



# Approches Computationnelles des Genres Littéraires

Introduction aux CLS - ENS

---

Jean Barré

14 mars 2024

Translitterae - ENS-PSL - Lattice

## Les genres littéraires

- Concept vieux comme la théorie littéraire  
Triade Lyrique / Épique / Dramatique
- Historiquement - Pratique bibliographique - BNF
- Objet d'étude de plusieurs disciplines : Rhétorique - Sociologie - Littérature
- Motifs textuels de genre ?

# Caractéristiques des genres littéraires

- **Éléments internes** -

Thèmes, Rythme de l'intrigue, cadre spatiotemporel, lexique

- **Éléments externes** -

Rôle de la réception - Horizon d'attente - lecteurs.

Ou rôle des institutions, éditeurs -> construction de produits culturels / littérature de masse

- **Dynamique naïve** -  
Cristallisation - Apogée - Transformation/Hybridation
  
- **Cycles du changements** -  
Phénomène Générationnel - ou Auctorial

## Méthode - 3 étapes

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

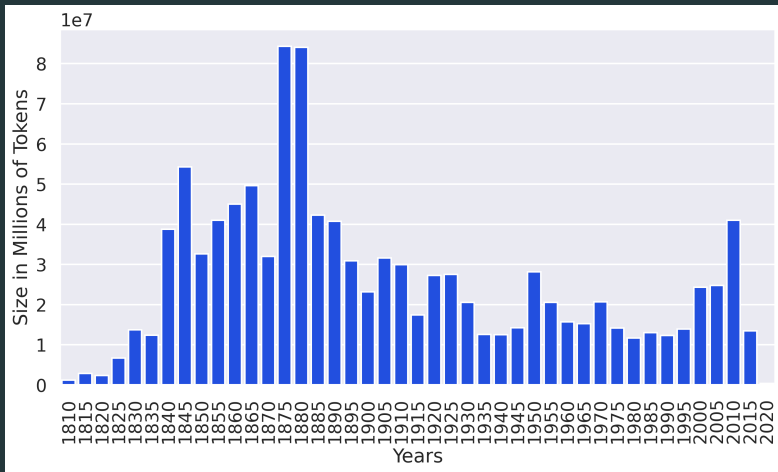


Figure 1 - Distribution temporelle des mots dans le corpus

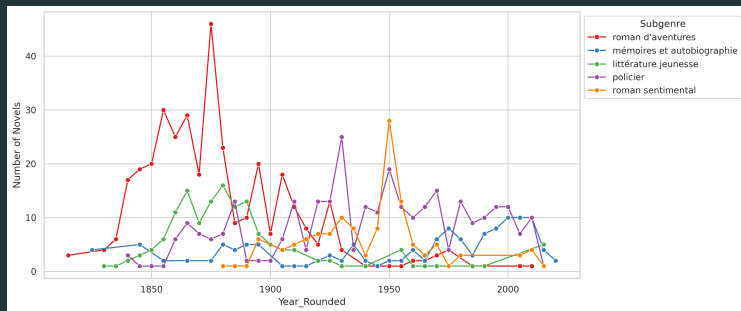


Figure 2 – Distribution temporelle des genres littéraires dans le corpus

## Un ensemble de caractéristiques textuelles

- **Lexique** : N mots les plus fréquents. Sac de mots.
- **Caractérisation** : Adjectifs et verbes qui caractérisent le personnage dans le récit.
- **Chronotope** - fr-BookNLP : LOC + FAC + VEH + TIME.



## Algorithme de TAL à l'échelle des romans

- NER / Reconnaissance d'entités nommées (PER, FAC, TIME, ORG, LOC, ...)
- Regroupe les dénominations des personnages :  
Arsène Lupin, Monsieur Lupin, Lupin, le gentleman cambrioleur  
-> ARSENE\_LUPIN
- Résolution de la co-référence : pronoms, noms, .. : Arsène, son ami, ce monsieur

# Classification automatique du roman policier

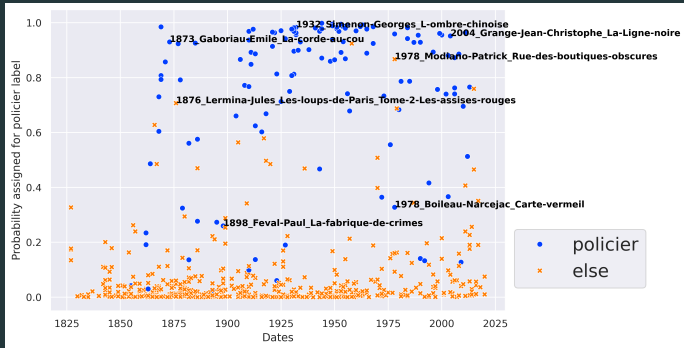


Figure 3 – Probabilité prédite d'appartenir au roman Policier

# Quelles caractéristiques discriminantes ?

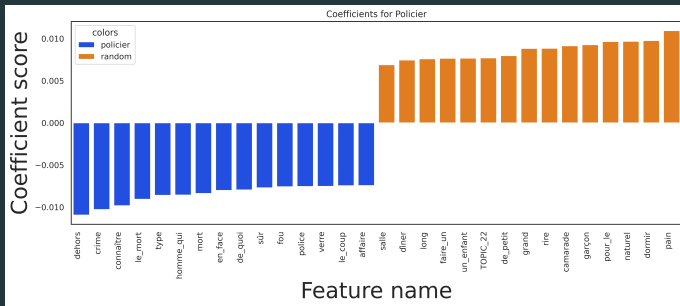
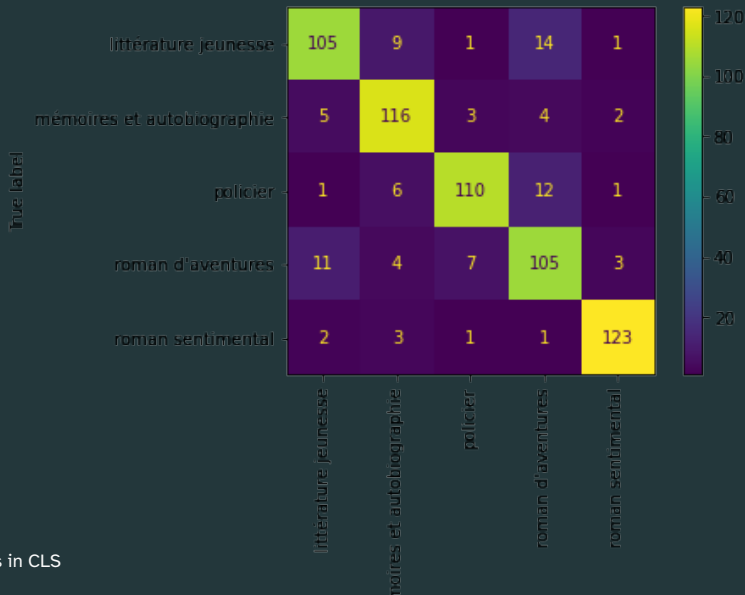
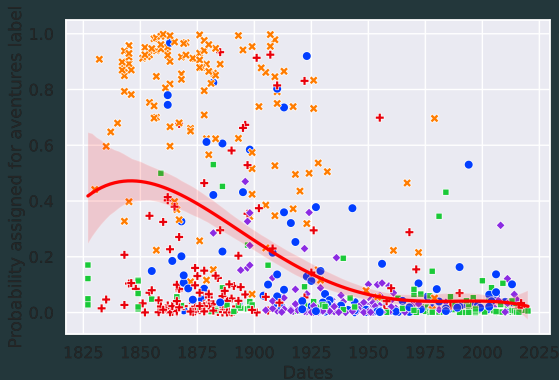


Figure 4 – Coefficients discriminants pour le modèle

# Généralisation de l'approche au sous-genres du corpus Chapitres : 86% d'efficacité

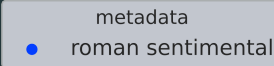
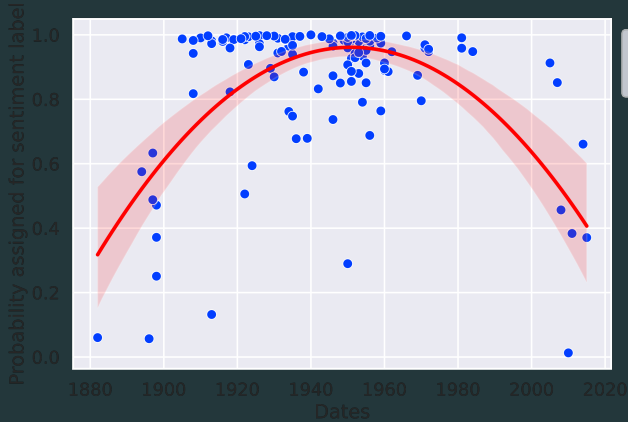


# Probabilité du roman d'aventures à travers le temps



- metadata
- policier
  - ✕ roman d'aventures
  - mémoires et autobiographies
  - ✕ littérature jeunesse
  - ◆ roman sentimental

# Probabilité du roman sentimental à travers le temps



1. Les genres sont un agrégats de styles d'auteurs / d'époque
2. Le modèle "triche" sur les mots les plus fréquents pour faire de l'attribution d'auteur ?
3. Quelles solutions ?

## Évaluation des modèles sans le biais d'auteur - BoW

	Precision	Recall	F1-score	Support	Accuracy
Children	0.75	0.75	0.75	130	
Memoirs	0.79	0.82	0.80	130	
Detective	0.67	0.68	0.67	130	
Adventure	0.60	0.65	0.62	130	
Romance	0.84	0.72	0.80	130	
Full Dataset				650	<b>0.72</b>

**Table 1** - Classification Report for BoW



# Évaluation des modèles sans le biais d'auteur - BoW

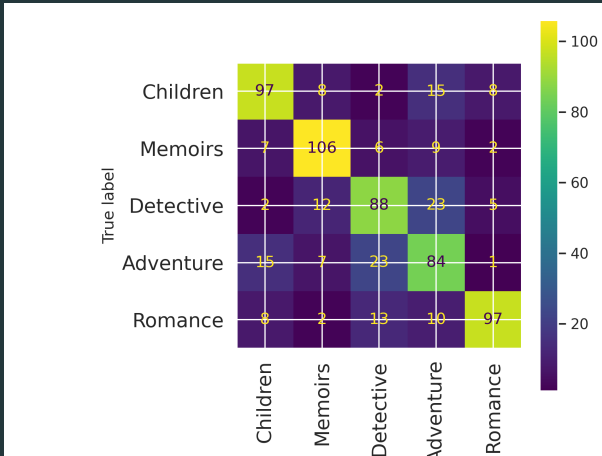


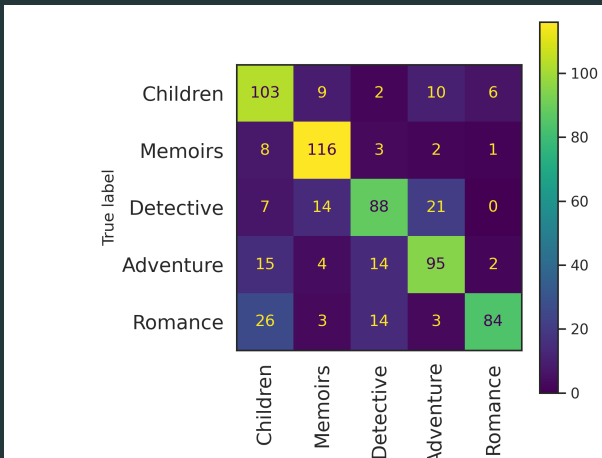
Figure 5 - Confusion Matrix for BoW

## Évaluation des modèles sans le biais d'auteur - BookNLP

	Precision	Recall	F1-score	Support	Accuracy
Children	0.65	0.79	0.71	130	
Memoirs	0.78	0.89	0.84	130	
Detective	0.68	0.70	0.70	130	
Adventure	0.73	0.73	0.73	130	
Romance	0.90	0.65	0.75	130	
Full Dataset				650	<b>0.75</b>

**Table 2** – Classification Report for BookNLP-fr features

# Évaluation des modèles sans le biais d'auteur - BookNLP

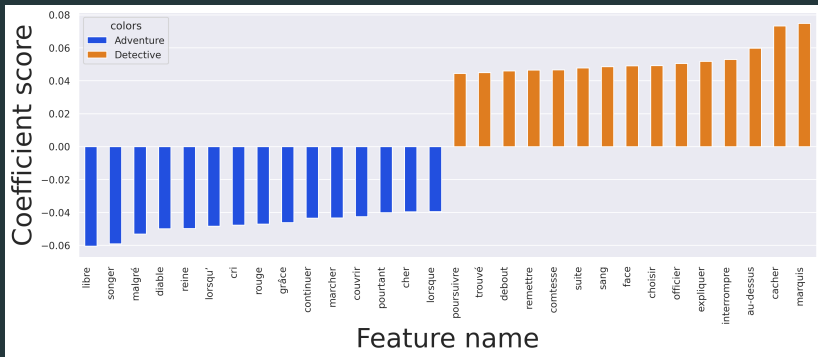


**Figure 6** – Confusion Matrix for BookNLP-fr features

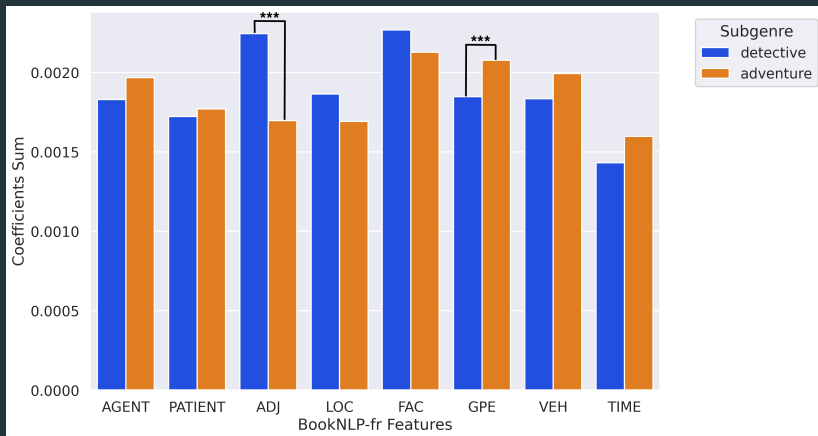
## Evaluation des features BookNLP pour la classification

BookNLP-fr features	Accuracy
LOC	0.45
FAC	0.59
VEH	0.42
GPE	0.47
TIME	0.50
PATIENT	0.52
AGENT	0.62
ADJ	0.50
Baseline	0.2

**Table 3** – BookNLP-fr features accuracy.



**Figure 7** – BoW discriminant features for Adventure vs Detective classification.



**Figure 8** – BookNLP-fr discriminant features for Adventures vs Detective classification. '\*\*\*' meaning  $p < 0.001$ .

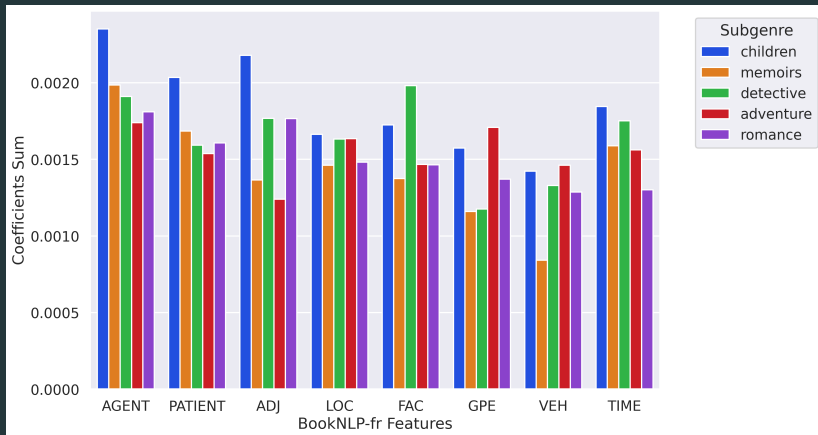


Figure 9 – BookNLP-fr discriminant features for the classification.

# Questions ?