

CARTOGRAPHIER LES LIEUX PARISIENS NOMMÉS DANS UN CORPUS DE ROMANS DU XIX^e SIÈCLE

PROPOSITION MÉTHODOLOGIQUE

Noémie Boeglin¹, Michel Depeyre¹, Thierry Joliveau¹, Yves-François Le Lay², Florian Latour¹

OBJECTIF : Réaliser de manière automatique un inventaire cartographique des lieux nommés dans un corpus de 31 romans du XIX^e siècle en proposant une méthodologie rigoureuse, opérationnelle et généralisable.



La maison du Chat-qui-pelote, Honoré de Balzac, 1830
Ferragus, Honoré de Balzac, 1833
La fille aux yeux d'or, Honoré de Balzac, 1835
Le Père Goriot, Honoré de Balzac, 1835
César Birotteau, Honoré de Balzac, 1837
Les Mystères de Paris, Eugène Sue, 1842
Sans Cravate ou les Commissionnaires, Paul de Kock, 1844
L'envers de l'histoire contemporaine, Honoré de Balzac, 1848
M. Choublanc à la recherche de sa femme, P. de Kock, 1856
Les Misérables, Victor Hugo, 1862

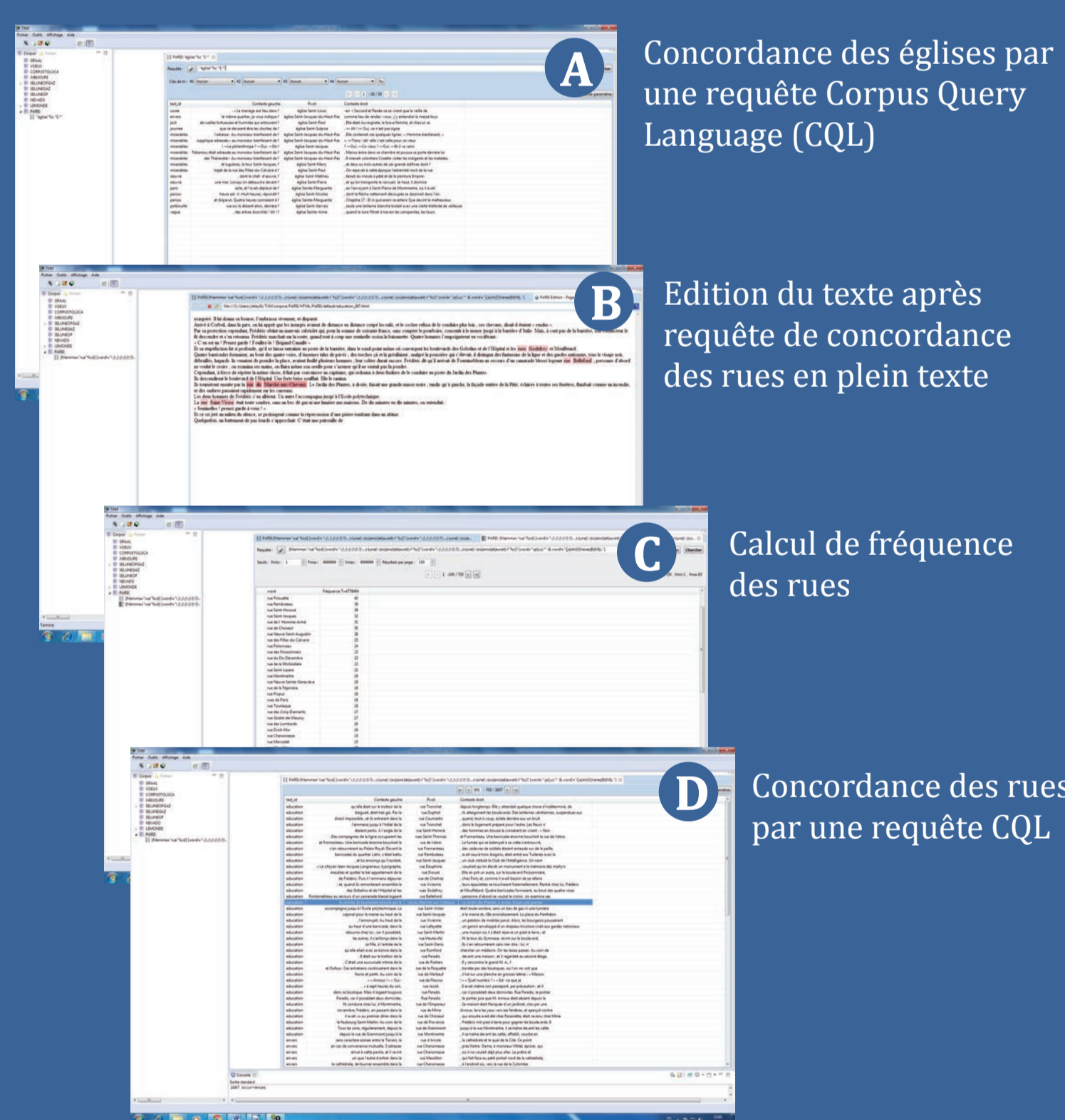
Les demoiselles de magasin, Paul de Kock, 1863
Paris au XX^e siècle, Jules Verne, 1863
L'éducation sentimentale, Gustave Flaubert, 1869
La Curée, Emile Zola, 1871
Le ventre de Paris, Emile Zola, 1873
Jack, Alphonse Daudet, 1876
L'Assommoir, Emile Zola, 1877
Une belle journée, Henry Céard, 1881
Pot-Bouille, Emile Zola, 1882
Le vingtième siècle, Albert Robida, 1883

Au Bonheur des dames, Emile Zola, 1883
Sapho, Alphonse Daudet, 1884
Bel-ami, Guy de Maupassant, 1885
L'œuvre, Emile Zola, 1886
La vie électrique, Albert Robida, 1892
Paris, Emile Zola, 1897
La Charpente, J.H. Rosny jeune, 1900
Mr. Bergeret à Paris, Anatole France, 1901
La Maternelle, Léon Frapié, 1904
La Vague rouge, J.H. Rosny aîné, 1910
Dans les rues, J.H. Rosny aîné, 1913

Les 31 romans

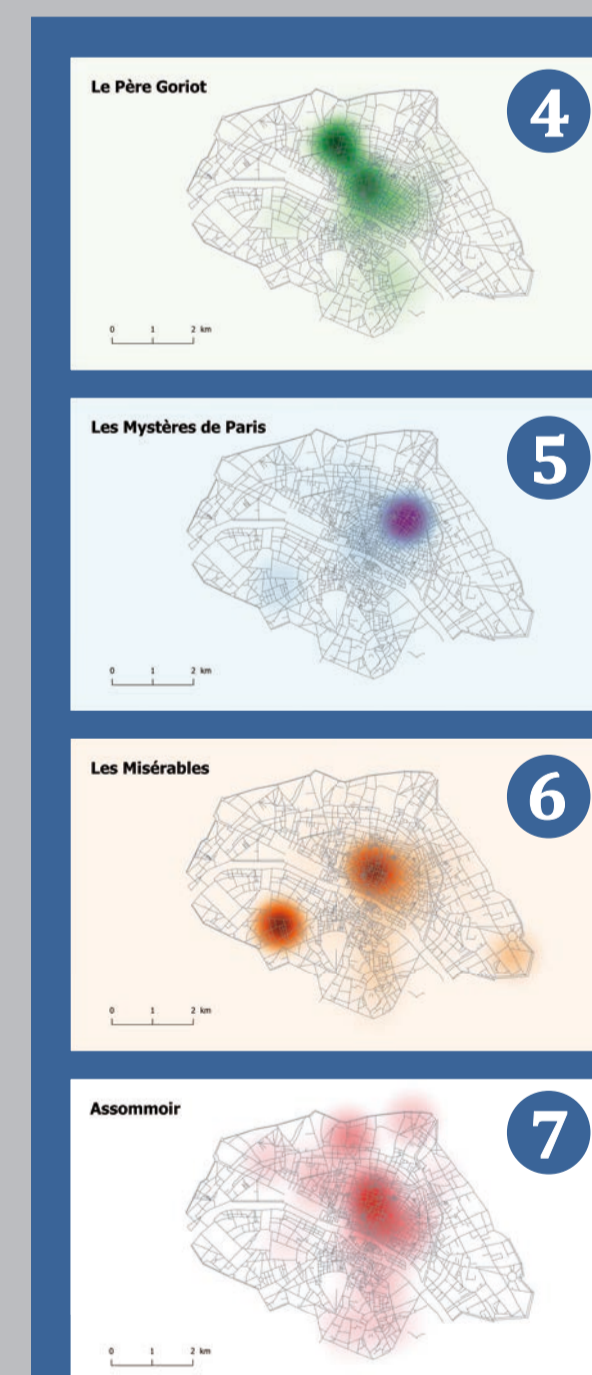
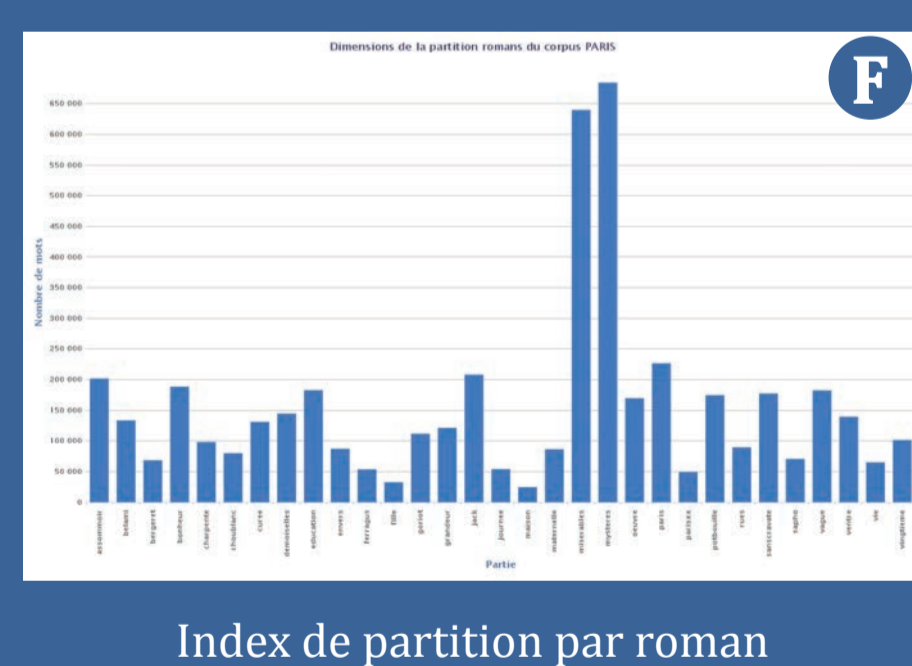
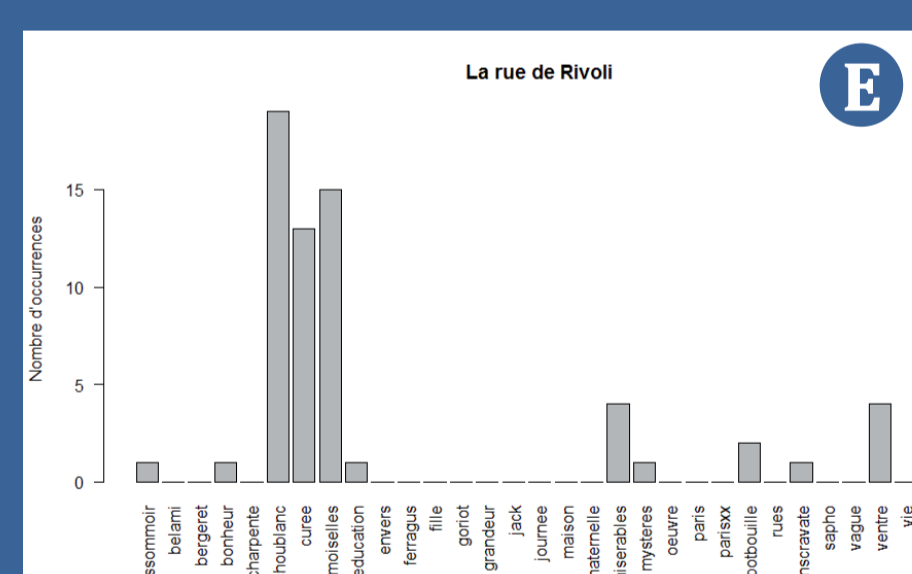
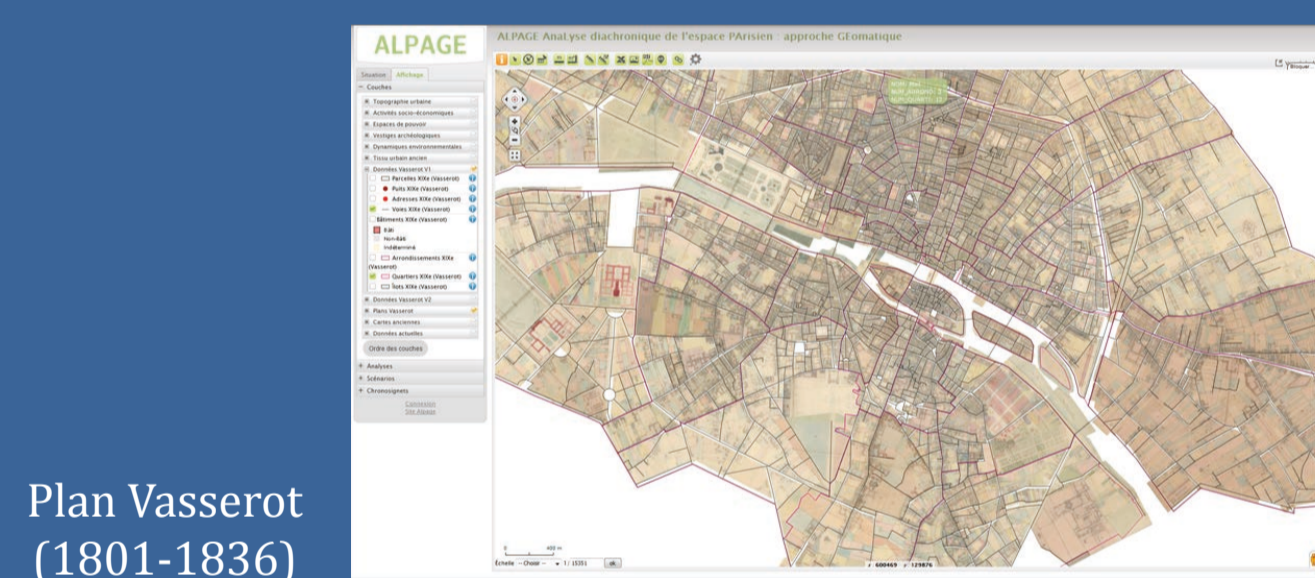
1. REPÉRAGE DES ENTITÉS NOMMÉES DANS LE TEXTE (GEOPARSING)

- Les versions numériques des textes des romans sont intégrées dans la plateforme d'analyse textométrique TXM (4779785 mots au total).
- Des requêtes TXM retrouvent automatiquement toutes les mentions d'entités spatiales nommées : voies (rues, avenues ...), cours d'eau, bâtiments (théâtre, hôpital, école ...), etc.
- TXM établit les concordances et fournit des statistiques par roman, auteur, date...
- A partir de ces concordances, il est possible de parcourir interactivement le texte.



2. CONSTITUTION D'UN INDEX GÉOHISTORIQUE

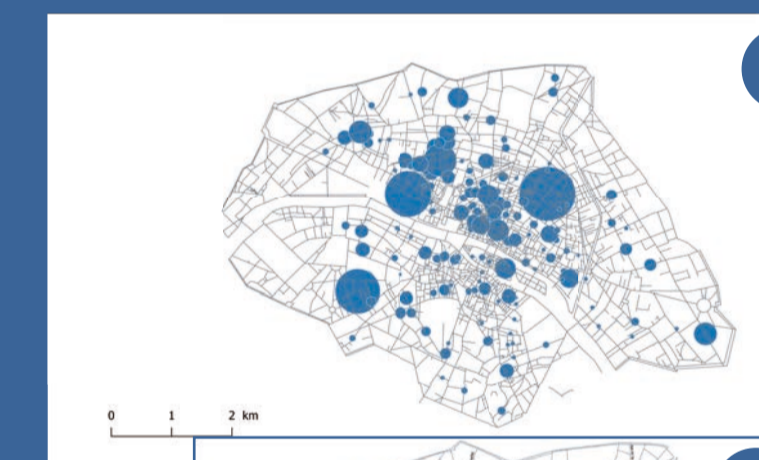
- Un Système d'Information Géographique (SIG) est constitué pour géoréférencer les entités spatiales nommées (plans historiques comme le Plan Vasserot sur le site Alpage, annuaires et autres sources historiques).
- Une jointure ou un géocodage permet d'associer les entités spatiales trouvées par TXM à celles du SIG.



3. GÉOVISUALISATION & ANALYSE SPATIALE

- Le SIG offre de multiples fonctions d'interrogation et de cartographie, par point ou par ligne, symbolique ou textuelle (nom de rue ou de l'entité) pour l'ensemble du corpus ou par roman, par auteur, par période, par quartier, etc.
- Les outils de mesure et d'analyse spatiale pourront aider à repérer et comparer les empreintes spatiales d'un roman ou d'un auteur (barycentres, ellipses de dispersion des entités nommées, etc.).

1 : Carte en symboles proportionnels de toutes les rues citées dans le corpus

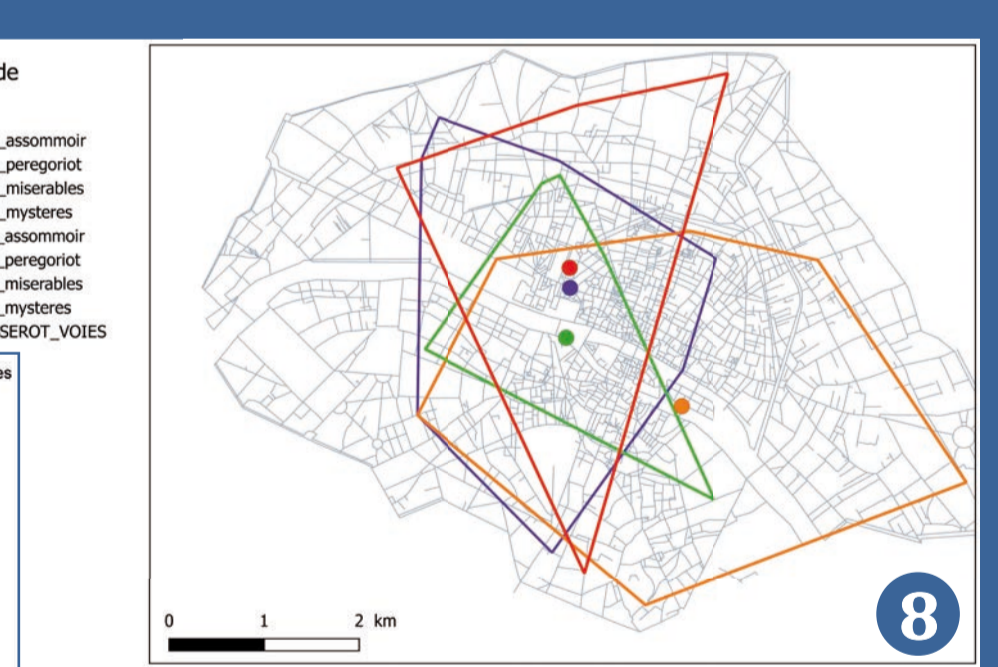
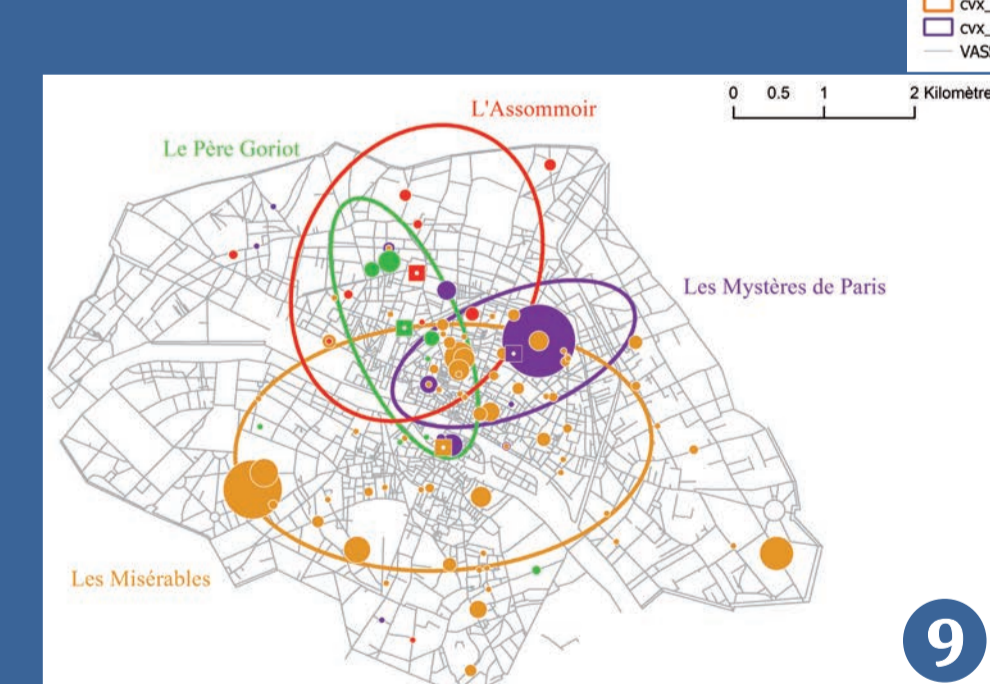


2 : Carte par interpolation spatiale du nombre total de rues dans le corpus



3 : Carte en libellés à la taille proportionnelle au nombre de rues dans le corpus

8 : Centres géométriques et enveloppes des rues citées dans 4 romans



9 : Symboles proportionnels, centres géométriques et ellipses d'écart-type pondérées par le nombre d'occurrences des rues de 4 romans

Visualisation des rues nommées dans un corpus de 31 romans (traitement partiel sur 1/3 des localisations)

PREMIER BILAN :

- Un projet encore récent
- Des requêtes TXM efficaces
- Un index géohistorique des lieux parisiens complexe à construire
- Même en l'état, une nouvelle manière d'explorer les lieux des romans

ET ENSUITE ?

- A terme construire une plateforme facilitant l'analyse spatiale et géographique des entités nommées dans les textes évoquant Paris au XIX^e.
- Usages envisagés :
 - o faciliter le tourisme littéraire, en donnant accès facilement *in situ* aux extraits de texte associés à des lieux explicitement nommés et développer une approche de *geocrowdsourcing* littéraire ;
 - o caractériser automatiquement le contenu sémantique associé aux entités nommées (qualificatifs employés, personnages concernés, ...)
 - o déduire les lieux implicites à partir des lieux explicitement nommés.

OUTILS UTILISÉS :
 La plateforme de textométrie TXM : <http://textometrie.ens-lyon.fr/>
 Les logiciels SIG QGIS et ArcGIS
DONNÉES UTILISÉES :
 Site Alpage : <http://alpage.huma-num.fr/fr/ressources/donnees-sig>
 APUR : <http://www.apur.org/article/donnees-disponibles-open-data>