

RESUME. L'article propose une méthode pour localiser, cartographier et analyser les occurrences de lieux citées dans un corpus de 31 romans du XIXe siècle dont l'action se situe pour tout ou partie à Paris. L'objectif à terme plus lointain est de construire une plateforme facilitant l'analyse cartographique d'autres textes évoquant Paris à la même époque. La méthode expérimentée associe l'usage de la plateforme d'analyse textuelle open-source TXM et l'élaboration d'un Système d'Information Géographique adapté, supportant un atlas toponymique des lieux parisiens entre 1830 et 1914 et fournissant une interface de géovisualisation. Les quatre étapes successives sont décrites : rassembler et corriger les versions numériques des textes des romans, extraire les Entités Spatiales Nommées par l'écriture de requêtes adaptées dans la plateforme textométrique TXM, localiser ces entités par la mobilisation de sources historiques en constituant un gazetteer adapté et proposer des formes de cartographie adaptées aux spécificités des espaces littéraires. Une discussion et des perspectives concluent le papier.

ABSTRACT. We propose a method to locate, map and analyze the occurrences of places cited in a corpus of 31 novels of the nineteenth century whose action takes place wholly or partly in Paris. The more distant term goal is to build a platform facilitating the cartographic analysis of other texts evoking Paris at the same time. The experienced method combines the use of a textual analysis platform (TXM) and the development of a convenient Geographic Information System supporting a gazetteer of Parisian places between 1830 and 1914 and an interface of geovisualization. The four stages are described: (a) collect and correct digital versions of novels, (b) extract spatial named entities by writing suitable queries in the TXM lexicometric platform, (c) locate these entities by using historical sources and constituting an appropriate gazetteer, and (d) offer mapping forms adapted to the specificities of literary spaces. Discussion and perspectives conclude the article.

MOTS-CLES : RIG – GEOPARSING – CARTOGRAPHIE – ANALYSE TEXTUELLE – TEXTOMETRIE – HISTOIRE URBAINE – PARIS – LITTÉRATURE – SIG – XIXe SIECLE

KEYWORDS: GEOPARSING – CARTOGRAPHY – GIS – TEXTOMETRY – LITERATURE – PARIS – URBAN HISTORY – 19TH CENTURY

1. La question posée

Dans le cadre de sa thèse d'histoire sur les paysages parisiens évoqués dans 31 romans du XIXe siècle, Noémie Boeglin effectue un inventaire cartographique des lieux nommés dans son corpus. Cet inventaire manuel s'avère fastidieux et approximatif. La réalisation de cartes pose aussi des problèmes complexes tant que l'on reste dans une approche manuelle, même informatisée, en particulier si l'on veut procéder à des visualisations nombreuses par roman, par auteur ou selon une logique temporelle. L'équipe de recherche s'est constituée afin de : 1) proposer une méthodologie rigoureuse permettant de répondre de manière opérationnelle à ce problème précis ; 2) réfléchir à la généralisation possible de cette méthode et envisager l'extraction et la cartographie automatique des noms de lieu dans les textes de fiction.

2. Etat de l'art et problématique

Le rapprochement entre la géographie et les disciplines littéraires et l'emprunt de concepts et de techniques cartographiques pour l'étude des espaces romanesques est déjà ancien (Moretti, 2000) ; (Moretti, 2008) ; (Bulson, 2007). F. Moretti en particulier a proposé une approche de « lecture distante » des textes utilisant des méthodes statistiques empruntées aux sciences humaines. Il propose ainsi une « géographie de la littérature » en replaçant par exemple dans la topographie concrète des lieux les personnages des romans du XIXe dont l'intrigue se passe à Londres ou à Paris, réactualisant sur une base systématique des propositions anciennes. Si le mode de production de F. Moretti était manuel, les méthodes informatiques de recherche d'information géographique (RIG) dans les textes, relevant plus ou moins de ce que l'on appelle en anglais *geoparsing*, se sont largement développées pour tenter d'extraire le plus automatiquement possible ces références textuelles aux lieux. Le développement du domaine des humanités numériques a contribué à accroître fortement l'intérêt pour ces approches en littérature et en histoire, comme l'illustrent par exemple les travaux de (Gregory et al., 2015) ou la publication récente du texte de référence de (Cooper, Donaldson, et Murrieta-Flores, 2016). Il n'est pas utile de développer ici les différentes méthodes de RIG, à laquelle de nombreux travaux ont été consacrés ces dernières années. On trouvera dans (Gaio et al., 2008) et (Sallaberry, 2014) une présentation de différentes techniques utilisables pour repérer et extraire les lieux de documents textuels. Les solutions diffèrent selon que ces lieux apparaissent sous forme de toponymes, appelés dans la littérature spécialisée Entités Spatiales Nommées, ou qu'ils sont évoqués de manière implicite ou allusive. L'objectif visé dans ce papier est de proposer une méthode pour trouver les Entités Spatiales Nommées dans les textes en utilisant les marqueurs explicites simples : « rue », « place », « théâtre »

etc. La recherche de toponymes aux formes plus floues nécessite de mobiliser des solutions plus complexes.

Le deuxième domaine dans lequel nous faisons des propositions est celui de la visualisation cartographique de ces lieux, qui reste sommaire et même parfois déficiente chez F. Moretti ou dans les travaux de critique littéraire. Au-delà de l'application des règles admises de la sémiologie classique, on peut faire l'hypothèse que la cartographie littéraire doit s'adapter à la nature originale du rapport à l'espace de son objet et qu'il est nécessaire de proposer des solutions nouvelles. Dans son projet d'Atlas littéraire de l'Europe¹, une équipe de l'Université de Zürich a tenté de décrire la spécificité de ces espaces fictionnels, dont le caractère imaginaire nécessite un mode de projection original. (Piatti et al., 2008) proposent une liste de solutions pour rendre compte sémiologiquement de ces espaces littéraires par nature incertains, flous, détournés... Notre contribution dans ce papier porte sur des propositions d'un ordre un peu différent, permettant de comparer la manière dont plusieurs romans décrivent le même espace, en l'occurrence l'espace parisien.

L'objectif de cette première phase de la recherche est en effet de retrouver automatiquement dans les textes de 31 romans d'auteurs variés (voir Tableau 1) les Entités Spatiales Nommées présentes et de les utiliser pour cartographier l'empreinte spatiale de ces romans ou de leurs auteurs. Compte tenu des délais de la première phase de ce projet, nous ne travaillons pour l'instant qu'avec des solutions informatiques existantes et disponibles, articulées manuellement et sans développement informatique spécifique. La méthode proposée se décompose en trois étapes : 1) le repérage des Entités spatiales dans les textes ; 2) la localisation des Entités Spatiales Nommées dans l'espace parisien ; 3) la géovisualisation et l'analyse spatiale des résultats.

<i>La maison du Chat-qui-pelote</i> , Honoré de Balzac, 1830	<i>Jack</i> , Alphonse Daudet, 1876
<i>Ferragus</i> , Honoré de Balzac, 1833	<i>L'Assommoir</i> , Emile Zola, 1877
<i>La fille aux yeux d'or</i> , Honoré de Balzac, 1835	<i>Une belle journée</i> , Henry Céard, 1881
<i>Le Père Goriot</i> , Honoré de Balzac, 1835	<i>Pot-Bouille</i> , Emile Zola, 1882
<i>César Birotteau</i> , Honoré de Balzac, 1837	<i>Le vingtième siècle</i> , Albert Robida, 1883
<i>Les Mystères de Paris</i> , Eugène Sue, 1842	<i>Sapho</i> , Alphonse Daudet, 1884
<i>L'envers de l'histoire contemporaine</i> , Honoré de Balzac, 1848	<i>Bel-ami</i> , Guy de Maupassant, 1885
<i>Les demoiselles de magasin</i> , Paul de Kock, 1863	<i>L'œuvre</i> , Emile Zola, 1886
<i>Paris au XXème siècle</i> , Jules Verne, 1863	<i>La vie électrique</i> , Albert Robida, 1892
<i>L'éducation sentimentale</i> , Gustave Flaubert, 1869	<i>Paris</i> , Emile Zola, 1897
<i>La Curée</i> , Emile Zola, 1871	<i>La Charpente</i> , J.H. Rosny jeune, 1900
<i>Le ventre de Paris</i> , Emile Zola, 1873	<i>Mr. Bergeret à Paris</i> , Anatole France, 1901
	<i>La Maternelle</i> , Léon Frapié, 1904
	<i>La Vague rouge</i> , J.H. Rosny aîné, 1910
	<i>Dans les rues</i> , J.H. Rosny aîné, 1913

¹ <http://www.literaturatlas.eu/en/>

Tableau 1. Les 31 romans du corpus

3. La réalisation

3.1. Le repérage des Entités Spatiales Nommées dans le texte

Il est d'abord nécessaire de rassembler et de corriger les versions numériques des textes des romans, souvent fautives car tirées de procédures OCR. On obtient un corpus de 4 778 464 mots, interrogeable dans son ensemble ou par fraction (Figure 1). Ce corpus est chargé dans la plateforme d'analyse textométrique TXM¹ (Heiden, 2010) (Heiden, Magué, et Pincemin, 2010).

TXM propose notamment un module logiciel appelé *Corpus Query Language* (CQP) qui traite des requêtes et fonctionne comme un moteur de recherche. Les requêtes sont construites en *Corpus Query Language* (CQL) pour exprimer un motif linguistique (un ou plusieurs mots) au moyen de leurs propriétés (forme graphique, lemme ou encore catégorie grammaticale). Ces requêtes permettent ainsi de trouver toutes les mentions d'Entités Spatiales Nommées des différentes catégories : voies (rue, avenue, boulevard, quai, etc.), cours d'eau (rivière, bassin, canal...), bâtiments (théâtre, hôpital, école...), parcs et autres. Le Tableau 2 en donne un exemple révélateur de la difficulté pratique de l'analyse de données textuelles. L'expression est d'abord conçue pour chercher toutes les occurrences d'une catégorie nommée, à savoir les rues. Pour en appréhender les diverses réalisations, il importe de neutraliser la casse (%c), les éventuels signes diacritiques (%d) et le pluriel : rechercher un lemme permet de saisir un mot sous ses formes variables (en les ramenant à une entrée du dictionnaire). Le reste de l'expression spécifie progressivement la requête. Il s'agit d'abord de préciser que la catégorie nommée doit être suivie d'un ou de plusieurs mots commençant par une majuscule ([word="\p{Lu}.*"]): la « rue Plumet » et la « rue du Grenier-Saint-Lazare » par exemple sont désormais considérées. Pour ne pas négliger la « rue du Temple » ou la « rue de l'Homme-Armé », la requête doit préciser qu'un ou deux mots ([]?) peuvent s'insérer entre la rue et le mot avec majuscule. Il faut encore en exclure quelques formes graphiques et plus particulièrement un éventuel signe de ponctuation qui signalerait le début d'une nouvelle proposition : des chaînes telles que « rue. Il » ou « rue... Ah » se voient ainsi exclues de la recherche.

Au terme d'un enrichissement progressif, la requête s'est complexifiée mais débouche sur un index des rues qui peut être traité dans les meilleurs délais. Certes cette requête *ad hoc* ne met en œuvre qu'une recherche ciblée et doit être adaptée lorsque de nouveaux questionnements se font jour. Mais plus généralement le langage CQL se prête bien à la recherche de tout motif linguistique (et non seulement des entités spatiales), si bien que son apprentissage peut être remobilisé ensuite selon un point de vue de sémantique interprétative (en s'intéressant par exemple aux adjectifs qui qualifient les rues parisiennes).

¹ <http://textometrie.ens-lyon.fr/>

```
[frlemma="rue"%cd] [word!="\.,\;\;|\:|\!|\?|\.\.\-|une|\-
|où|ainsi|et|aurait|-l"%c]? [word!="\.,\;\;|\:|\!|\?|\.\.\-|une|\-
|où|ainsi|et|aurait|-l"%c]? [word="p{Lu}.*" &
word!="Ça|Ah|O|Venez|Et|M|L."]
```

Tableau 2. Exemple de requête TXM pour extraire les rues des 31 romans

TXM fournit une liste de concordances, c'est-à-dire l'ensemble des Entités Nommées trouvées par la requête, encadrées par les passages du texte où elles apparaissent. (Tableau 3). Il est possible de naviguer dans la liste des concordances obtenues et d'accéder en permanence au texte lui-même à partir de l'extrait correspondant. Cela donne à l'analyste un outil précieux pour explorer son corpus textuel à partir des Entités Spatiales Nommées et analyser le contenu littéraire et les descriptions associées aux différents lieux. Cette fonction a permis par exemple de rechercher rapidement les mentions des passages et galeries couvertes, pour faire une analyse approfondie de cet élément urbanistique très important du Paris de l'époque.

text_id	Contexte gauche	Pivot	Contexte droit
education	qu'elle était sur le trottoir de la	rue Tronchet	depuis longtemps. Elle y attendait quelque chose d'indéterminé, de
education	blaguait, était très gai. Par la	rue Duphot	, ils atteignirent les boulevards. Des lanternes vénitiennes, suspendues aux
education	direct impossible ; et ils entraient dans la	rue Caumartin	, quand, tout à coup, éclata derrière eux un bruit
education	l'emmena jusqu'à l'hôtel de la	rue Tronchet	, dans le logement préparé pour l'autre. Les fleurs n'
education	étaient partis. A l'angle de la	rue Saint-Honoré	, des hommes en blouse le croisèrent en criant : « Non
education	. Des compagnies de la ligne occupaient les	rues Saint-Thomas	et Fromanteau. Une barricade énorme bouchait la rue de Valois.
education	et Fromanteau. Une barricade énorme bouchait la	rue de Valois	. La fumée qui se balançait à sa crête s'entrouvrit,
education	s'en retournèrent au Palais-Royal. Devant la	rue Fromanteau	, des cadavres de soldats étaient entassés sur de la paille.
education	barricades du quartier Latin, s'était battu	rue Rambuteau	, avait sauvé trois dragons, était entré aux Tuileries avec la
education	, et lui annonça qu'il existait,	rue Saint-Jacques	, un club intitulé le Club de l'Intelligence. Un nom
education	« Le citoyen Jean-Jacques Langreneux, typographe,	rue Dauphine	, voudrait qu'on élevât un monument à la mémoire des martyrs
education	meubles et quitter le bel appartement de la	rue Drouot	. Elle en prit un autre, sur le boulevard Poissonnière,
education	de Frédéric. Puis il l'emmena déjeuner	rue de Chartres	, chez Party et, comme il avait besoin de se refaire
education	; et, quand ils remontèrent ensemble la	rue Vivienne	, leurs épaulettes se touchaient fraternellement. Rentré chez lui, Frédéric
education	des Gobelins et de l'Hôpital et les	rues Godefroy	et Mouffetard. Quatre barricades formaient, au bout des quatre voies
education	Fontainebleau au secours d'un camarade blessé logeant	rue Bellefond	, personne d'abord ne voulut le croire ; on examina ses
education	le ranima. Ils tournèrent ensuite par la	rue du Marché-aux-Chevaux	. Le Jardin des Plantes, à droite, faisait une grande
education	accompagna jusqu'à l'Ecole polytechnique. La	rue Saint-Victor	était toute sombre, sans un bec de gaz ni une lumière
education	caporal pour le mener au haut de la	rue Saint-Jacques	, à la mairie du XIIe arrondissement. La place du Panthéon
education	, l'annonçait. Au haut de la	rue Vivienne	, un peloton de mobiles parut. Alors, les bourgeois poussèrent
education	au haut d'une barricade, dans la	rue Lafayette	, un gamin enveloppé d'un drapeau tricolore criait aux gardes nationaux
education	retourna chez lui ; car il possédait,	rue Saint-Martin	, une maison où il s'était réservé un pied-à-terre ; et
education	les autres, il s'enfonça dans la	rue Hauteville	, fit le tour du Gymnase, revint sur le boulevard,
education	sa fille, à l'entrée de la	rue Saint-Denis	. Ils s'en retournèrent sans rien dire ; lui, n'
education	qu'elle allait avec sa bonne dans la	rue Rumford	chercher un médecin. On les laissa passer. Au coin de
education	. Il était sur le trottoir de la	rue Paradis	, devant une maison ; et il regardait au second étage,
education	. C'était une succursale intime de la	rue de Poitiers	. Il y rencontra le grand M. A., l'
education	et Dufour. Ces entretiens continuèrent dans la	rue de la Roquette	, bordée par des boutiques, où l'on ne voit que
education	fiacre et partit. Au coin de la	rue de Marbeuf	, il lut sur une planche en grosses lettres : « Maison
education	» « Arnoux ! » « Oui :	rue de Fleurus	! » « Quel numéro ? » « Est -ce que je
education	, » à sept heures du soir,	rue Jacob	. Il avait même son passeport, par précaution ; et il
education	dans sa boutique. Mais il logeait toujours	rue Paradis	, car il possédait deux domiciles. Rue Paradis, le portier
education	Paradis, car il possédait deux domiciles.	Rue Paradis	, le portier jura que M. Arnoux était absent depuis la
education	fit conduire chez lui, à Montmartre,	rue de l'Empereur	. Sa maison était flanquée d'un jardin, clos par une
education	novembre, Frédéric, en passant dans la	rue de Mme	Arnoux, leva les yeux vers ses fenêtres, et aperçut contre
education	il avait vu au premier dîner dans la	rue de Choiseul	, qui ensuite avait été chez Rosanette, était revenu chez Mme
education	le faubourg Saint-Martin. Au coin de la	rue de Provence	, Frédéric mit pied à terre pour gagner les boulevards. Il
education	Tous les soirs, régulièrement, depuis la	rue de Grammont	jusqu'à la rue Montmartre, il se traîne devant les cafés
education	depuis la rue de Grammont jusqu'à la	rue Montmartre	, il se traîne devant les cafés, affaibli, courbé en
envers	sans caractère assises entre le Terrain, la	rue d'Arcole	, la cathédrale et le quai de la Cité. Ce point
envers	en cas de convenance mutuelle. S'adresser	rue Chanoinesse	, près Notre- Dame, à monsieur Millet, épicier, qui
envers	émut à cette parole, et il revint	rue Chanoinesse	, où il ne voulait déjà plus aller. Le prêtre et
envers	un que l'autre d'entrer dans la	rue Massillon	, qui fait face au petit portail nord de la cathédrale,
envers	la cathédrale, de tourner ensemble dans la	rue Chanoinesse	, à l'endroit où, vers la rue de la Colombe

Tableau 3. Les concordances obtenues avec la requête du Tableau 2

La plateforme TXM permet des analyses statistiques sur les occurrences de mots ou de syntagmes en fonction d'entités définies par l'utilisateur : romans, auteurs, période d'édition ou de narration, etc. (Figures 1 et 2).

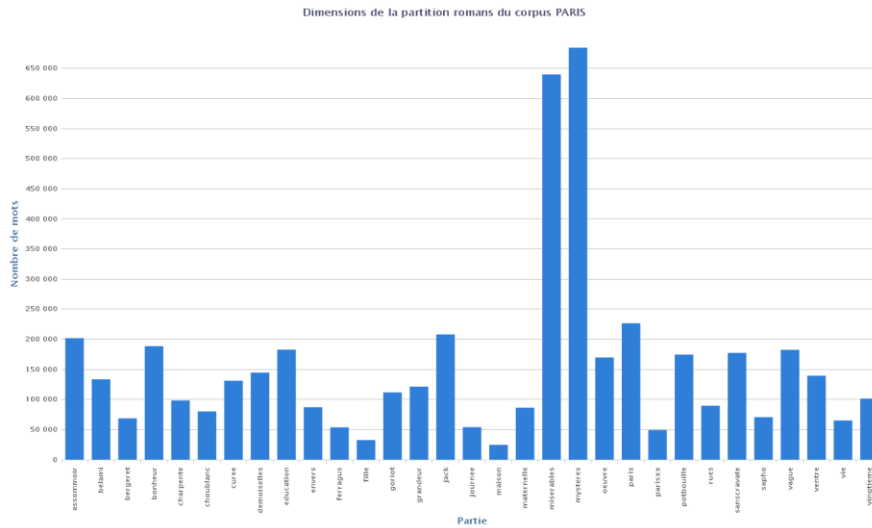


Figure 1. Le nombre de mots par roman

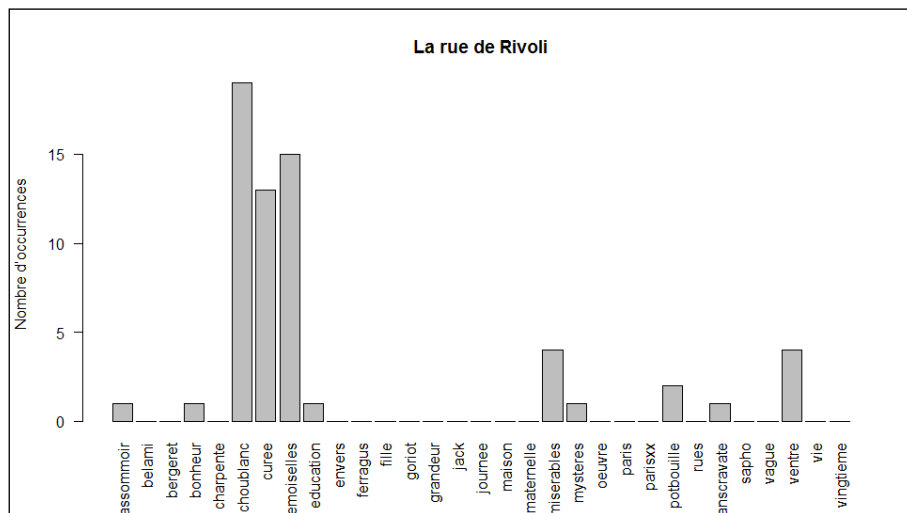


Figure 2. Le nombre de mentions de la rue de Rivoli par roman

3.2. La constitution d'un index géohistorique des lieux littéraires parisiens

La localisation des Entités Spatiales Nommées trouvées dans les textes nécessite une recherche approfondie dans les sources historiques. Il n'existe par exemple pas d'index constitué des rues de Paris portant indication des changements de nom au cours du temps. Il n'y a pas non plus d'annuaire général ou global des théâtres ou des églises existant aux différentes époques. *Gazetteer* est le nom couramment donné dans la littérature aux atlas toponymiques. Les *Gazetteers* généralistes et contemporains tels que Geoname¹ s'avèrent de peu de secours dans ce contexte. Toute cartographie romanesque parisienne nécessite la constitution d'un *Gazetteer* historique, intégrant le plus grand nombre possible d'Entités Spatiales Nommées précisément localisées à Paris. Les informations existent dans de nombreux documents dispersés. Certaines informations se trouvent directement sur le Web (sur Wikipedia ou DbPedia ou sur divers sites spécialisés), d'autres dans des documents conservés dans les archives, dont certains sont publiés sous forme de collections en ligne. Les localisations précises des entités se trouvent principalement dans les cartes et plans historiques de Paris. Il est donc indispensable de constituer parallèlement au corpus des textes de roman un corpus de cartes devant répondre à cet objectif de localisation. Il servira aussi à documenter l'espace de référence des romans en facilitant le repérage cartographique des occurrences de lieu trouvées dans les textes.

Pour le début de la période (1830), nous avons pris comme référence principale les données du plan Vasserot, disponibles au téléchargement en ligne sur le site du projet Alpage² (Figure 3) (Noizet, Bove, et Costa, 2013). Au plan lui-même, superbe document accessible en format raster, ont été ajoutés par le projet Alpage les limites des bâtiments et les filaires de voies au format vecteur. On constate cependant que seul un faible nombre de noms de rues sont renseignés dans la base. Pour la période de la fin du siècle et les années 1900, alors que les limites de Paris se sont fortement étendues suite à l'annexion en 1860 des communes ou parties de communes périphériques et aux fortes transformations d'Hausmann dans le centre de Paris, nous nous appuyons sur un plan de rue courant de l'époque. D'autres pourront être ajoutés. Ces plans sont projetés dans le référentiel contemporain du plan de rue disponible en *opendata* sur le site de l'APUR³ afin de permettre ultérieurement une visualisation des localisations romanesques sur les plans actuels, dans une optique de développement d'un tourisme littéraire.

Un premier travail de référencement des rues a montré que les données disponibles dans la base de données d'Alpage et dans le fichier de rues contemporains ne permettaient de retrouver qu'environ un tiers des noms de rues présents dans les 31 romans. Un travail de recherche est en cours pour compléter ce *gazetteer* et l'étendre, au-delà de la voirie, aux églises et aux théâtres en priorité. La structure de données adoptée permettra d'associer toutes les Entités Spatiales Nommées trouvées dans les romans à un nom de rue ayant existé à l'époque en fondant l'association sur la date du roman ou l'époque du récit. Il est à noter que

¹ <http://www.geonames.org/>

² <http://alpage.huma-num.fr>

³ <http://www.apur.org/article/donnees-disponibles-open-data>.

constituer un tel *gazetteer* des lieux parisiens du XIXe reste une tâche d'ampleur raisonnable même s'il concerne un espace urbain vaste et complexe. Il est d'abord finalisé vers la cartographie des espaces romanesques d'une période donnée, 1830-1914. Il peut aussi être partagé avec d'autres projets visant la connaissance de l'espace parisien à la même époque. Il peut enfin être conçu comme un projet collaboratif. On sait l'intérêt grandissant pour les initiatives de localisation des œuvres culturelles, qu'elles soient cinématographiques ou littéraires (Joliveau, 2009), dont le site *Placing Literature*¹ est une bonne illustration. Il est possible d'envisager ce *gazetteer* comme une application de ce qu'on pourrait appeler suite à (Goodchild, 2007) une démarche d'information géographique littéraire volontaire.

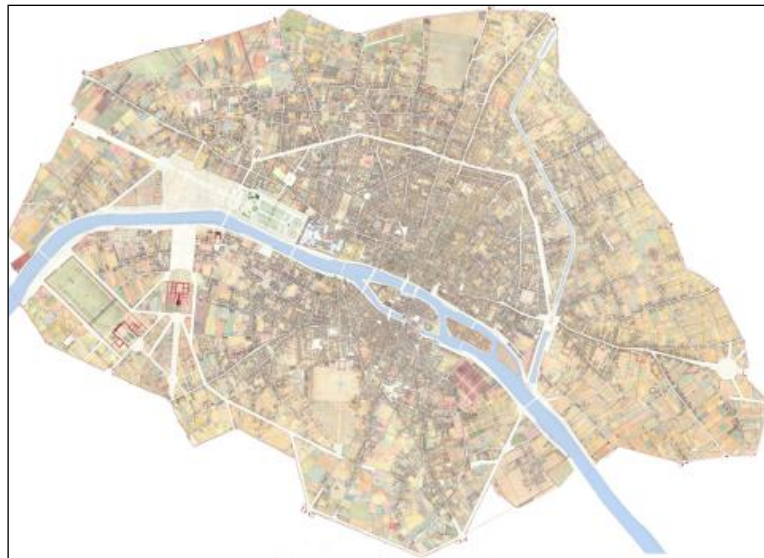


Figure 3. Le plan Vasserot tiré du site *Alpage* (source : <http://alpage.huma-num.fr>)

3.3. La géovisualisation et l'analyse spatiale

Les Entités Spatiales Nommées localisées dans les romans peuvent facilement être associées à leur position géographique dans un SIG. Il devient alors aisé de construire des requêtes thématiques et spatiales, de produire des géovisualisations et des cartes concernant l'ensemble du corpus, un roman, un auteur, une période ou un quartier particuliers. Nous avons fait des tests à partir des Entités Spatiales Nommées de type « rue » déjà localisées, qui correspondent à un tiers environ des 900 rues repérées avec TXM dans les 31 romans. Plusieurs formes de géovisualisation ont été testées pour représenter les nombres de citations des rues : par ligne ou par point, par symboles textuels, qualitatifs ou quantitatifs (Figures 4 à

¹ <http://www.placingliterature.com/>

7). Le choix de la meilleure visualisation dépend des objectifs et du mode de rendu (carte papier, visualisation interactive à l'écran ...). Les modes d'implantations par symboles proportionnels ponctuels ou linéaires sont plus clairs mais sont moins évocateurs du phénomène que les variations de taille de label, en particulier quand ils conservent les orientations des rues (Fig. 4).

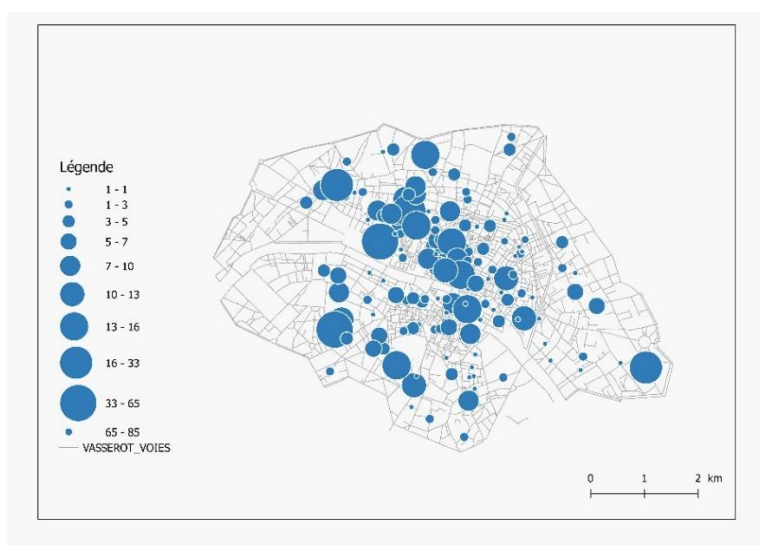


Figure 4. Nombre de citations de rue dans les romans par symboles proportionnels en implantation ponctuelle (échantillon 1/3)

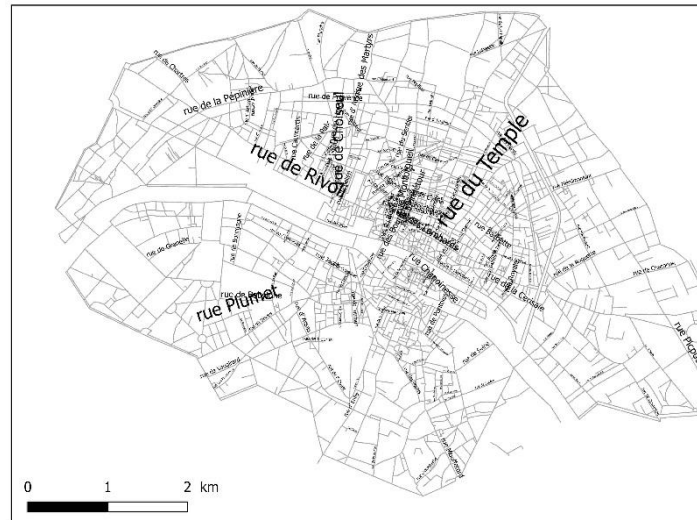


Figure 5. Nombre de citations des rues dans les romans par taille de label proportionnelle en implantation linéaire (échantillon 1/3)

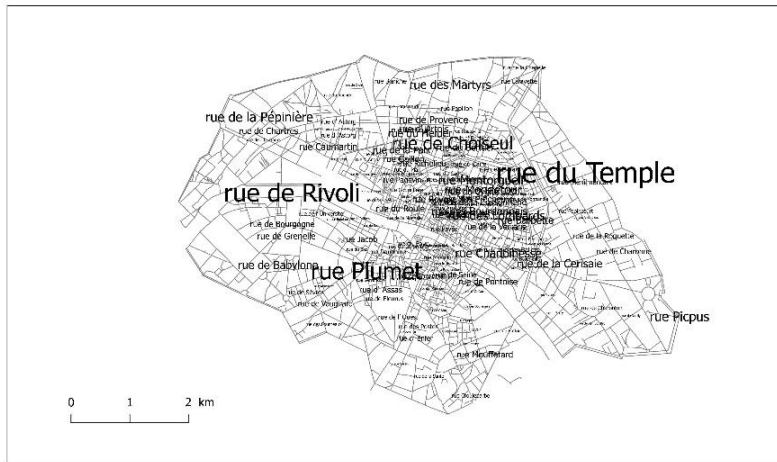


Figure 6. Nombre de citations de rues dans les romans par taille de label proportionnelle en implantation linéaire (échantillon 1/3)



Figure 7. Nombre de citations de rues dans les romans par symboles proportionnels en implantation linéaire (échantillon 1/3)

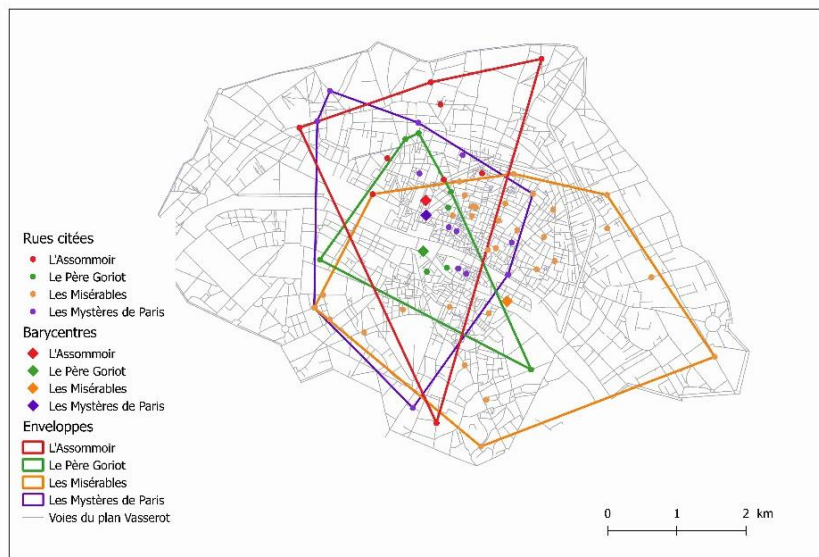


Figure 8. Les citations de rues dans quatre romans (semis de point, enveloppes convexes et barycentres) (échantillon 1/3)

L'usage de fonctions de mesure et d'analyse spatiale permet de synthétiser les informations de manière à comparer les empreintes spatiales de différents romans ou différents auteurs. Les Figures 8 à 10 illustrent l'emploi de différents indicateurs classiques de statistiques spatiales pour résumer des semis de points : enveloppes convexes, centres moyens ou barycentres pondérés, ellipses de dispersion standard et interpolations spatiales à partir des rues nommées dans quatre romans. Cela permet de simplifier le tracé des figures. On constate sur l'échantillon de voies concerné que les empreintes de *L'Assommoir*, du *Père Goriot*, *des Misérables* et des *Mystères de Paris* se chevauchent au centre de la ville et se différencient en allant vers la périphérie. Ces calculs nécessitent cependant tous de réduire le linéaire de rue à son centroïde et conduisent à perdre une information importante.

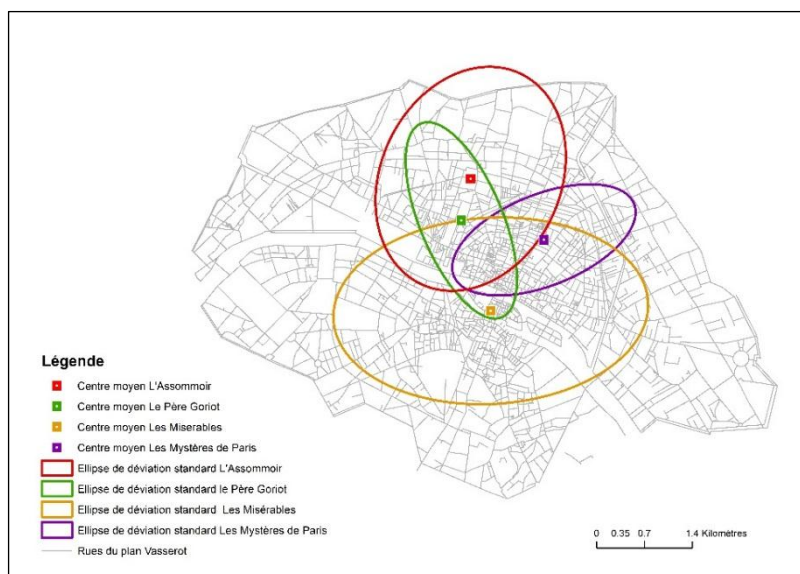


Figure 9. Les citations de rues dans quatre romans (centres moyens et ellipses spatiales de dispersion) (échantillon 1/3)

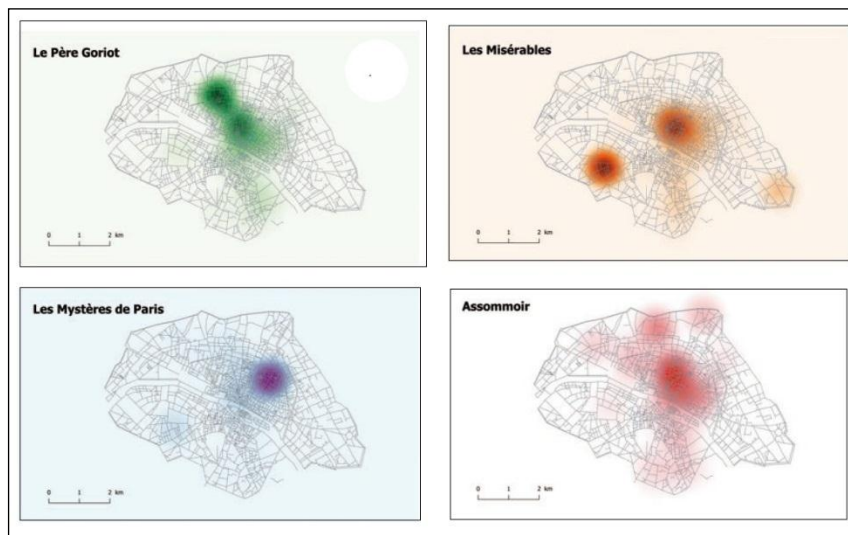


Figure 10. Les citations de rues dans quatre romans par interpolation spatiale (échantillon 1/3)

4. Discussion et perspectives

Les premiers résultats de ce projet démontrent l'intérêt de notre proposition méthodologique pour produire une cartographie romanesque des romans parisiens du XIXe siècle à partir d'une recherche automatique des Entités Spatiales Nommées. Les requêtes TXM semblent fonctionner de manière efficace et répondre aux besoins. Des problèmes liés à des caractères non pris en compte dans les requêtes ont été corrigés par les auteurs de la plateforme. D'autres problèmes inattendus sont apparus, telle que l'existence de deux signes typographiques différents pour les apostrophes qui ont contraint à homogénéiser le corpus. Des analyses complémentaires doivent maintenant être menées pour repérer des cas où les rues non trouvées ne proviendraient pas des limites actuelles de notre Gazetteer : graphies erronées, structures de phrases complexes et résoudre la question des noms de rues inventés par l'auteur. Le choix de TXM pour extraire des Entités Nommées était lié à deux motivations principales. D'abord la plateforme est relativement simple à mettre en œuvre sans compétences informatiques trop spécialisées. Ensuite, elle propose une interface interactive pour explorer les textes par l'intermédiaire des Entités Spatiales Nommées, ce qui est particulièrement utile dans notre cas d'usage. Il existe d'autres outils de reconnaissance d'entités nommées ou d'autres plateformes de traitement de corpus textuels qui intègrent ce type de fonction comme GATE par exemple (<https://gate.ac.uk/>). Un travail d'analyse comparée de leurs performances et de l'intérêt qu'ils présentent pour un projet comme le nôtre est à faire.

Le travail de constitution de l'index géographique des Entités Spatiales Nommées n'étant pas terminé, il n'est pas encore possible de valider l'intérêt des géovisualisations pour l'analyse des empreintes spatiales des différents romans ou auteurs. Mais même sans disposer de cartes ou d'indicateurs exhaustifs, il reste

indéniable que l'accès direct au corpus de texte par l'intermédiaire des noms de lieux transforme fortement les manières d'explorer l'espace et les paysages romanesques parisiens. Disposer d'outils d'interrogation, de visualisation et d'analyse spatiale s'avère utile pour émettre des hypothèses et synthétiser les résultats liés à la spatialisation des romans.

La prochaine étape est d'utiliser la plate forme textométrique pour caractériser automatiquement le contenu sémantique associé aux Entités Spatiales Nommées en recherchant les qualificatifs employés pour les décrire. Par exemple, il sera possible de préciser si une rue est grande, sombre, animée ... ou bien quels personnages sont convoqués dans tel ou tel lieu. Une autre piste serait de représenter la dynamique temporelle des lieux dans la chronologie du récit. Il en effet possible d'associer à chacun des citations sa place dans le texte par le chapitre ou même par son rang dans l'ordre des mots. L'objectif à terme serait de construire une plateforme facilitant l'analyse spatiale et géographique des Entités Spatiales Nommées ouvertes à tous les romans évoquant Paris à la même époque. Cette plateforme pourrait être mobilisée par les opérateurs du tourisme littéraire, en donnant un accès *in situ* aux extraits de texte associés à des lieux explicitement nommés. Comme nous l'avons déjà écrit, le *geocrowdsourcing* (ou information géographique volontaire) permettrait de documenter littérairement l'espace parisien et de spatialiser les œuvres littéraires traitant Paris au XIXe siècle.

Pour conclure, soulignons que l'équipe ne pense pas que la dimension géographique des romans ni même les questions de localisation romanesque se réduisent à la cartographie d'Entités Spatiales Nommées trouvées dans les textes littéraires. Mais celles-ci, qui sont assez aisées à localiser de manière automatique, peuvent servir de point d'ancrage pour cartographier des lieux non nommés explicitement et explorer la dimension géographique des textes romanesques, moyennant la création de nouvelles méthodes de recherche d'information géographique textuelle.

NB. Merci aux relecteurs pour leurs remarques judicieuses qui ont permis d'améliorer ce texte.

Bibliographie

Bulson E. *Novels, Maps, Modernity: The Spatial Imagination, 1850-2000*. 1^{re} éd.[s.l.] : Routledge, 2007. 188 p.ISBN : 0-415-97648-0.

Cooper D., Donaldson C., Murrieta-Flores P. *Literary Mapping in the Digital Age*. [s.l.] : Routledge, 2016. 457 p.ISBN : 978-1-317-10455-1.

Gaio M., Sallaberry C., Etcheverry P., Marquesuzaa C., Lesbegueries J. « A global process to access document's contents from a geographical point of view ». *J. Vis. Lang. Comput.* 2008. Vol. 19, n°n° 1, p. 3–23.

Goodchild M. F. « Citizens as sensors ; the world of volunteered geography ». *GeoJournal*. 2007. Vol. 69, n°4, p. 211-221.

Gregory I., Donaldson C., Murrieta-Flores P., Rayson P. « Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research ». *International Journal of Humanities and Arts Computing* [En ligne]. mars 2015. Vol. 9, n°1, p. 1-14. Disponible sur : < <https://doi.org/10.3366/ijhac.2015.0135> >

Heiden S. « The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme ». In : *24th Pacific Asia Conference on Language, Information and Computation*. [s.l.] : Institute for Digital Enhancement of Cognitive Development, Waseda University, 2010. p. 389-398. Disponible sur : < <https://halshs.archives-ouvertes.fr/halshs-00549764/document> > (consulté le 16 juin 2016)

Heiden S., Magué J.-P., Pincemin B. « TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement ». In : *10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*. [s.l.] : Edizioni Universitarie di Lettere Economia Diritto, 2010. p. 1021-1032. Disponible sur : < <https://halshs.archives-ouvertes.fr/halshs-00549779/document> > (consulté le 16 juin 2016)

Joliveau T. « Connecting Real and Imaginary Places through Geospatial Technologies: Examples from Set-Jetting and Art-Oriented Tourism ». *Cartographic Journal*. février 2009. Vol. Vol. 46, n°1, Cinematic Cartography Special Issue, p. 36-45.

Moretti F. *Atlas du roman européen*. Paris : Editions du Seuil, (édition originale italienne : 1997), 2000.

Moretti F. *Graphes, cartes et arbres. Modèles abstraits pour une autre histoire de la littérature*. Paris : Les prairies ordinaires (Edition originale en anglais. 2005), 2008.

Noizet H., Bove B., Costa L. *Paris, de parcelles en pixels : Analyse géomatique de l'espace parisien médiéval et moderne*. Saint-Denis (Seine-Saint-Denis); Paris : Coédition PU Vincennes, 2013. 354 p. ISBN : 978-2-84292-364-8.

Piatti B., Bär H., Reuschel A., Hurni L., Cartwright W. « Mapping Literature: Towards a Geography of Fiction ». In : *Proceedings of the Art and Cartography Conference, Berlin Heidelberg: Springer*. Vienna : [s.n.], 2008. Disponible sur : < <http://www.literaturatlas.eu/2012/03/23/ein-literarischer-atlas-europas-poster/> > (consulté le 8 novembre 2010)

Sallaberry C. *Recherche d'information géographique dans les corpus textuels*. Collection Systèmes d'Information Géographique. [s.l.] : ISTE, 2014. 190 p. ISBN : 978-1-78405-012-2.