



PSL 

Lattice
FRANCE

Opérationnaliser la Canonicité

Une Étude Quantitative dans la Littérature du XIX^e et XX^e
Siècle

Jean Barré

20 décembre 2023

Lattice Lab : ENS-PSL-CNRS

<http://tinyurl.com/5t4z92kw>

The image shows a screenshot of a journal article page with a teal background. At the top left, the word "Article" is written in a light blue font. Below it, the text "Vol. 8, Issue 3, 2023 · October 17, 2023 EDT" is displayed. The main title, "Operationalizing Canonicity: A Quantitative Study of French 19th and 20th Century Literature", is centered in a large, white, serif font. Below the title, the authors' names "Jean Barré, Jean-Baptiste Camps, Thierry Poibeau" are listed in a smaller white font. A horizontal row of eight rounded rectangular tags follows, containing the following terms: "Literary history", "Canon studies", "Computational Literary Studies", "French Literature", "machine learning", "distant reading", "natural language processing", "text mining", and "stylometry". At the bottom left of the teal area is the Creative Commons BY license logo (CC BY) and the text "4.0 · <https://doi.org/10.22148/001c.88113>".

Article

Vol. 8, Issue 3, 2023 · October 17, 2023 EDT

Operationalizing Canonicity: A Quantitative Study of French 19th and 20th Century Literature

Jean Barré, Jean-Baptiste Camps, Thierry Poibeau

Literary history Canon studies Computational Literary Studies French Literature machine learning
distant reading natural language processing text mining stylometry


 CC BY-4.0 · <https://doi.org/10.22148/001c.88113>

Figure 1 : Journal of Cultural Analytics

« Comment la sélection des œuvres et des noms destinés à l'immortalité est-elle faite ? » (Lanson, 1895)[1]

Formation du canon dans le champ socioculturel :

- « Tradition sélective » (Pollock, 1999)[2]
- Le canon « incarne la légitimité littéraire elle-même » (Casanova, 2008)[3]
- La canonisation en tant que processus sociologique (Bourdieu, 1992)[4]
- Existe-t-il des éléments empiriques / textuelles de ces recherches ?

Études littéraires computationnelles & le canon littéraire

- Lecture à distance. (Moretti, 2000)[5]
- Les études littéraires ne sont-elles pas vraiment familières avec les lignes structurantes principales de l'histoire littéraire ? (Underwood, 2019)[6]

Principales questions de recherche :

- Peut-on opérationnaliser le processus de canonisation ?
- Peut-on lier la canonicité à certaines dynamiques textuelles ?
- Les propriétés textuelles sont-elles un phénomène causal ? Ou le produit du processus de canonisation ?

Le canon littéraire français à travers sa réception contemporaine

- Le canon scolaire

Détermination des facteurs canoniques

Le canon littéraire français à travers sa réception contemporaine

- Le canon scolaire
- Le canon académique

Le canon littéraire français à travers sa réception contemporaine

- Le canon scolaire
- Le canon académique
- Le canon de l'agrégation

Le canon littéraire français à travers sa réception contemporaine

- Le canon scolaire
- Le canon académique
- Le canon de l'agrégation
- Le canon des éditeurs

Le canon littéraire français à travers sa réception contemporaine

- Le canon scolaire
- Le canon académique
- Le canon de l'agrégation
- Le canon des éditeurs
- **Le canon de la critique**

Le canon littéraire français à travers sa réception contemporaine

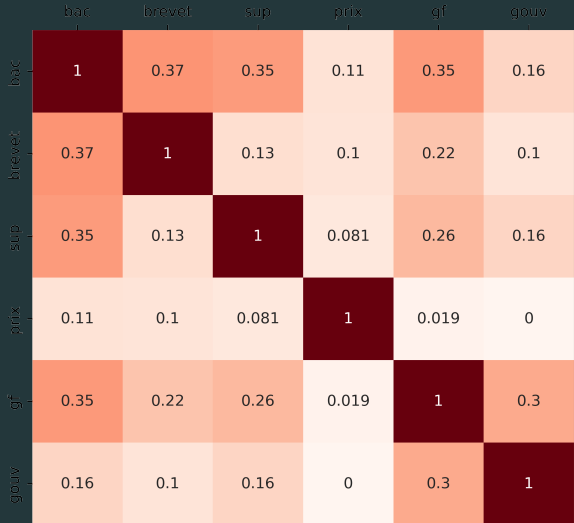
- Le canon scolaire
- Le canon académique
- Le canon de l'agrégation
- Le canon des éditeurs
- Le canon de la critique
- Le canon politique

Granularité double

- Échelle du roman - 306 éléments (10% du corpus)
- Échelle de l'auteur - 1173 éléments (40%)
- L'institution scolaire en tant que fabricant de canon ?

Carte de similarité cosinus

- Forte similarité entre les facteurs basés sur l'école
- La liste des prix littéraires est clairement différente



Corpus des chapitres (Leblond, 2022)[7] :

- 2960 romans

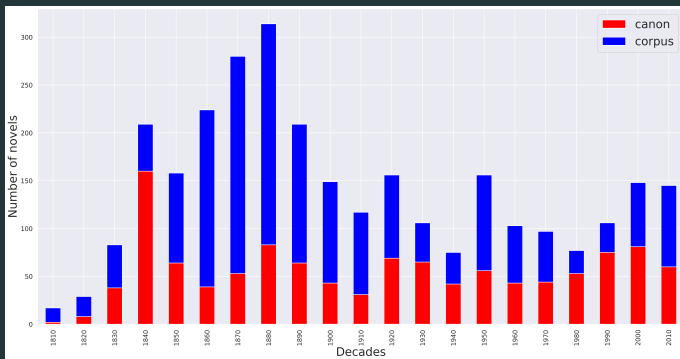


Figure 3 : Distribution du nombre de romans au fil du temps

Sac de mots

- 1000 n-grammes les plus fréquents
- Accent sur les mots fonctionnels -> manière inconsciente d'écrire (stylométrie) -> marqueurs inconscients de la sélection canonique ?

Classification binaire

- SVM - Scikit-Learn (Pedregosa, 2011) [8]
- GroupKFold - Pipeline de validation croisée traitant du biais idiolectal
- Baselines
- Mesures (précision équilibrée, score f1, ...)

Précision équilibrée : **0.708** Baseline :0.496

	précision	rappel	score f1	support
canon	0.728	0.668	0.697	306
non_canon	0.691	0.748	0.719	304
ensemble du corpus				610

Table 1 : Résultats de l'évaluation du modèle, échelle du roman

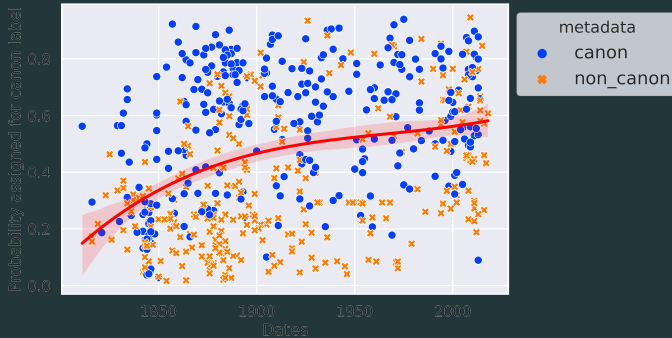


Figure 4 : Probabilité prédite d'être canonique, échelle du roman

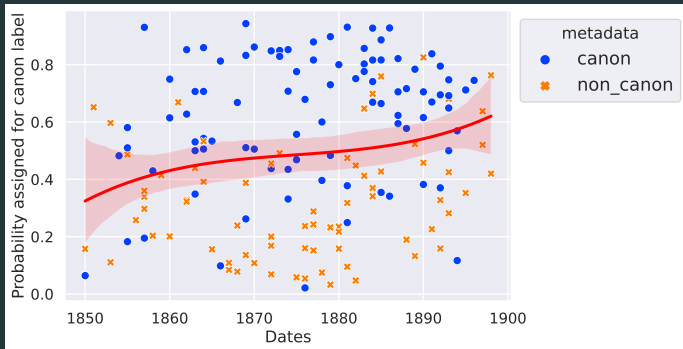


Figure 5 : Probabilité prédite d'être canonique, 1850-1900, échelle du roman

Précision équilibrée : 0.741 Baseline : 0.516

	précision	rappel	score f1	support
canon	0.721	0.645	0.681	1173
non_canon	0.782	0.836	0.808	1787
ensemble du corpus				2960

Table 2 : Résultats de l'évaluation du modèle, échelle de l'auteur

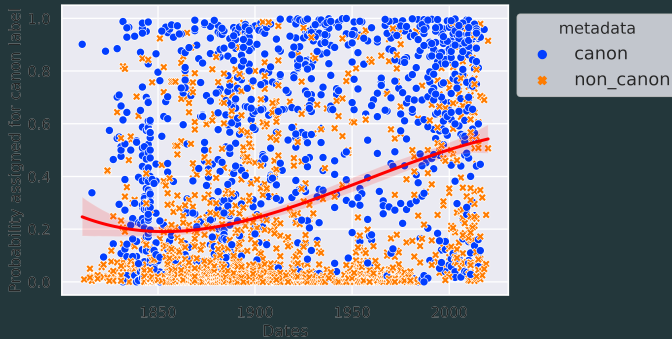


Figure 6 : Probabilité prédite d'être canonique, échelle de l'auteur

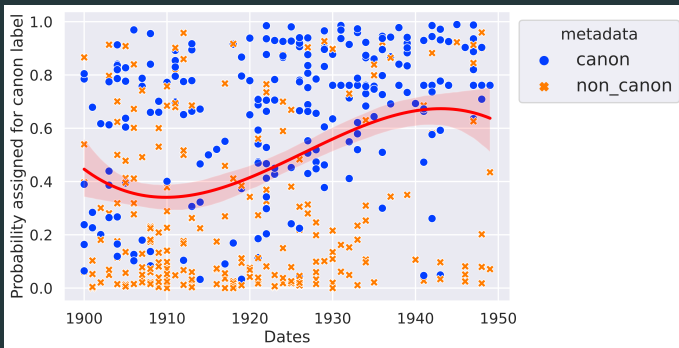
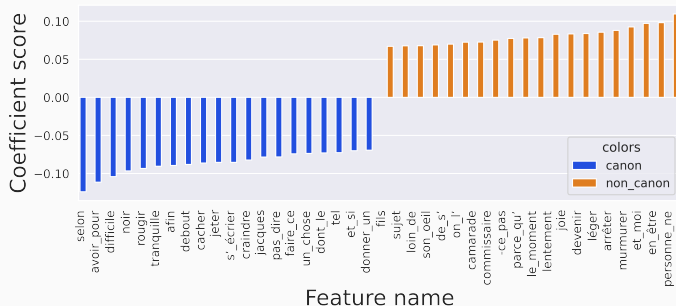


Figure 7 : Probabilité prédite d'être canonique, échelle de l'auteur

Coefficient discriminant du modèle

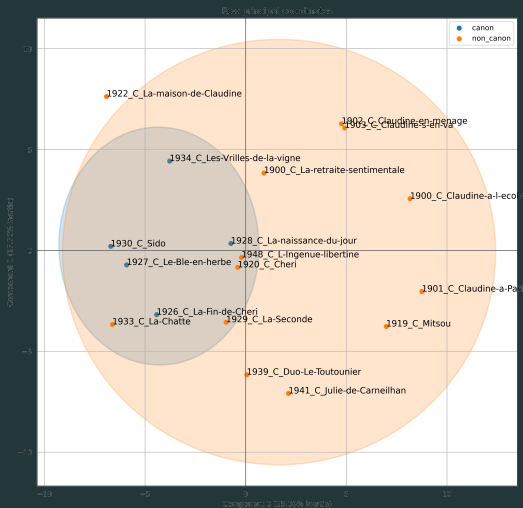


Aperçu des coefficients du modèle

- Complexité de la phrase (auxiliaires, conjonctions, noms substantifs)
- Un registre plus familier pour les romans non canonisés

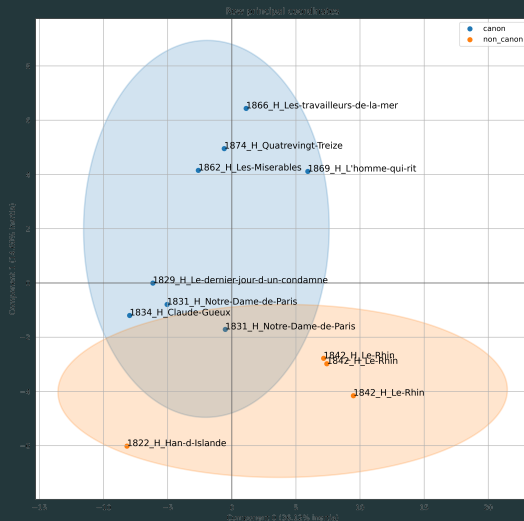
ACP Colette

- La série *Claudines*
- Pas seulement un décalage idiolectal/chronolectal

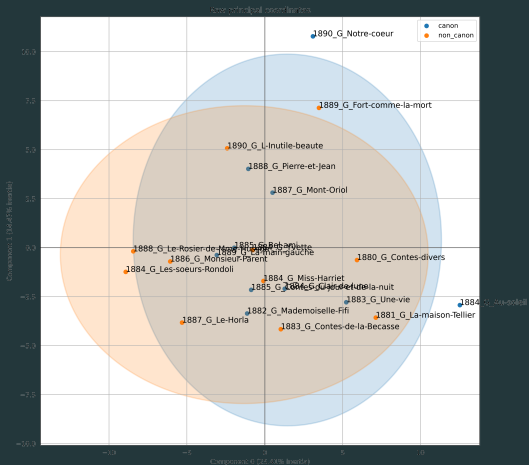


ACP Victor Hugo

- Les trois volumes du *Le Rhin* - Un guide de voyage
- *Han d'Islande*
Le jeune Hugo



ACP Maupassant



- Le canon comme **Grammaire Culturelle** (Altieri, 1983)[9]
- Le canon agit comme un cadre façonnant non seulement la création littéraire actuelle, mais influençant également la manière dont nous percevons rétrospectivement le passé, l'alignant sur les normes contemporaines.
- une norme linguistique émerge de mécanismes de sélection latents biaisés liés au rôle du système éducatif dans le processus de formation du canon.

- Récupérer les métadonnées diachroniquement : récupérer le contexte social de la réception au fil du temps. Pas le même canon
- Fragmenter notre vue du canon : Éditions, manuels, prestige académique, revues littéraires, ...
- Embeddings / LLMs ?
- Récupération et analyse d'extraits « canoniques » en lecture proche

Conclusion

- Définition opérationnalisable de la notion de canon littéraire
- Un modèle statistique peut prédire la canonicité avec une précision de 70% à 74%
- Le modèle produit des critères relativement valides pour une période spécifique, mais échoue pour deux siècles de production littéraire
- Cette norme détectée soutient la recherche socioculturelle sur le processus de canonisation et lui ajoute un aspect formel.
- Nous supposons que cette norme est le résultat de mécanismes de sélection latents biaisés qui produisent une valeur littéraire et une « immortalité » littéraire.
- La littérature est également déterminée par ses propres institutions et conventions, par ses propres mécanismes de production et de reproduction.

Merci de votre attention !

Bibliographie i



Gustave Lanson.

Hommes et livres : études morales et littéraires.

Hachette livre-bnf, 1895.



Griselda Pollock.

Differencing the canon : feminist desire and the writing of art's histories.

Re visions. Routledge, 1999.



Pascale Casanova.

La république mondiale des lettres.

Number 607 in Points Série essais. Éditions du Seuil, Édition revue et corrigé e edition, 2008.



Pierre Bourdieu.

Les règles de l'art.

Éditions du Seuil, 1992.



Franco Moretti.

Conjectures on world literature.

New Left Review, 2000.



Ted Underwood.

Distant horizons : digital evidence and literary change.

The University of Chicago Press, 2019.



Aude Leblond.

Corpus chapitres, December 2022.



F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay.

Scikit-learn : Machine learning in Python.

Journal of Machine Learning Research, 12 :2825–2830, 2011.



Charles Altieri.

An idea and ideal of a literary canon.

10(1) :37–60.

Données et code disponible sur github : *https://github.com/crazyjeannot/canonization_process/*

LaTeX et PDF : *https://github.com/crazyjeannot/redaction_canonization_process*

N'hésitez pas à m'écrire

jean.barre@ens.psl.eu,