

Modeling the Birth of a Literary Archetype: The Detective in French Fiction

From character extraction to genre periodization (1866 - 2020)

Jean Barré Olga Seminck Antoine Bourgois Thierry Poibeau
Culture Lab Kick Off

École normale supérieure - Université PSL

LaTTiCe (CNRS - UMR 8094)

Langues, Texte, Traitements Informatiques, Cognition

- **Linguistics:** Natural Language Processing
- **9 permanent members in the CultureLab**

Computational Linguistics for Digital Humanities

- **BookNLP-fr/Propp Project:** since 2020
 - Pipeline to extract automatically character representations from books
 - **Distant reading:** large-scale corpora, beyond the canon

Back to the Detectives...



Introduction

General aim: Modeling the construction and the evolution of the **Detective Archetype** in French literature using computational tools.

Research questions - The Detective Figure

- **RQ1 - Archetype unity:** Is it possible to detect an archetypal detective in 150 years of fiction?
- **RQ2 - Archetype emergence:** When does the detective figure evolve from a secondary character into the main character?
- **RQ3 - Semantic trajectory:** How does the detective archetype *change* across the evolution of the detective novel's subgenres?

Methods

Idea

Create a tool to recognize characters that are detectives.

→ **Detective Detector**

How?

1. We use the **Propp-fr** pipeline to identify the characters of a text.
 - mentions
 - coreference chains
 - attributes

Propp-fr : Output example

[#37] T131 J' avais amené T1482 cette jeune femme au bal de
T615 madame de Lanty . Comme T1482 elle venait pour la
première fois dans T356 cette maison , T131 je T1482 lui
pardonnai T1482 son rire étouffé ; mais T131 je T1482 lui fis
vivement T131 je ne sais quel signe impérieux qui T1482 la rendit
tout interdite et T1482 lui donna du respect pour T1482 son
voisin. T1482 Elle s'assit près de T131 moi . T992 Le vieillard ne

Figure 1: Coreference chains in Sarrasine (Balzac, 1830)

Methods

Idea

Create a tool to recognize characters that are detectives.

→ **Detective Detector**

How?

1. We use the **Propp-fr** pipeline to identify the characters of a text.
 - mentions
 - coreference chains
 - attributes
2. Train a classifier to recognize detectives with examples
 - 185 detectives
 - 419 non-detectives

Methods

Detective Detector... What for?

Identify all the detectives in a large text collection spanning over 150 years.

- Test if detective figures are ‘consistent’ enough to learn a representation that covers them all
- Study the emergence of the detective genre
- Study the semantic trajectory of the genre via the attributes

‘

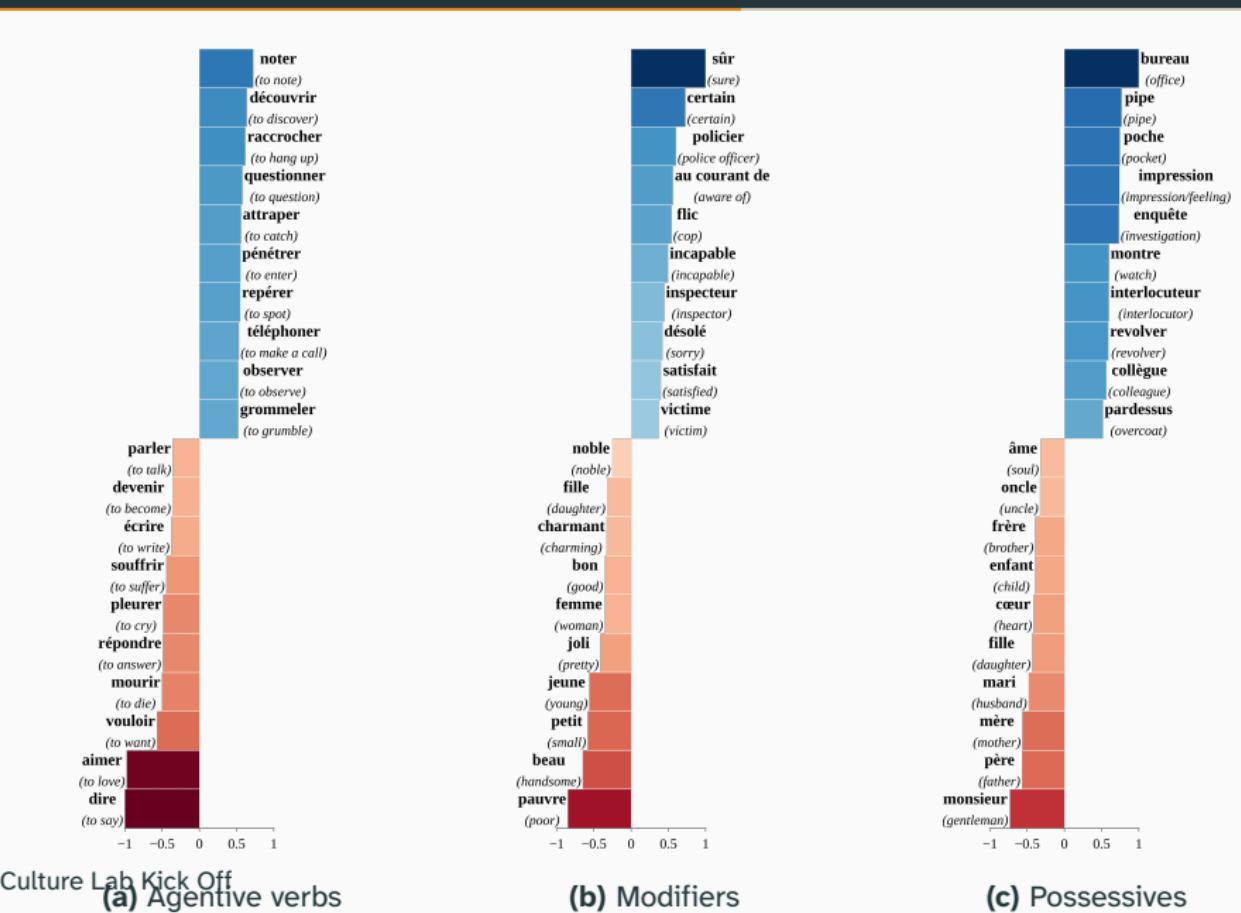
Results - Detective Detector Evaluation

RQ1 - Archetype unity: Is it possible to detect an archetypal detective in 150 years of fiction?

Model	B.Acc.	F1 (Det.)
BoW + LogReg	0.836	0.75
BoW + SVM	0.895	0.84
CamemBERT + LogReg	0.906	0.85
CamemBERT + SVM	0.923	0.887
LOGO (Author)	0.908	0.873
LOGO (10 yrs)	0.915	0.879
LOGO (50 yrs)	0.892	0.858

Table 1: Detector performance

Lexical specificity of the archetype



Culture Lab Kick Off
(a) Agentic verbs

(b) Modifiers

(c) Possessives

Large-scale inference: emergence and centrality

RQ2 - Archetype emergence: When does the detective figure evolve from a secondary character into the main character?

Narrative centrality

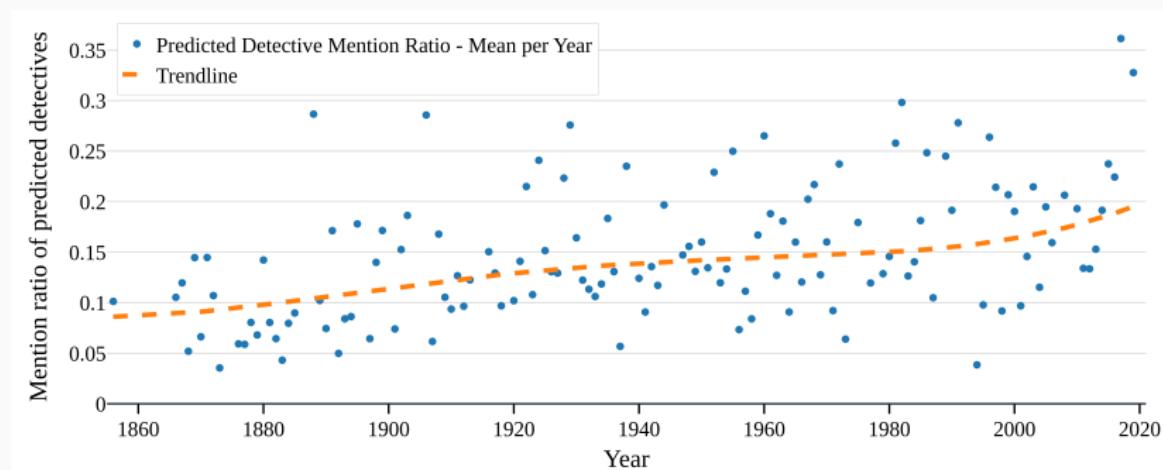


Figure 3: Mention-ratio -> growing narrative pivot role

Archetype Evolution

RQ3 - Semantic trajectory: How does the detective archetype *change* across the evolution of the detective novel's subgenres?

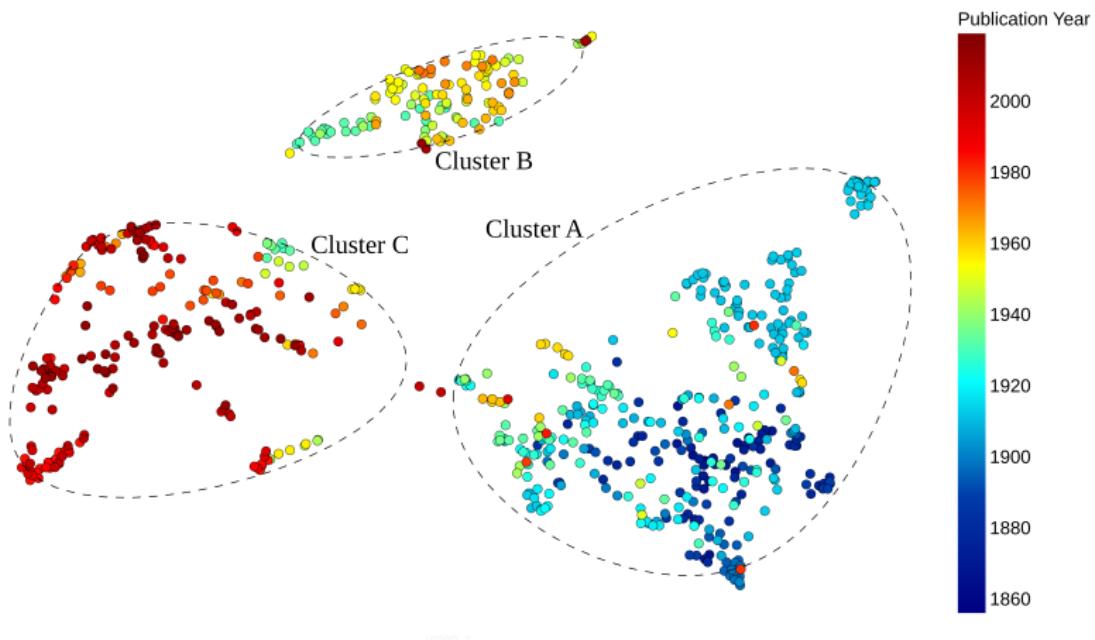


Figure 4: UMAP + k -means ($k=3$) on 713 detectives

Table Lexical specificities per period

A Classical (late 19th/early 20th)	B Empathic/procedural	C Hardboiled/neo-polar
Verbs: say, exclaim, answer, declare, murmur, reply, question	Verbs: question, phone, suppose, look, smoke, drink, remember	Verbs: feel, catch, put down, fuck, spot, put on, run into
Modifiers: dear, young, brave, honest, mysterious	Modifiers: sure, gloomy, police, drunk, tired	Modifiers: alone, responsible, aware, cop, guy
Possessives: friend, revolver, office, duty, master	Possessives: pipe, office, overcoat, hat, pocket	Possessives: jacket, phone, car, face, father/mother

Table 2: Distinctive attributes of the three periods

Interpretation and significance

- **Robust unity** of the archetype (stable detection, temporal LOGO)
- **Increasing centrality** of the detective as narrative driver
- **Semantic trajectory** through the genre history:
 - A: dialogic rationality (investigation through questioning)
 - B: empathy/procedural (Maigret, social embeddedness)
 - C: violent immersion and moral ambiguity (Série noire, neo-polar)
- **Revised Periodization** of the French Detective Novel: 1866-1931, 1931-1950, 1950-2020

Thank you !

Q & A

Contact:

jean.barre@ens.psl.eu
olga.seminck@cnrs.fr

Representations and Classifier

Representations

- **BoW** (1,000 MFW of attributes) > sparse vectors
- **CamemBERT_{LARGE}**: contextual averages (1024d) of attributes per character

Learning

- SVM and Logistic Regression (stratified 5-CV)
- **LOGO** (character, author, time periods) to prevent leakage
- Main metric: *balanced accuracy*

Limitations and future work

- **Coreference chains** (fusion/split errors) > noisy attributes
- **Unigram attributes**: no negation/gradation/disambiguation
- **Future**: multi-word expressions, explicit negation, sense disambiguation, “degree of detectiveness” (continuous), role intersections