

# Analyse computationnelle de la naissance du roman policier

---

Jean Barré

20 Mars 2025

École normale supérieure - Université PSL

LaTTiCe lab

# Introduction

## Approches des Genres Littéraires

- Tension formaliste vs. contextuelle.
- Approche computationnelle des genres littéraires.
- Résultats (Underwood, 2019) : Stabilité du genre, validant les approches transcendentales et formalistes.

## Pourquoi le roman policier ?

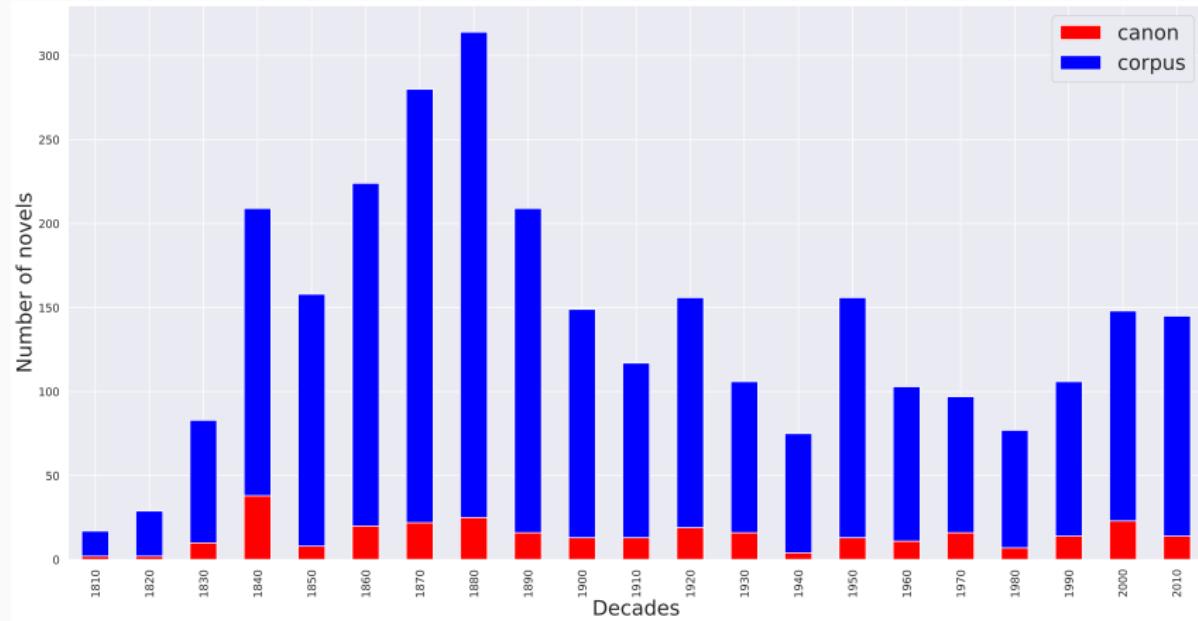
- Genre à forte identité architextuelle et intertextuelle.
- Longue tradition éditoriale depuis la collection du masque (Pigasse, 1927)
- Corpus abondant et clairement balisé, idéal pour une analyse computationnelle des évolutions génériques.

## De la marginalité à la reconnaissance

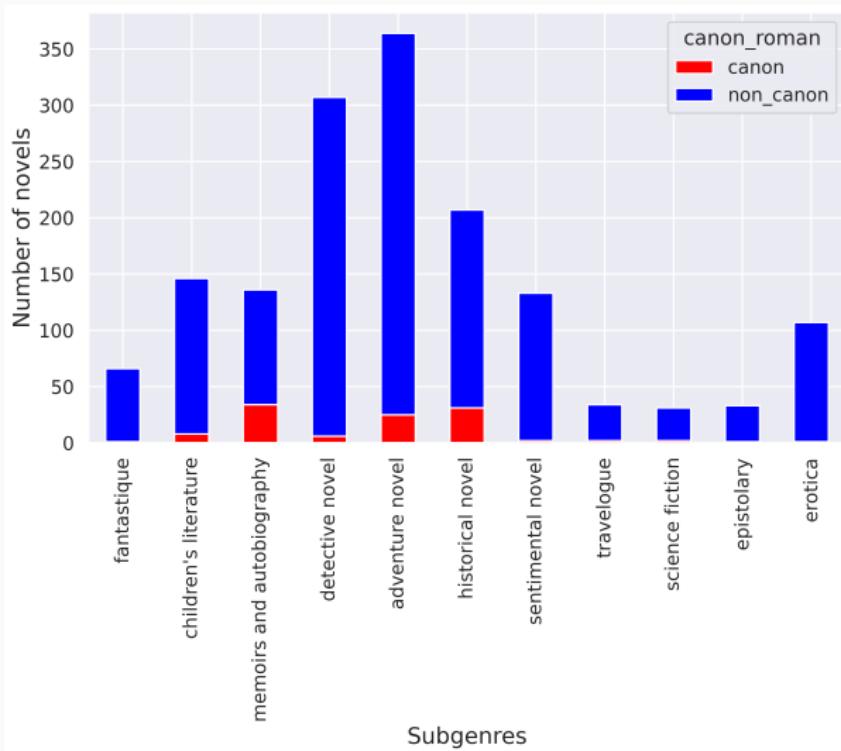
- Apparition progressive à partir de la fin du XIX<sup>e</sup> siècle, issue du roman feuilleton (notamment Gaboriau).
- Structuration narrative spécifique : « un récit consacré avant tout à la découverte méthodique et graduelle, par des moyens rationnels, des circonstances exactes d'un crime mystérieux » (Messac, 2011).
- Canonisation symbolique : reconnaissance critique de Simenon par André Gide (1939), pléiade,...

## Logique architextuelle (Genette, 1986 ; Letourneau, 2017)

- Réseau interconnecté de textes partageant des conventions communes.
- Chaque auteur enrichit et diversifie l'architexte policier : Poe (énigme), Gaboriau (enquête méthodique), Leroux (Suspense), Christie (whodunit), Simenon (psychologie).
- Renouvellement continu des attentes du lecteur par une logique d'intertextualité sérielle (Maigret, Rouletabille).



**Figure 1 – Distribution temporelle du corpus**



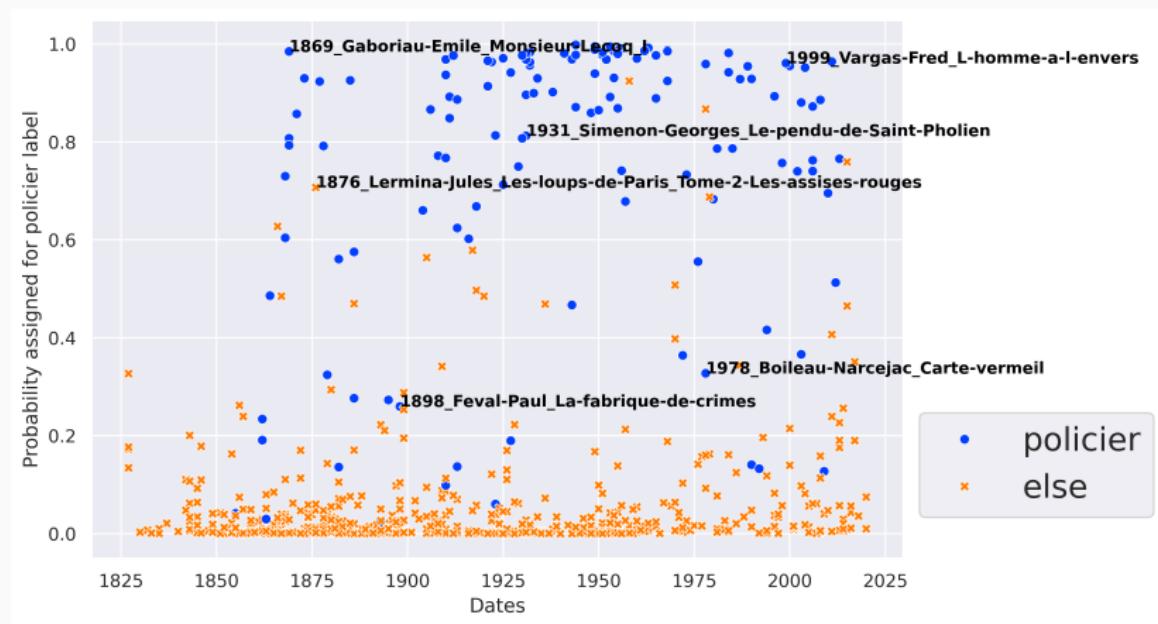
**Figure 2 – Les genres littéraires du corpus**

## En pratique - Exemple du roman policier

### Méthode - 3 étapes

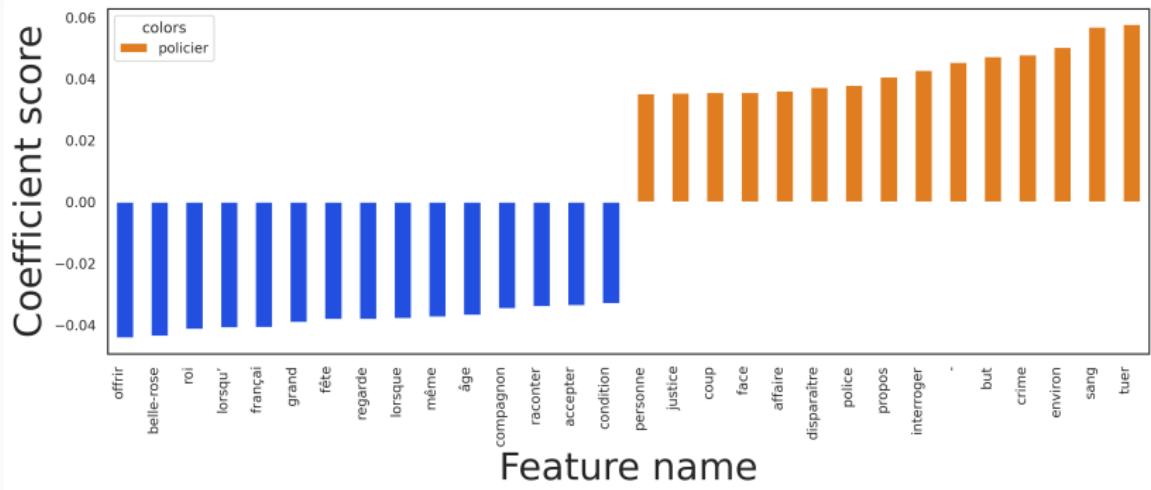
- Extraire des informations textuelles
- Apprentissage Machine - Modélisation -> SVM
- Analyses / Interprétations -> Humain

# Classification Automatique du Roman Policier



**Figure 3 – Classification automatique du roman policier francophone**

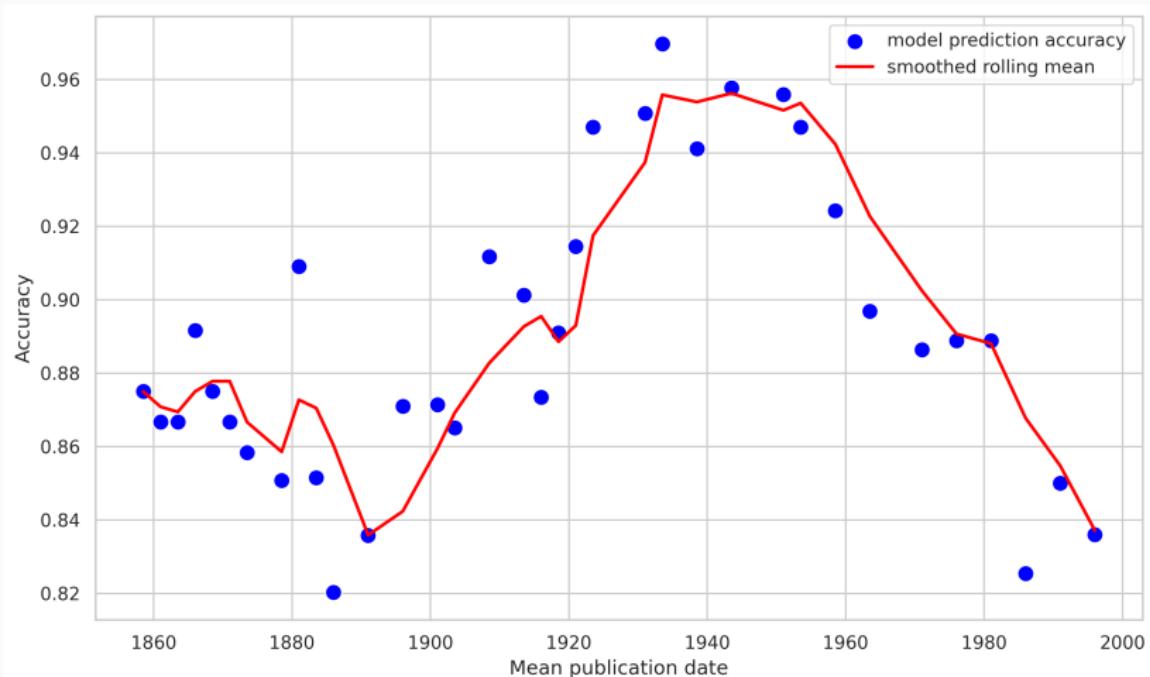
# Caractéristiques discriminantes



**Figure 4 – Caractéristiques discriminantes pour prédire le policier**

- **Histoire anhistorique** : Risque de projeter notre définition contemporaine du roman policier sur des textes anciens.
- **Mutation historique du genre** :
  - Deux modèles SVM entraînés avant et après WWII donnent des résultats très différents.
  - Exemple : un modèle entraîné sur les romans policiers-feuilletons (Gaboriau, Boisgobey) avant 1939 obtient 88.4% de précision, et reste performant (75%) sur les romans d'après-guerre.
  - À l'inverse, un modèle entraîné sur les romans postérieurs à 1939 obtient une précision élevée (88.1%) et chute à (55%) sur les œuvres antérieures.
  - Cette asymétrie montre que le proto roman policier (1860-1939) est différent et encore en cours de constitution par rapport à sa forme moderne.
- **Question ouverte** : Comment le roman policier s'est-il progressivement démarqué du roman-feuilleton pour devenir un genre autonome ?

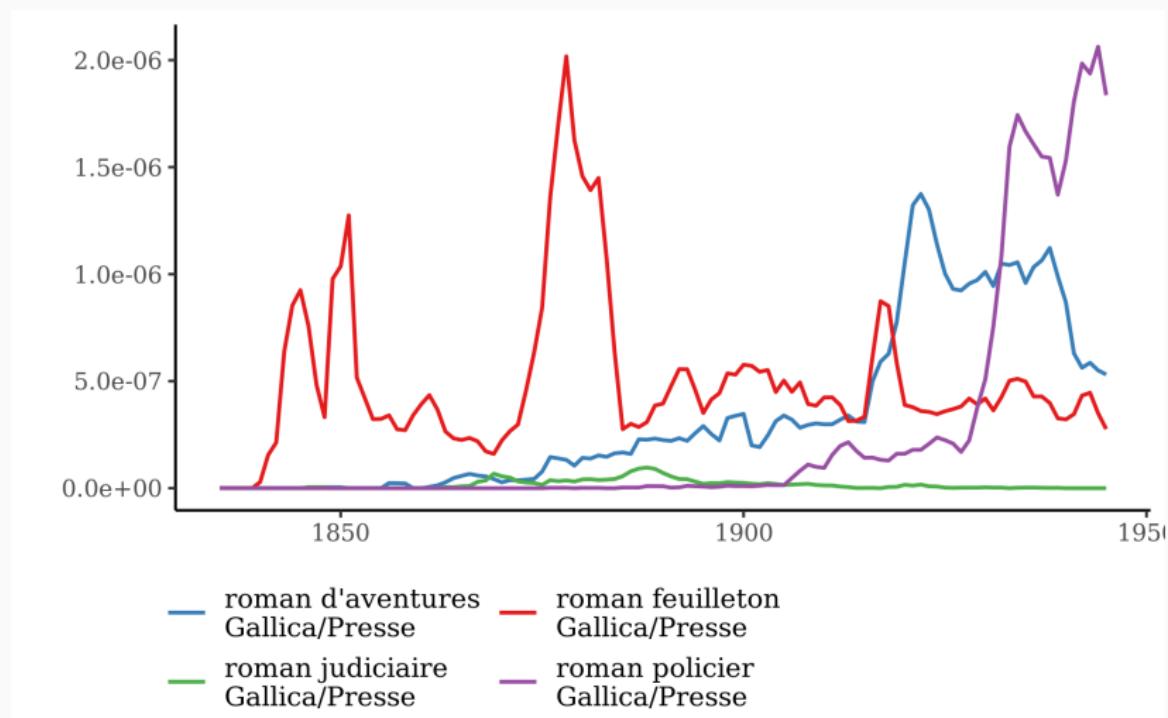
# Évolution de la Prédiction du Genre



**Figure 5 – Évolution de la prédiction du genre tous les 25 ans**

## Limites 2 : Apparition Tardive du Terme

- Le terme *roman policier* n'apparaît que très tard historiquement.



## Question Centrale

- Comment apparaît le genre et comment se structure-t-il au fil du temps ?

## Hypothèse du Génie ?

- Inspirations : Poe, Balzac, Gaboriau.
- « Balzac, pionnier en tout, a écrit *Une ténébreuse affaire*, le premier roman policier » (Fortassier, 1955)
- « Poe est l'inventeur du roman policier » (Borges, 1936)
- « Émile Gaboriau est le père du roman policier » (Messac, 1929)

## Approche Cumulative de la Naissance des Genres

- La répétition et l'accumulation des pratiques discursives créées le genre.
- (Rieder, 2012)  
*Etudier les débuts du genre ne consiste pas du tout en une recherche de points d'origine mais plutôt en l'observation d'une accrétion de répétition, échos, imitations, identifications et distinctions qui témoignent d'une conscience émergente d'un réseau conventionnel de ressemblances*

# Approche par Ressemblance de Famille

## (Wittgenstein, 1953)

*Prenez, par exemple, les activités que nous appelons « jeux », [...] vous ne verrez pas quelque chose de commun à tous, mais des ressemblances, des affinités [...]. Nous voyons un réseau complexe de ressemblances qui se chevauchent et se croisent [...]. Ces ressemblances sont comme celles entre les membres d'une famille.*

## Application aux genres littéraires

- Le genre n'est pas une catégorie rigide aux frontières fixes.
- Il forme un réseau dynamique de relations entre textes.
- Les similarités varient en intensité et en nature d'une œuvre à l'autre.
- Aucun trait unique ne définit entièrement le genre, mais plusieurs traits se croisent et se chevauchent.

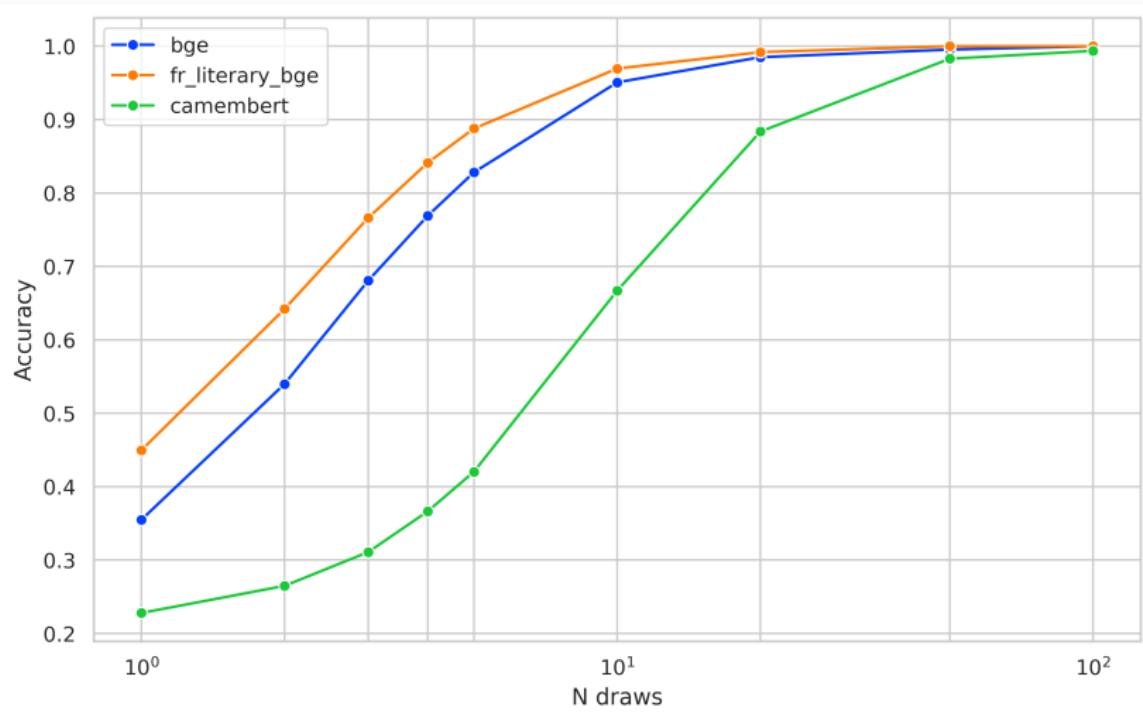
# Ressemblance de Famille dans le Roman Policier

- Éléments déterminants communs : présence d'un crime, enquête méthodique, rôle central du détective.
- Aucun roman policier ne possède tous les traits du genre à la fois, mais chaque œuvre partage un ensemble variable de caractéristiques génériques.
- **Personnages récurrents** : Rouletabille (Gaston Leroux), Tiraclair et Lecoq (Émile Gaboriau), Maigret (Simenon).
- **Variations stylistiques et narratives :**
  - *L'Affaire Lerouge* (Gaboriau, 1866)
  - *Le Mystère de la Chambre Jaune* (Leroux, 1907) reprend la logique d'énigme et introduit une forte dimension ludique (chambre close).
  - Les romans de *Maigret* (Simenon, à partir de 1931) renouvellent l'approche, tout en conservant l'enquêteur central
- Chaque roman participe à une définition mouvante du genre, à travers un réseau de ressemblances partielles.

## Les embeddings et la similarité cosinus comme indicateur de l'intertextualité

- Modèle d'encodage contextuel de pointe comme outil de représentation textuelle
- Ajustement fin du modèle BGE-M3-Embedding sur le langage littéraire français
- **Requête** : un paragraphe; **Pos** : 5 paragraphes suivants; **Neg** : 5 paragraphes aléatoires
- **Hypothèses** : les embeddings capturent des notions de style individuel, mais aussi de similarité thématique

# Vérification de la validité du modèle



**Figure 7 – Évaluation de l'encoder**

# Réseau de Ressemblance - Échelle d'un roman

version HTML



**Figure 8 – Réseau de ressemblance pour le Mystère de la chambre jaune,(Leroux, 1908)**

## Exemple de similarité- Morceau 41

*mask regarda le ciel , le trouva à sa convenance et , sans doute , à la mienne , car il me prit sous le bras et me dit : « Allons! ... J' ai besoin de marcher . - Eh bien ! lui demandai -je . Ça se débrouille ? ... - Oh ! fit -il , oh ! Il n' y a rien de débrouillé du tout ! ... C' est encore plus embrouillé qu' avant ! Il est vrai que j' ai une idée ... - Dites - la . - Oh ! je ne peux rien dire pour le moment ... Mon idée est une question de vie ou de mort pour deux personnes au moins ... - Croyez -vous à des complices ? - Je n' y crois pas ... »*

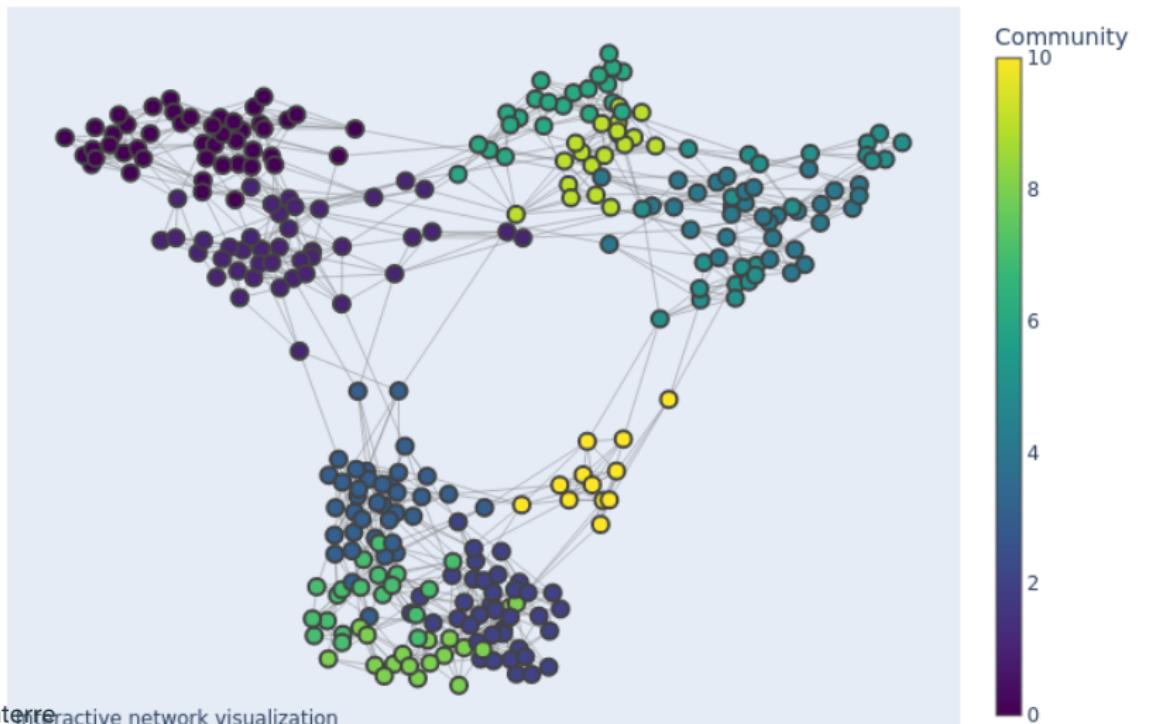
## Exemple de similarité- Morceau 176

*J' en eus la preuve quand , en descendant la côte d' mask , il me dit : « mask mask est arrivé mask mask avant moi; il a commencé son enquête avant moi; il a eu le temps de savoir des choses que je ne sais pas et a pu trouver des choses que je ne sais pas ... Où a -t -il trouvé cette canne -là ? ... » Et il ajouta : « Il est probable que son soupçon - plus que son soupçon , son raisonnement - qui va aussi directement à mask mask , doit être servi par quelque chose de palpable qu' il palpe , lui , et que je ne palpe pas , moi ... Serait -ce cette canne ? ... Où diable a -t -il pu trouver cette canne -là ? ... »*

# Réseau de Ressemblance - Échelle du genre

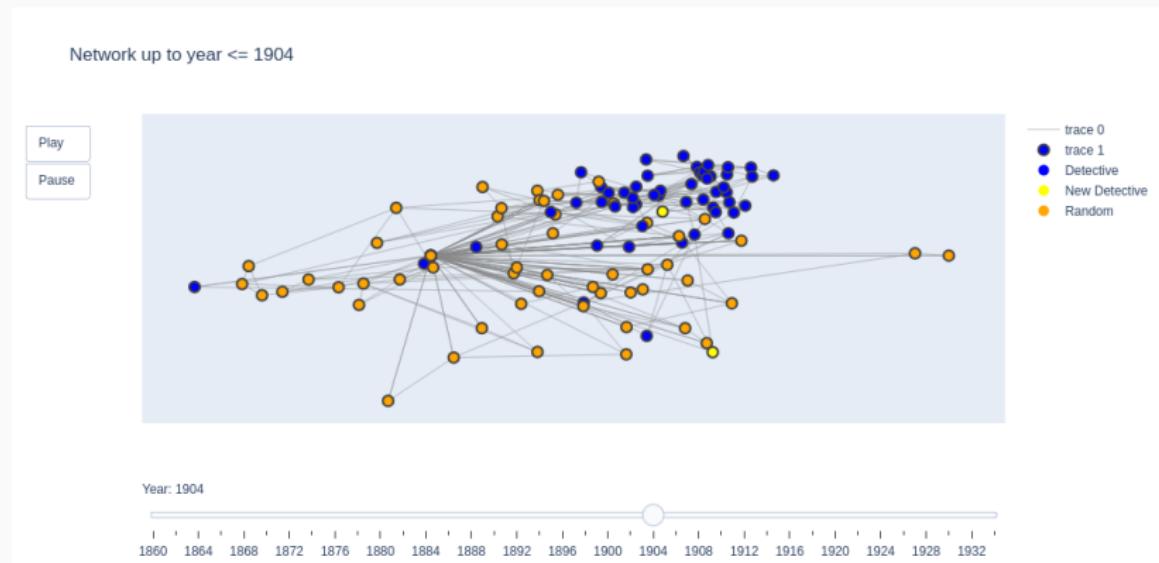
version HTML

Network of Novel Similarities (Top-5 Nearest Neighbors & Louvain Communities)



# Réseau de Ressemblance - Échelle du corpus

version HTML



**Figure 10** – Approche cumulative du réseau de ressemblance du roman policier

**Le roman policier se démarque du roman-feuilleton en trois points :**

1. **Persistance du thème criminel** : plus cohérent et central que dans le roman-feuilleton.
2. **Centralité du détective** : l'enquêteur passe d'un rôle secondaire à un rôle principal, moteur de la résolution du crime.
3. **Structure narrative centrée sur l'investigation** : récit organisé autour de la progression de l'enquête, allant des indices connus vers l'inconnu, ponctué de fausses pistes et de rebondissements.

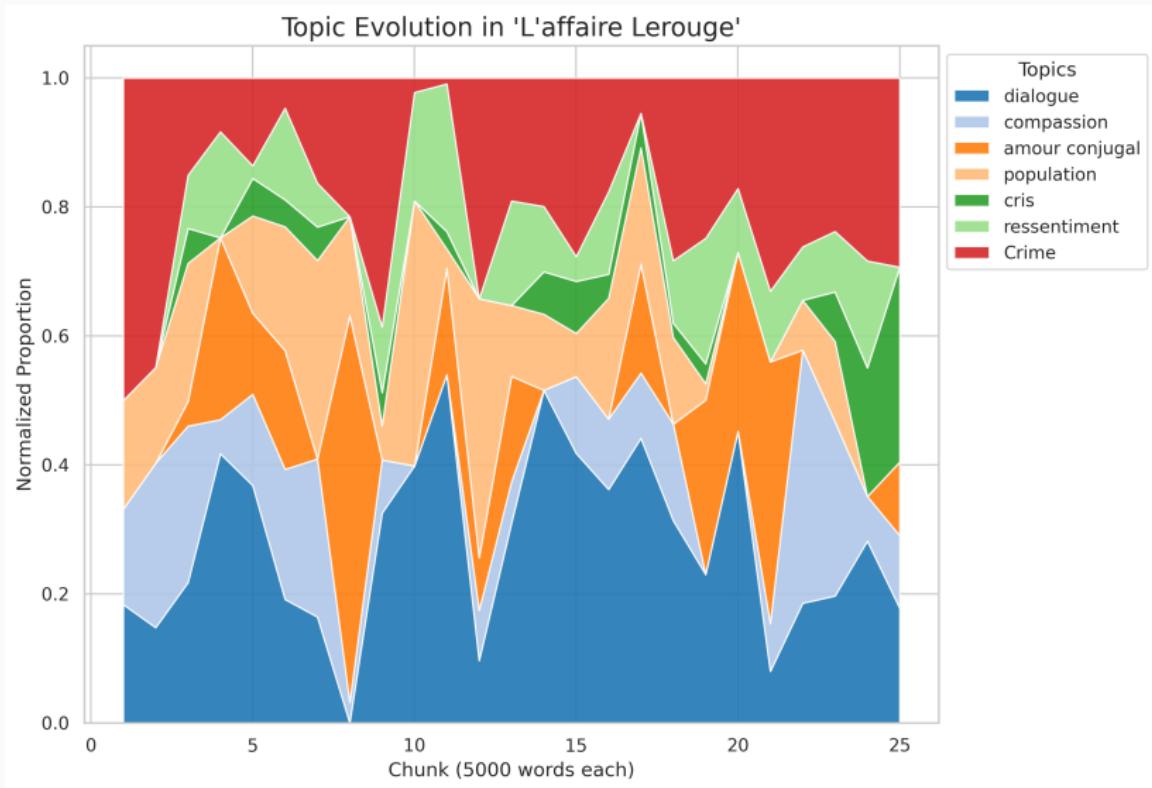
## Principe de l'analyse thématique LDA

- **Latent Dirichlet Allocation (LDA)** : méthode probabiliste d'extraction de thèmes dans un corpus.
- Les documents sont considérés comme des **distributions de thèmes**, eux-mêmes distributions de mots.

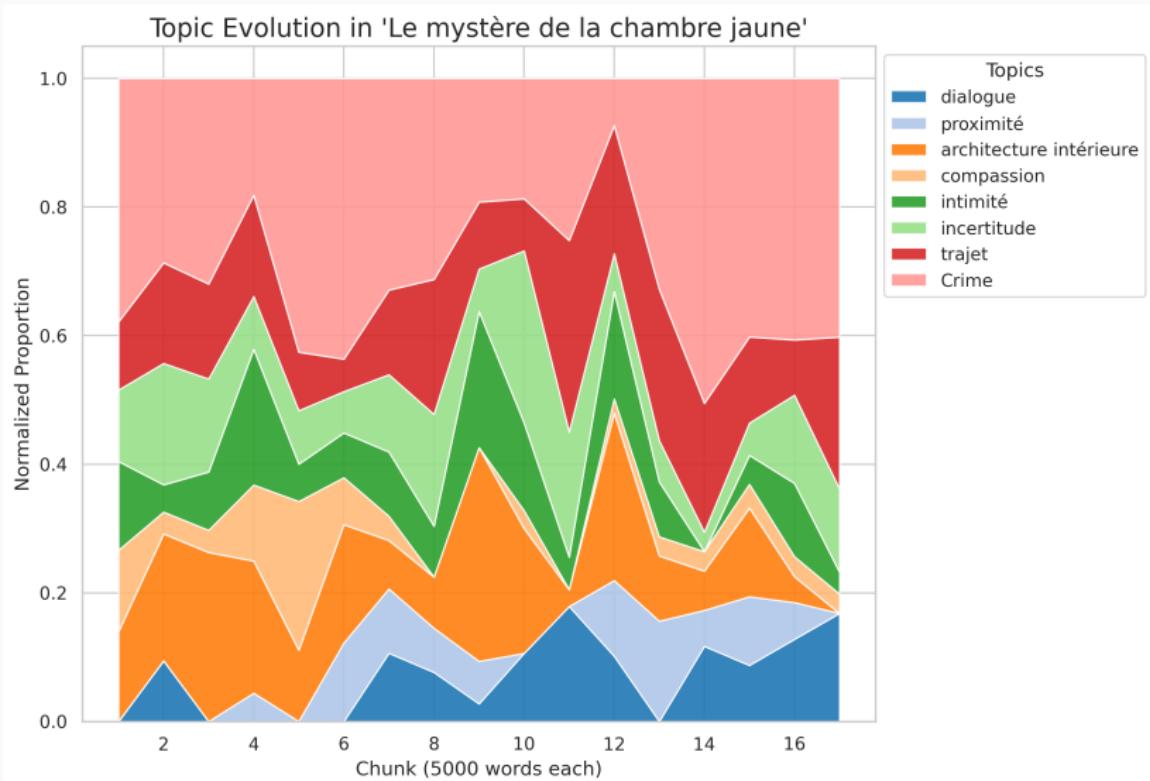
## Prétraitements appliqués au corpus

- **Lemmatisation** des mots du corpus.
- Filtrage par **catégories grammaticales** (noms, verbes, adjectifs, adverbes).
- Élimination des **noms propres**, des mots rares et des mots outils (stopwords).
- Segmentation des romans en **chunks** de 5000 mots pour étudier l'évolution intra-textuelle.
- Récupération de 100 thèmes dans le corpus chapitres, et de leur proportion dans chaque roman / morceaux de roman

# Évolution du Thème Criminel sur l'Affaire Lerouge



# Évolution du Thème dans le Mystère de la Chambre Jaune



IA Nanterre **Figure 12** – Évolution du thème dans le mystère de la chambre jaune

# Évolution du Thème dans le 120 rue de la Gare

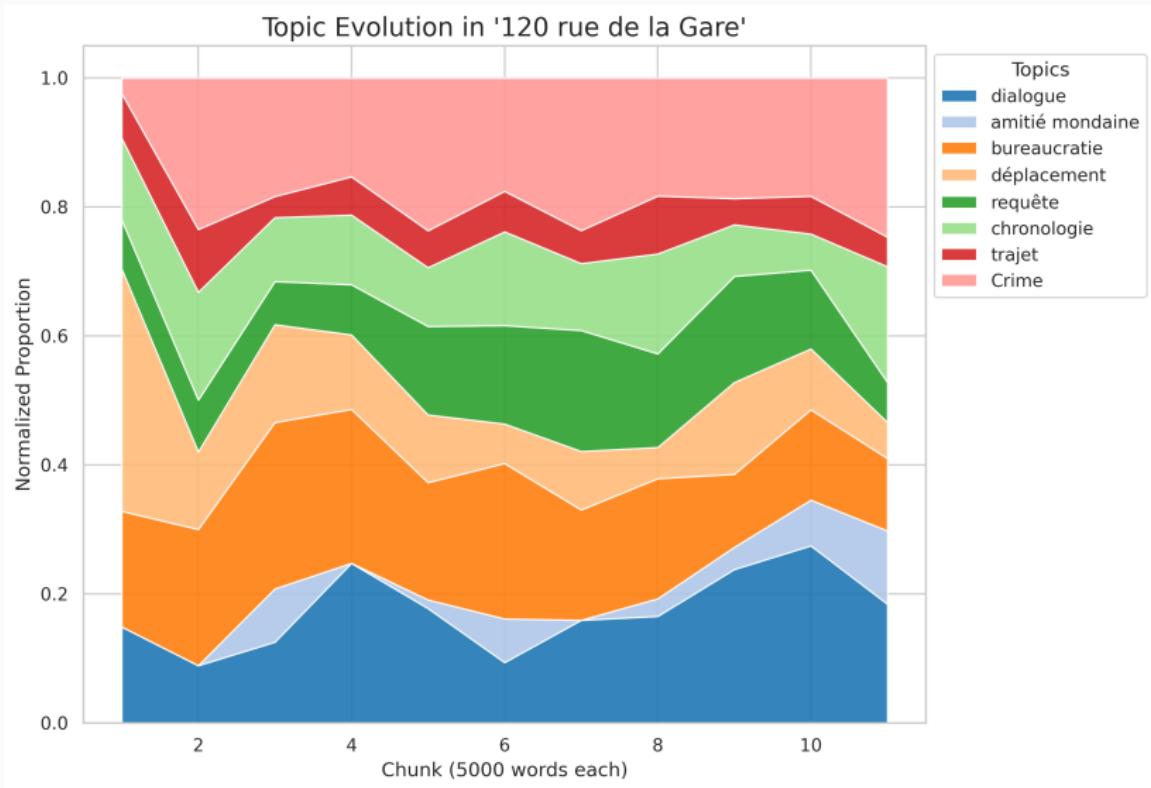


Figure 13 – Évolution du thème dans 120 rue de la Gare

# Calcul de la Cohérence du Thème Criminel

## Objectif de la mesure

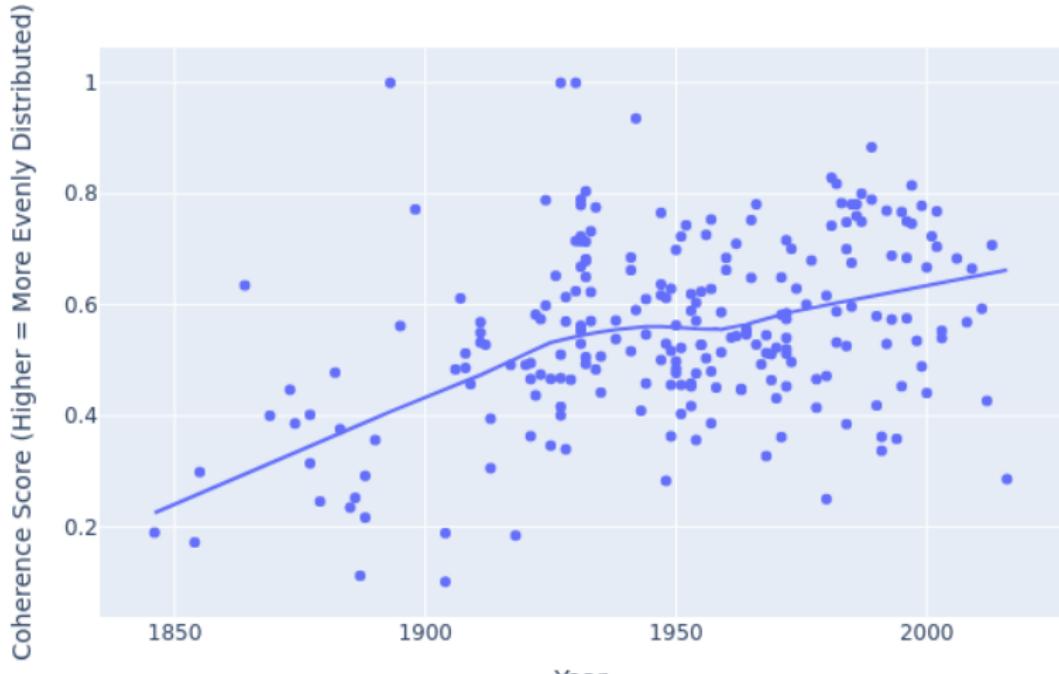
- Mesurer l'**homogénéité de la distribution** du thème criminel au fil du roman.
- Identifier si le thème criminel apparaît de manière régulière ou épisodique.

## Méthode : Score de Cohérence (Mean/Max)

$$\text{Cohérence} = \frac{\mu}{x_{\max}}$$

- $\mu = \frac{1}{N} \sum_{i=1}^N x_i$  : moyenne des proportions du thème criminel sur les  $N$  chunks.
- $x_{\max} = \max(x_1, x_2, \dots, x_N)$  : proportion maximale atteinte par le thème criminel dans un chunk du roman.
- $x_i$  : proportion du thème criminel dans le chunk  $i$ .

# Persistance du Thème Criminel

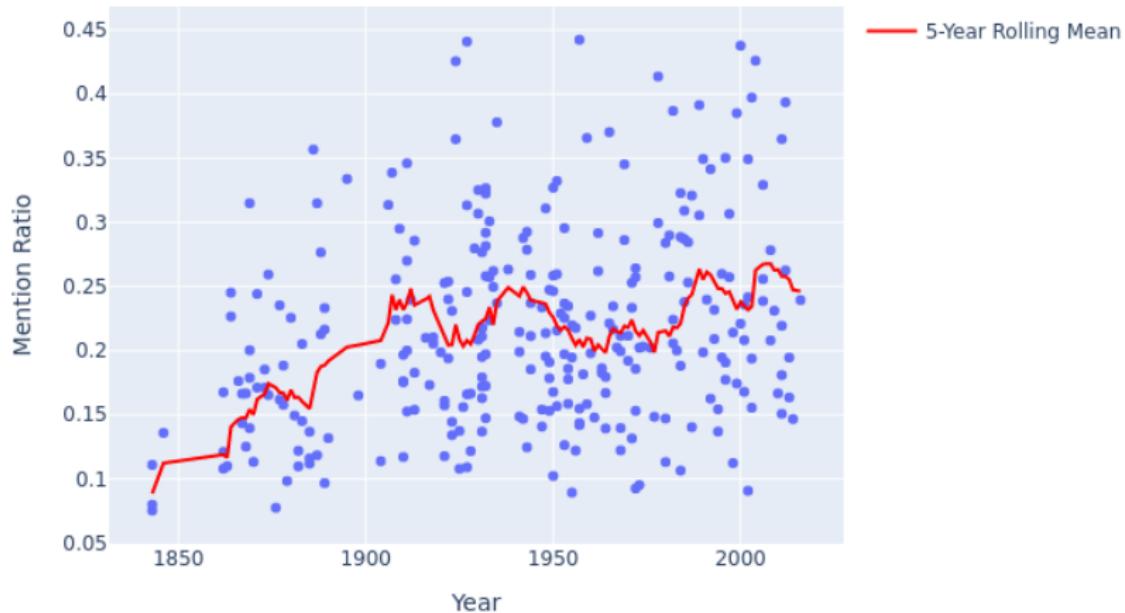


**Figure 14 – Persistance du thème criminel au fil du temps**

## BookNLP-fr en Action

- Extraction automatique des chaînes de coréférence à partir des données BookNLP-fr.
- Deux approches de sélection du détective :
  - **Option 1 :** Choix du personnage principal (identifiant 0) et utilisation de son mention\_ratio.
  - **Option 2 :** Vérification des deux premiers personnages (identifiants 0 et 1) pour identifier un détective en évaluant leurs mentions communes à l'aide d'une liste de termes adaptés (par exemple : "agent, sûreté, limier, homme de loi, juge d'instruction, brigadier, magistrat, ...").
- Filtrage des valeurs aberrantes : élimination des cas dont le mention\_ratio excède 0,45.
- Visualisation temporelle de la domination du détective, avec ajout d'une moyenne roulante sur 5 ans pour suivre l'évolution de sa présence dans l'intrigue.

# Domination progressive du Personnage DéTECTive



**Figure 15 –** Évolution de la domination du personnage détective dans les romans policiers

# Analyse de l'intrigue dans le roman policier

- Identification des mécanismes de suspense, du crime à la résolution
- Axes de travail en cours :
  - Détection des séquences narratives clés : scène de crime, découverte, tentative de résolution et révélation finale
  - Étude de l'agencement narratif comme indicateur spécifique du roman policier

# Merci !

**Questions ?**

jean.barre@ens.psl.eu

<https://crazyjeannot.github.io/>