

Beyond Canonicity

Modeling Canon/Archive Literary Change in French Fiction

Jean Barré, Thierry Poibeau

7 december 2023

École Normale Supérieure - PSL University - Lattice Lab

Computational Literary Studies & Canon / Archive

- « Distant Reading » (Moretti, 2000) [1]
- Do we understand the outlines of literary history? (Underwood, 2019) [2]
- Canon/Archive - going beyond canonicity : literature as a dynamic system

Beyond Canonicity

- Literary variability (Tynianov, 1927) [3] as a framework to study literature
- Naïve inquiry : which authors brought novelty to fiction (formal/semantic?)
- Is our contemporary canon grasping those innovations?

Corpus

- Chapitres Corpus (Leblond, 2022)[4] : 2961 novels
- Canonical tags : (Barré, 2023)[5]

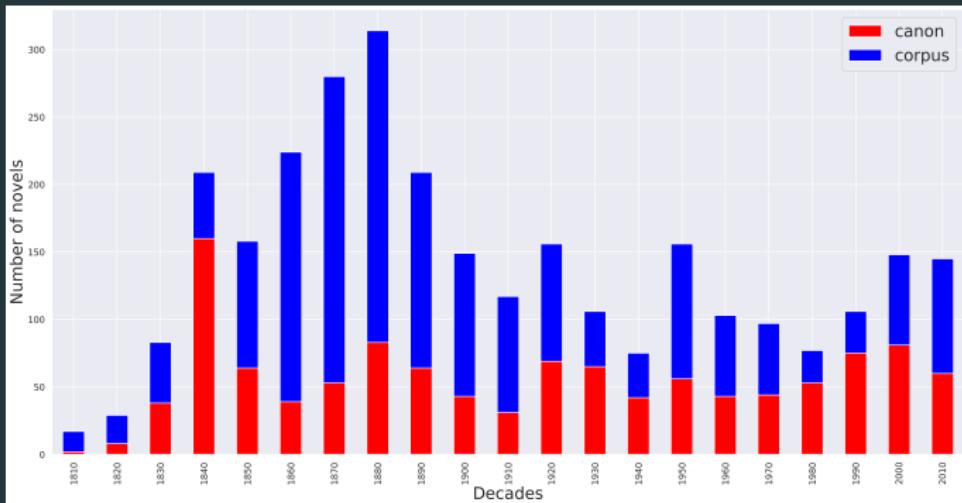


Figure 1 – Corpus distribution

Between Formal & Semantic Textual Features

- **Characterization** : operationalizing the Character Space (Woloch, 2003) [6] with coreference resolution & Fr-BookNLP.
Characterizing adjectives and verbs throughout the narrative.

Between Formal & Semantic Textual Features

- **Characterization** : operationalizing the Character Space (Woloch, 2003) [6] with coreference resolution & Fr-BookNLP.
Characterizing adjectives and verbs throughout the narrative.
- **Chronotope** : fr-BookNLP NER : LOC + FAC + VEH + TIME.
(Bamman, 2019) [7]

Textual Features

Between Formal & Semantic Textual Features

- **Characterization** : operationalizing the Character Space (Woloch, 2003) [6] with coreference resolution & Fr-BookNLP.
Characterizing adjectives and verbs throughout the narrative.
- **Chronotope** : fr-BookNLP NER : LOC + FAC + VEH + TIME.
(Bamman, 2019) [7]
- **Themes** : Fifty main topics (BERTTopics) and their proportions for each novel are identified and retrieved

Textual Features

Between Formal & Semantic Textual Features

- **Characterization** : operationalizing the Character Space (Woloch, 2003) [6] with coreference resolution & Fr-BookNLP.
Characterizing adjectives and verbs throughout the narrative.
- **Chronotope** : fr-BookNLP NER : LOC + FAC + VEH + TIME.
(Bamman, 2019) [7]
- **Themes** : Fifty main topics (BERTTopics) and their proportions for each novel are identified and retrieved
- **Lexicon** : 1000 most frequent words. Unconscious structural writing.

Each text is a concatenated vector : Doc2Vec + Topics prop + BoW

Literary Variability

- Cohort renewal stronger than author change (Underwood, 2022)
[8]

Literary Variability

- Cohort renewal stronger than author change (Underwood, 2022) [8]
- Generation : 30 years (Moretti, 2007), (Béhar, 1996) [9, 10] as a cycle of change ?

Literary Variability

- Cohort renewal stronger than author change (Underwood, 2022)[8]
- Generation : 30 years (Moretti, 2007), (Béhar, 1996) [9, 10] as a cycle of change ?
- Formal innovation vs. Imitation

Literary Variability

- Cohort renewal stronger than author change (Underwood, 2022)[8]
- Generation : 30 years (Moretti, 2007), (Béhar, 1996) [9, 10] as a cycle of change?
- Formal innovation vs. Imitation
- Kullback/Leibler Divergence - Measuring the pace of change

$$KL(P||Q) = \sum_x P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

Measuring change - entropy graph

- Canon more influent & more innovative
- x-axis : high score means high novelty
- y-axis : low score means high influence

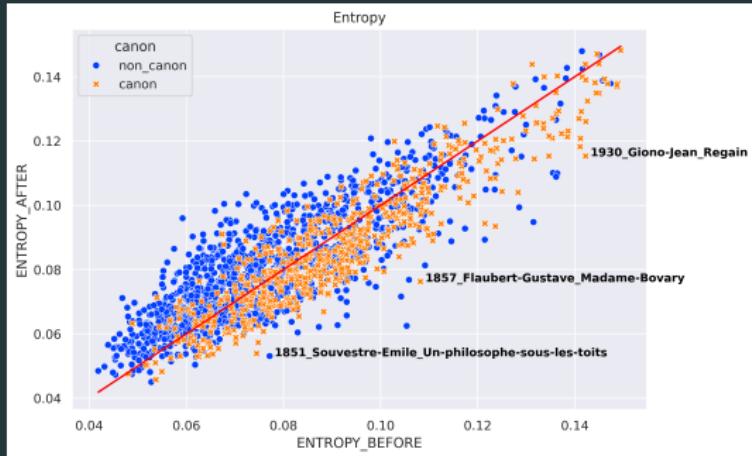


Figure 2 – Entropy graph, surprise with previous cohort (x-axis), surprise with next cohort (y-axis), Canon label in orange

Novelty Signal

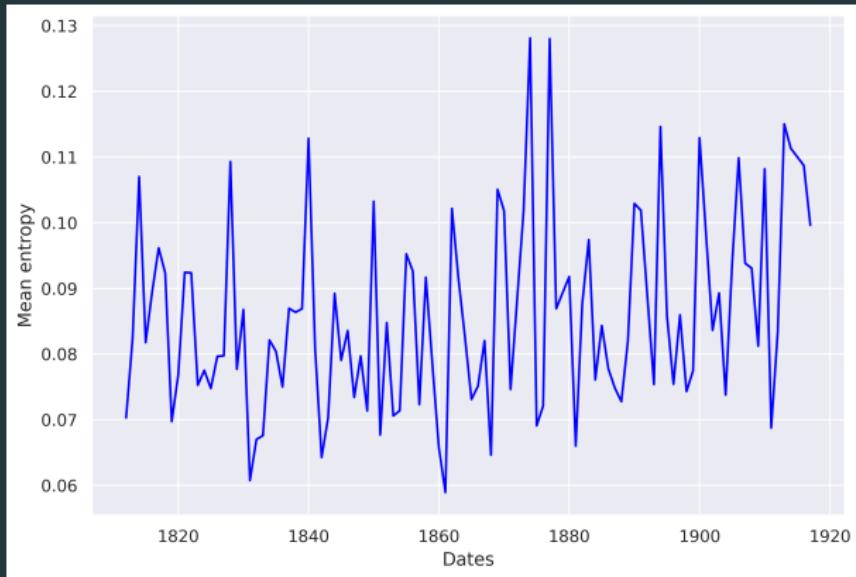


Figure 3 – Mean novelty per birth-year

- Each peak : Moment of innovation
- Which authors contributing to it?

Three Peaks Under Study

- 1828

	index	dates	novelty
1871_Verne-Jules_Une-Ville-flottante	1828	0.096299	
1872_Verne-Jules_Le-tour-du-monde-en-quatre-vi...	1828	0.106026	
1874_Verne-Jules_Le-Chancellor	1828	0.111852	
1875_Verne-Jules_L-Ile-Mysterieuse	1828	0.231755	
1876_Verne-Jules_Michel-Strogoff	1828	0.113568	

- 1877

	index	dates	novelty
1882_Toudouze-Gustave_La-Seductrice	1877	0.078932	
1897_Roussel-Raymond_La-doublure	1877	0.153972	
1900_Roussel-Raymond_Chiquenaude	1877	0.273218	
1910_Roussel-Raymond_Impressions-d-Afrique	1877	0.145402	
1911_Chateaubriant-Alphonse-de_Monsieur-des-Lo...	1877	0.061054	

- 1900

	index	dates	novelty
1935_Crevel-René_Le-roman-cassé	1900	0.170642	
1939_Sarraute-Nathalie_Tropismes	1900	0.158931	
1995_Sarraute-Nathalie_Ici	1900	0.161942	
1997_Sarraute-Nathalie_Ouvrez	1900	0.173511	

Canon vs Archive : Novelty

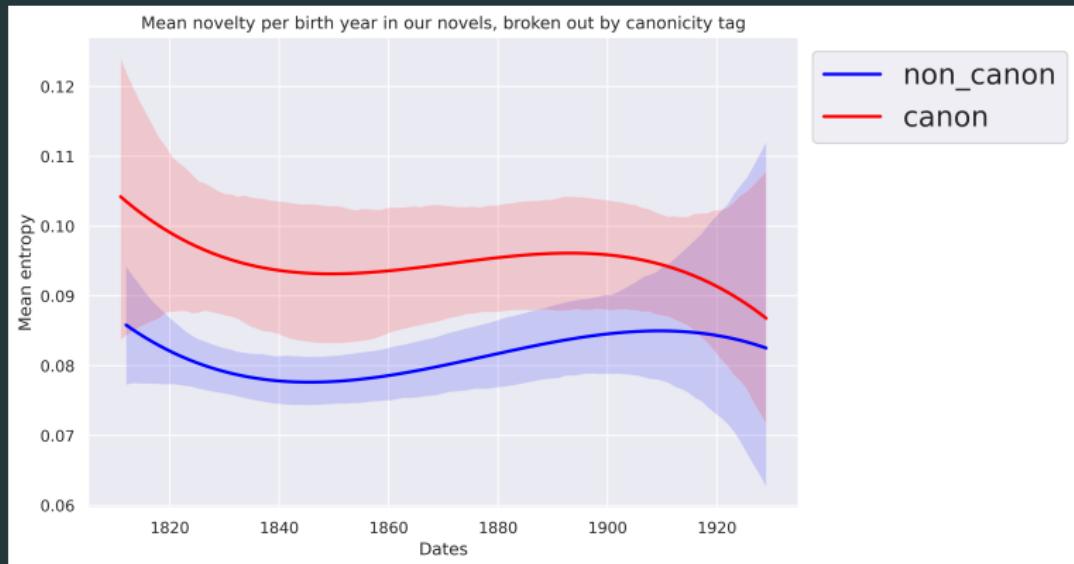


Figure 4 – Mean novelty per birth-year, broken out by canonical tag

Future Work

Close Reading is all you need

- Assumptions which still need validation :
which feature is relevant for each peak?
- Are peaks of "variability" correlated with the emergence of specific sub-genres?
- Are canonical works/authors driving novelty? Yes on average but not for all literary periods

Thank you!

Bibliography i

-  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.
-  Yuri Tynianov.
On Literary Evolution (1927), pages 267–282.
Academic Studies Press, Boston, USA, 2019.
-  Aude Leblond.
Corpus chapitres, December 2022.

Bibliography ii

-  Jean Barré, Jean-Baptiste Camps, and Thierry Poibeau.
Operationalizing canonicity : A quantitative study of french 19th and 20th century literature.
Journal of Cultural Analytics, 8(3), 2023-10-17.
-  Alex Woloch.
The One vs. the Many.
Princeton University Press, 2003.
-  David Bamman, Sejal Popat, and Sheng Shen.
An annotated dataset of literary entities.
pages 2138–2144, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

Bibliography iii

-  Ted Underwood, Kevin Kiley, Wenyi Shang, and Stephen Vaisey.
Cohort succession explains most change in literary culture.
Sociological Science, 9 :184–205, 2022-05-02.
-  Franco Moretti and Franco Moretti.
Graphs, maps, trees : abstract models for literary history.
Verso, paperback edition edition, 2007.
-  Henri Béhar.
La littérature et son golem.
Number 58 in Vol. 1 : Travaux de linguistique quantitative. H. Champion, 1996.

Q & A

Feel free to reach us!

jean.barre@ens.psl.eu

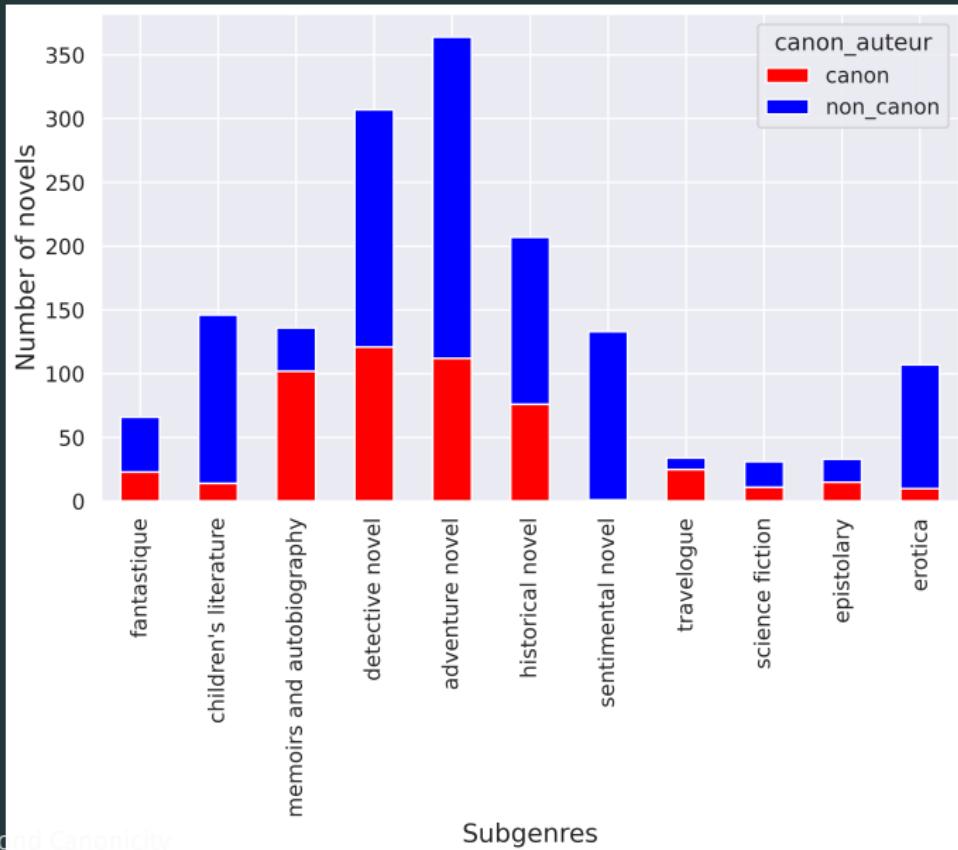
thierry.poibeau@ens.psl.eu

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

Data & Code :

https://github.com/crazyjeannot/CHR_beyond_canicity

Annexes



Measuring change - entropy graph

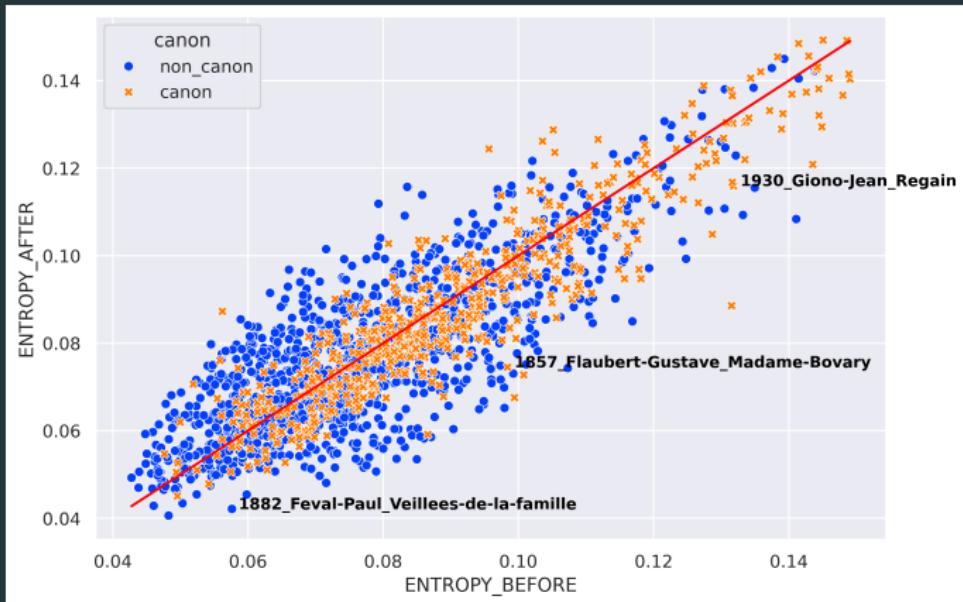


Figure 6 – Entropy graph, surprise with previous cohort (x-axis), surprise with next cohort (y-axis), canon in orange