



Pour une détection automatique de l'espace personnage dans le roman

Jean Barré, Pedro Cabrera Ramirez, Frédérique Mélanie-Becquet, Ioana Galleron
26 juin 2023

Lattice : ENS-PSL-CNRS

« À quelle fréquence, à quel moment et pendant combien de temps un personnage apparaît-il dans le texte ? » (Woloch, 2003) [1]

Mais au fait, qu'est-ce qu'un personnage ?

Conceptualisation de l'espace personnage dans le roman

- Le personnage - projection psychologique vs être de papier (Barthes, 1970) [2]
- L'espace personnage : « la rencontre entre une personnalité humaine individuelle et une position déterminée dans l'ensemble du récit » (Woloch, 2003) [1].
- Distribution de l'attention accordée à un personnage au fil du récit

Pour une détection automatique de l'espace personnage

Opérationnalisation de l'espace personnage

- Personnage : au delà de l'entité nommée
- Prise en compte de l'agentivité
- Détection de la coréférence - ensemble des mentions du personnage

Un algorithme de TAL pour la littérature ?

- Partie Française d'un projet multilingue
- Annotation sur un corpus de romans du XIX^e et XX^e siècle
- Détection d'entités (PER, FAC, TIME, ORG, LOC)
- Clustering d'entités
- Résolution de la coréférence pour chaque cluster

fr-BookNLP : Résolution de la coréférence

T1435 Mon frère , T10 m' embrassa tendrement, mais
T1435 il ne T10 me parla point, de sorte que T10 j' eus tout
le loisir, dont T10 j' avais besoin, pour rêver à T10 mon
infortune. T10 J' y trouvai d'abord tant d'obscurité que T10 je
ne voyais pas de jour à la moindre conjecture. T10 J' étais
trahi cruellement. Mais par qui ? T1462 Tiberge fut le premier

Figure 1 : Chaînes de coréférence dans Manon Lescaut

Résolution de la coréférence :

Détection des personnages :

	précision	rappel	F_1
PER	85.0	92.1	88,4

Métriques	F_1	
<i>MUC</i>	88,0	
<i>B³</i>	69,2	<i>Average 76.4</i>
<i>CEAF_e</i>	71.8	

Cas d'étude : Manon Lescaut

Manon Lescaut : Un récit test pour Booknlp

- Diversité des personnages (nombreux personnages secondaires)
- Discontinuité des personnages
- Pluralité des dénominations des personnages

Named characters

1436 Manon (256)/Lescaut (47)/ma chère Manon (5)/mademoiselle Manon (3)/Manon Lescaut (3)/Lescaut et moi (2)/Lescaut et Manon (2)/l' infidèle Manon (1)/belle Manon (1)/Chère Manon (1)/ta Manon (1)/la pauvre Manon (1)/La pauvre Manon (1)/la perfide Manon (1)/Inconstante Manon (1)/la malheureuse Manon (1)/chère Manon (1)

Figure 2 : Entités associées au nom propre « Manon »

Détecter la coréférence au fil du récit

- Inférences de fr-BookNLP sur tout le récit
- Fenêtre roulante de 1 000 tokens - Signal de l'espace personnage
- Intermittence des personnages secondaires ?

Tiberge - Baseline

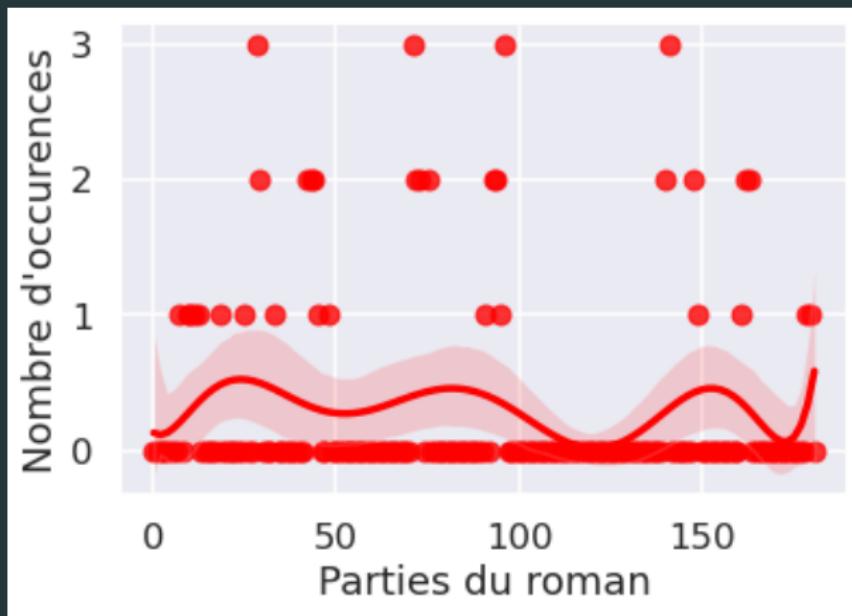


Figure 3 : Occurrences du token Tiberge au fil du roman

Tiberge

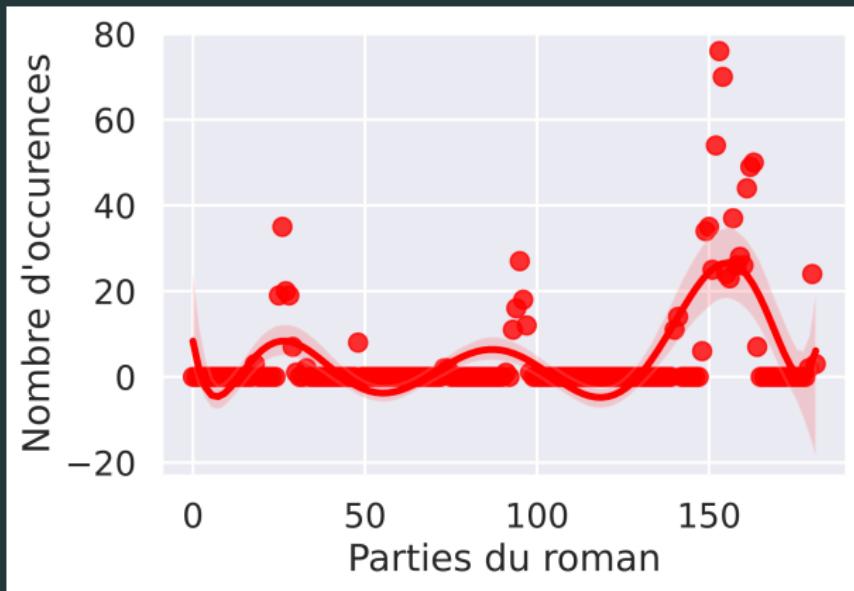


Figure 4 : Occurrences de la coréférence de Tiberge au fil du roman

Le lieutenant de police

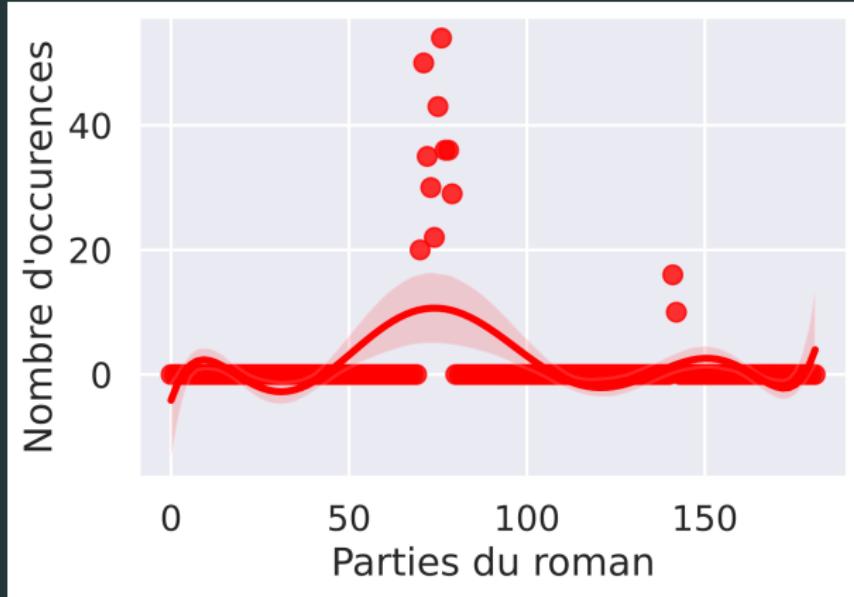


Figure 5 : Occurrences de la coréférence du lieutenant de police au fil du roman

Limites

Limites

- Doublons ? post-traitement
- Démographie des personnages ? (Lavocat, 2020) [3]
- Métalepse vs Personnage - Un seul niveau diégétique

Espace personnage comme élément clé du récit

- Impact des personnages sur le récit (Greimas, 1966) [4]
- Lier les entités entre elles (D'où vient un personnage ?, où va-t-il ? Quand ?)
- Construire le système des personnages - Addition des espaces personnage du récit

Études multi-échelle en lecture +- distante :

- Caractériser un roman
- Caractériser le style d'un auteur
- Caractériser un sous-genre romanesque

Questions?

Projet à Berkeley :

- [*https://github.com/dbamman/litbank*](https://github.com/dbamman/litbank)
- [*https://github.com/booknlp/booknlp*](https://github.com/booknlp/booknlp)

Partie française développée au Lattice :

- [*https://www.lattice.cnrs.fr/projets/booknlp/*](https://www.lattice.cnrs.fr/projets/booknlp/)
- [*https://github.com/lattice-8094/fr-litbank/*](https://github.com/lattice-8094/fr-litbank/)

Bibliography i

-  Alex Woloch.
The One vs. the Many.
Princeton University Press, 2003.
-  Roland Barthes.
S/Z.
Number 70 in Points Essais. Éditions du Seuil, 1996.
-  Françoise Lavocat.
L'étude des populations fictives comme objet et le « style démographique » comme nouveau concept narratologique.
Atelier de théorie littéraire, 2020.
-  Algirdas Julien Greimas.
Sémantique Structurale : Recherche de Méthode.
Larousse, Paris, 1966.