Fr-BookNLP : Applications

Frédérique Mélanie-Becquet, Olga Seminck, Jean Barré

5 décembre 2023

Lattice : ENS-PSL-CNRS

Fr-BookNLP in Action

Study on Gender with Fr-BookNLP

Study on Gender : Research Questions

- How does the author's gender influence the gender of characters ?
- What is the distribution between male and female characters ?
- Which stereotypes are associated with male and female characters ?
- How do gender stereotypes evolve over time ?

Inspiration

- Underwood, Ted, David Bamman, Sabrina Lee. *The Transformation of Gender in English-Language Fiction.* Journal of Cultural Analytics, 3, 2, 2018. *doi:https://doi.org/10.22148/16.019*

Study on Gender : Publications of Lattice

- Male and female novelists : an automatic study of gender of authors and their characters.
  Marco Naguib, Marine Delaborde Blandine Andrault, Anaïs Bekolo & Olga Seminck.
  TALN2022.
- A Computational History of Gender in French Fiction : Machine Learning & Literature.
  Jean Barré.
  DHAI Intensive Week 2022.
- Gender bias in French literature.
  Laurine Vianne, Yoann Dupont & Jean Barré.
  CHR2023.

Predicting a character's gender

French BookNLP gives us the coreference chains :

- "M. Barton ; lui ; l' ; l' ; M. Barton ; Excellent homme ; il ; M. Barton ; sa ; il ; M. Barton ; lui ; il ; ..."
- "Juliette ; Juliette ; lui ; son ; son ;..."

Accuracy between 0.60 and 0.91 :

- using mentions : 0.91 (CHR2023), 0.64 (TALN2022)

Distribution of Male and Female Characters

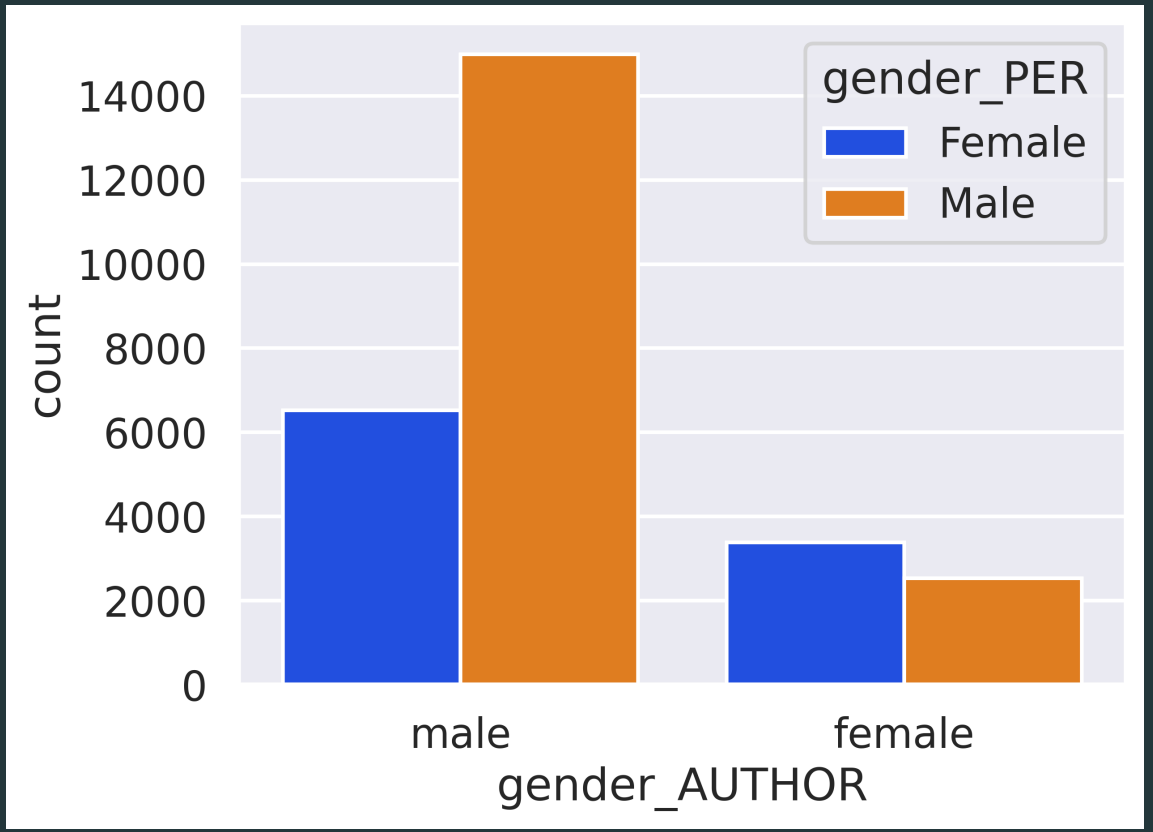
As a function of the gender of the author (TALN2022)...

	Mentions			Chains		
	Masc.	Fem.	Mixed/N.R.	Masc.	Fem.	Mixed/N.R.
Men	35,73%	24,24%	40,03%	19,23%	11,03%	69,74%
Women	27,75%	36,83%	35,42%	18,84%	16,58%	64,58%

- Male authors write more about male characters than female characters.
- Female authors have female characters with longer chains (leading characters).

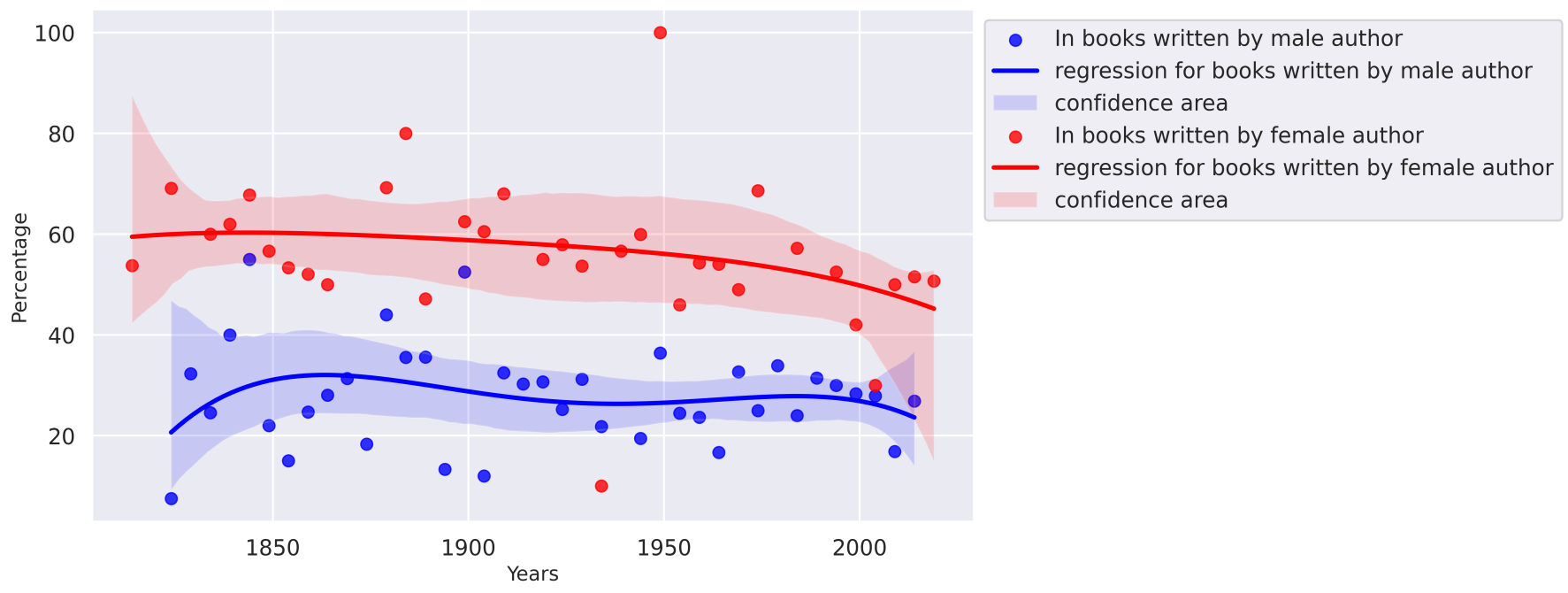
Distribution of Male and Female Characters

As a function of the gender of the author (DHAI Intensive Week 2022)...



Distribution of Male and Female Characters

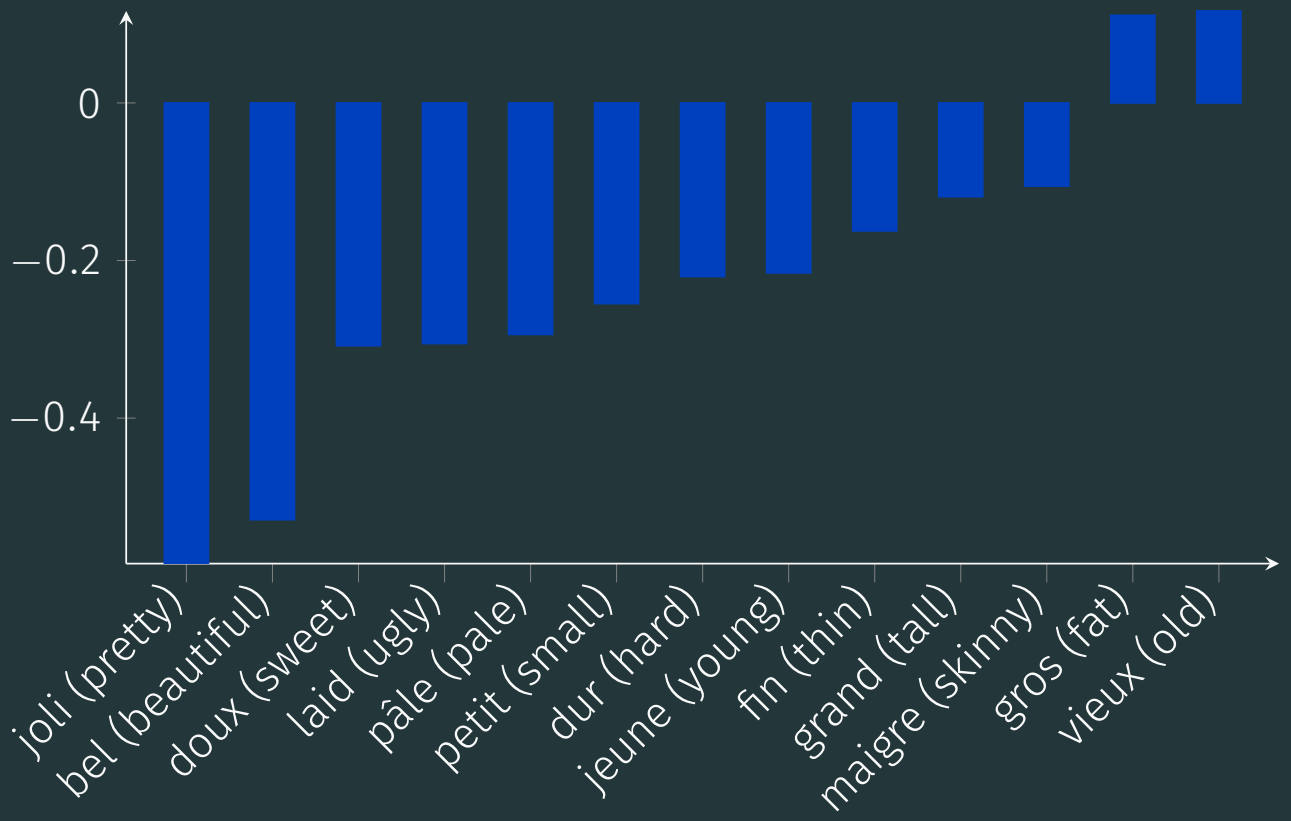
As a function of the gender of the author (DHAI Intensive Week 2022)...



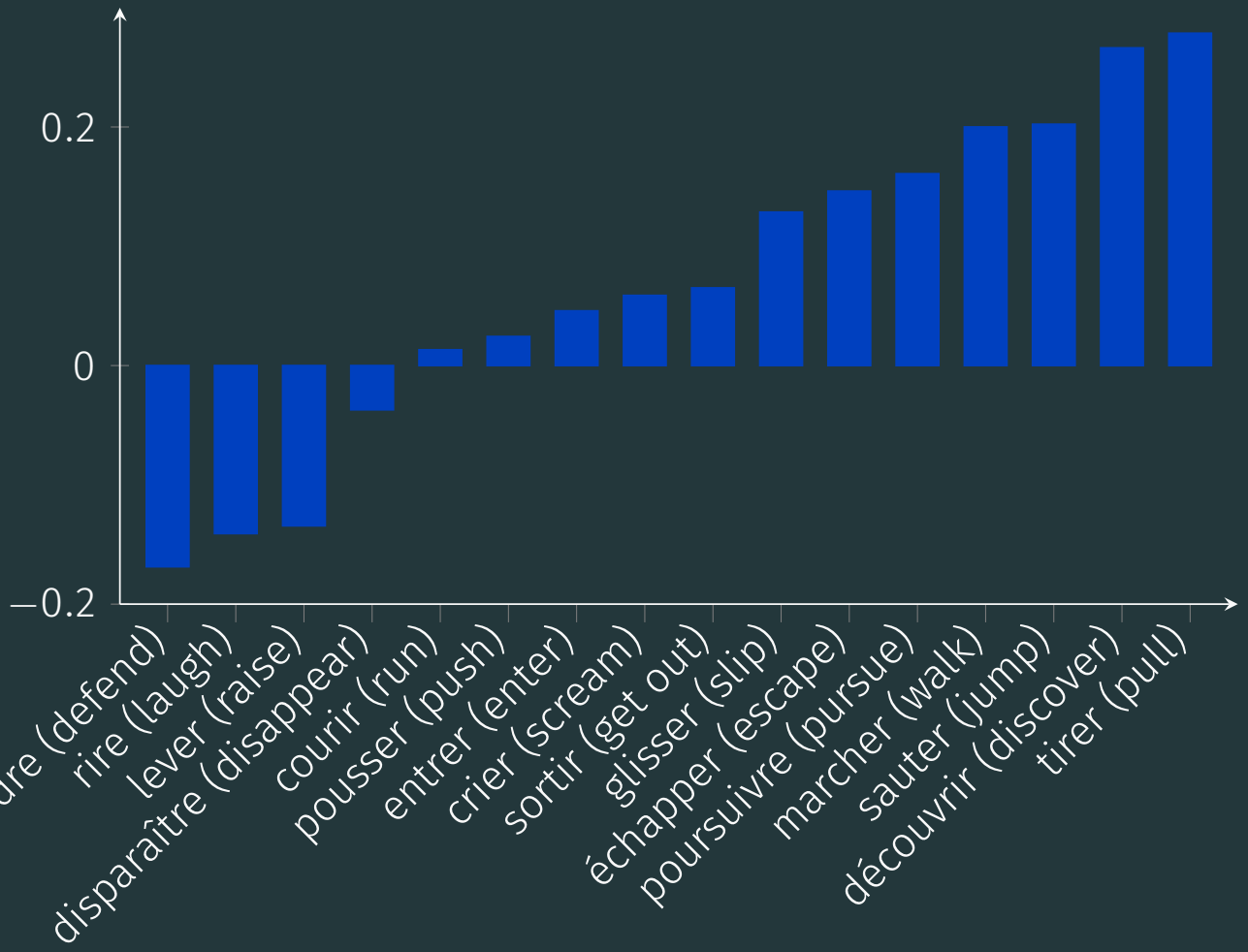
Distribution of Male and Female Characters

Predicting the gender of a character using adjectives describing the characters : 0.78 accuracy (CHR2023)

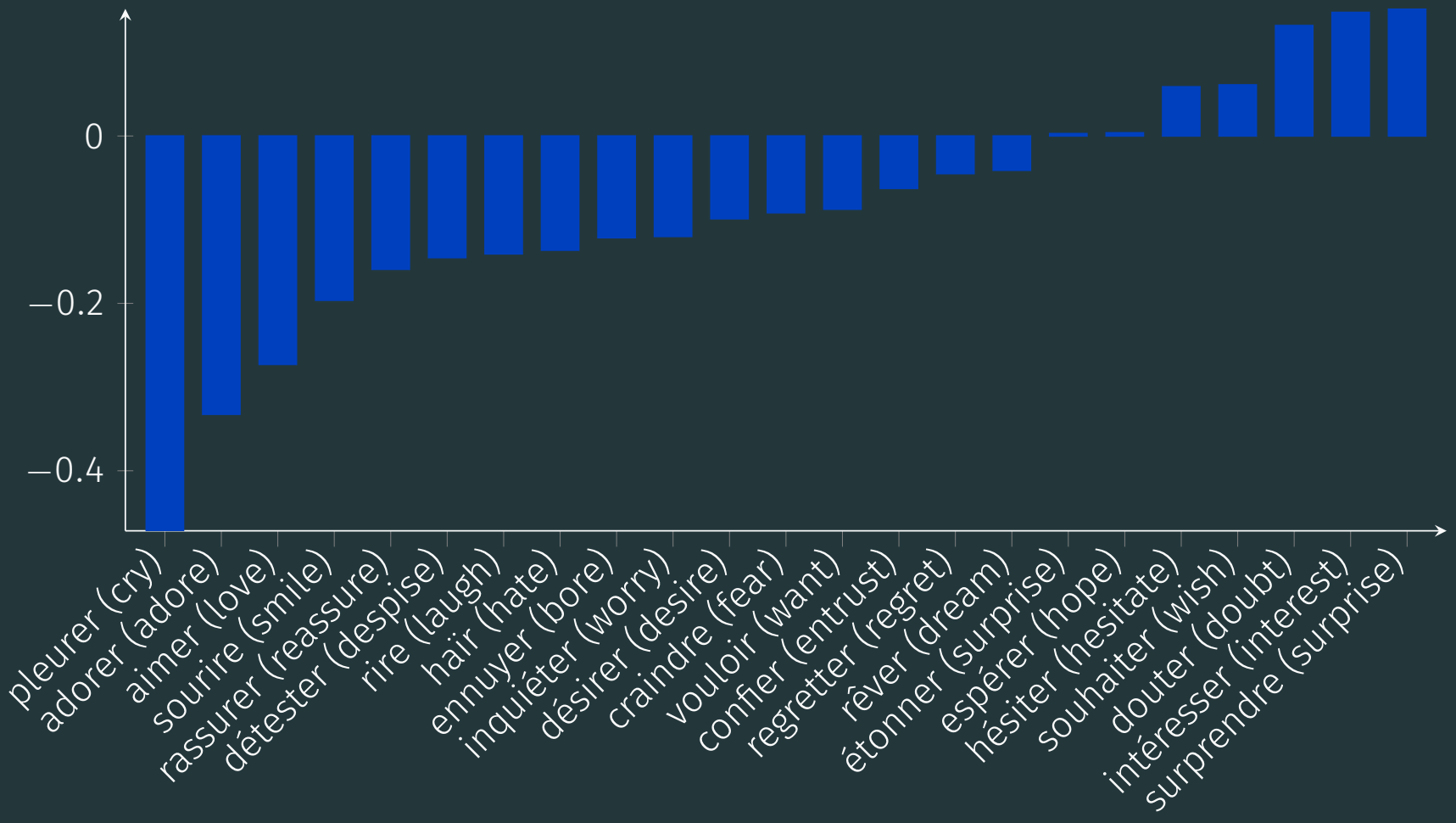
Gender Stereotypes CHR2023 :
Adjectives Associated with Male and Female Characters



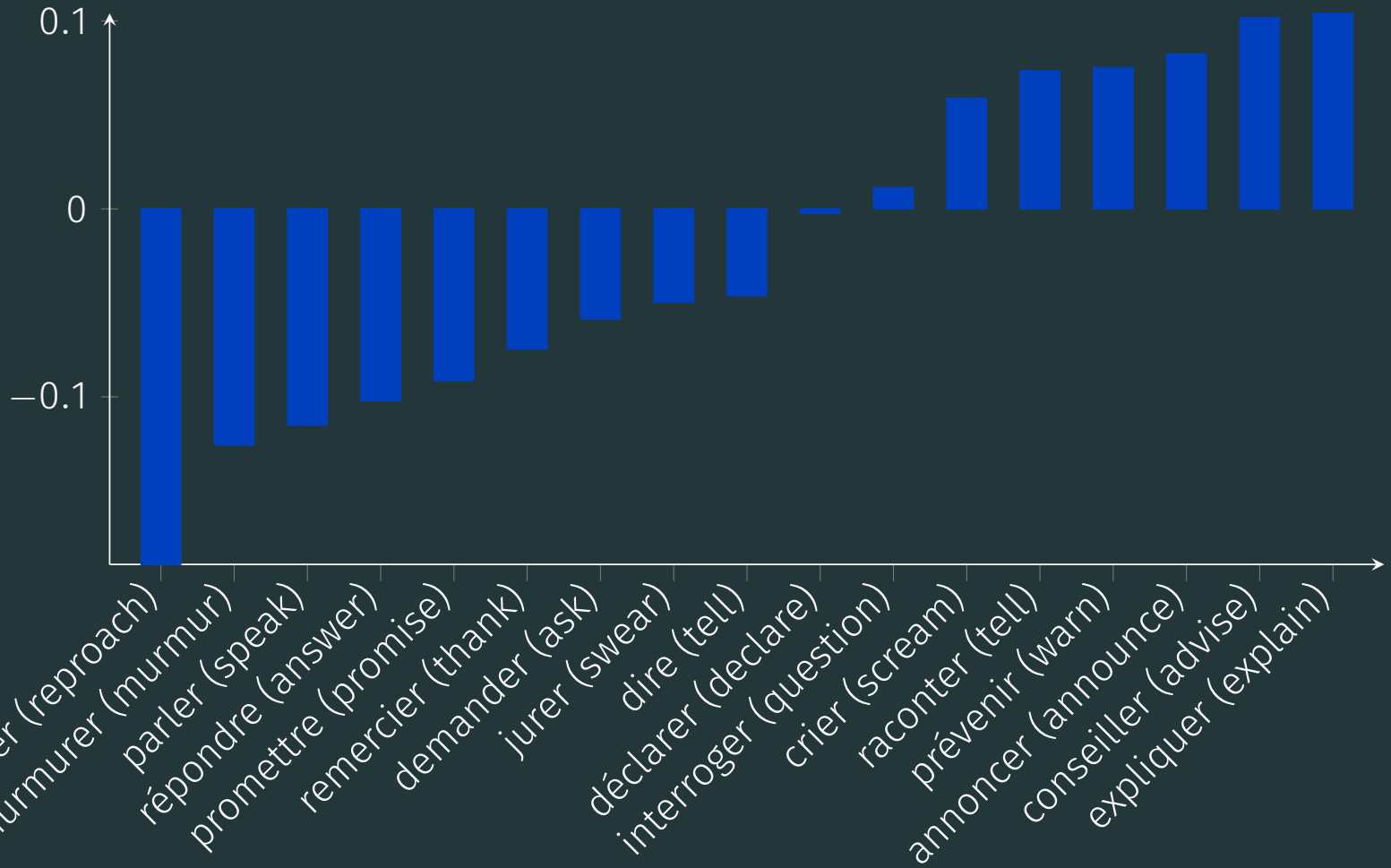
Gender Stereotypes CHR2023 :
Action Verbs Associated with Male and Female Characters



Gender Stereotypes CHR2023 :
Emotional Verbs Associated with Male and Female Characters



Gender Stereotypes CHR2023 :
Speech Verbs Associated with Male and Female Characters



Gender Stereotypes TALN2022 : Most Specific Vocabulary of Male and Female Authors

Some examples :

♂ : baron, guerre, Monseigneur, grand-duc, peuple, franc, gendarme, pape, barbare, guerrier, journal, chef, français, romain, soldat, ministre, million, haut, administration, préfet, cadavre, ventre, nation, rival, prince, trappeur, prêtre, coup, homme, patron
♀ : Madame, sentiment, bonheur, bijou, éprouver, monsieur, mère, coeur, aimer, affection, âme, douleur, lettre, Mademoiselle, charme, heureux, devoir, croire, inspirer, cousin, pensée, enfant, bonté, cruel, vie, amie, malheureux, regret, amour, marquise, généreux, tort, malheur, doux

Gender Stereotypes TALN2022 : Most Specific Vocabulary of Male and Female Authors

Male jobs :

♂ : baron, guerre, Monseigneur, grand-duc, peuple, franc, gendarme, pape, barbare, guerrier, journal, chef, français, romain, soldat, ministre, million, haut, administration, préfet, cadavre, ventre, nation, rival, prince, trappeur, prêtre, coup, homme, patron

policeman, pope, warrior, chieff, soldier, minister, prefect, trapper, priest, boss

Gender Stereotypes TALN2022 : Most Specific Vocabulary of Male and Female Authors

War/violence :

♂ : baron, guerre, Monseigneur, grand-duc, peuple, franc, gendarme, pape, barbare, guerrier, journal, chef, français, romain, soldat, ministre, million, haut, administration, préfet, cadavre, ventre, nation, rival, prince, trappeur, prêtre, coup, homme, patron

war, people, barbarian, warrior, soldier, corps, nation, rival, blow/shot

Gender Stereotypes TALN2022 : Most Specific Vocabulary of Male and Female Authors

Hierarchy/Titles :

♂ : baron, guerre, Monseigneur, grand-duc, peuple, franc, gendarme, pape, barbare, guerrier, journal, chef, français, romain, soldat, ministre, million, haut, administration, préfet, cadavre, ventre, nation, rival, prince, trappeur, prêtre, coup, homme, patron

baron, My lord, Grand Duke, chief, minister, high, administration, prefect, nation, prince, boss

Gender Stereotypes TALN2022 : Most Specific Vocabulary of Male and Female Authors

Emotions :

♀ : Madame, sentiment, bonheur, bijou, éprouver, monsieur, mère, coeur, aimer, affection, âme, douleur, lettre, Mademoiselle, charme, heureux, devoir, croire, inspirer, cousin, pensée, enfant, bonté, cruel, vie, amie, malheureux, regret, amour, marquise, généreux, tort, malheur, doux

feeling, happiness, feel, heart, to love, affection, soul, pain, happy, to inspire, thought, goodness, cruel, unhappiness, regret, love, generous, wrong, misfortune, soft

Gender Stereotypes TALN2022 : Most Specific Vocabulary of Male and Female Authors

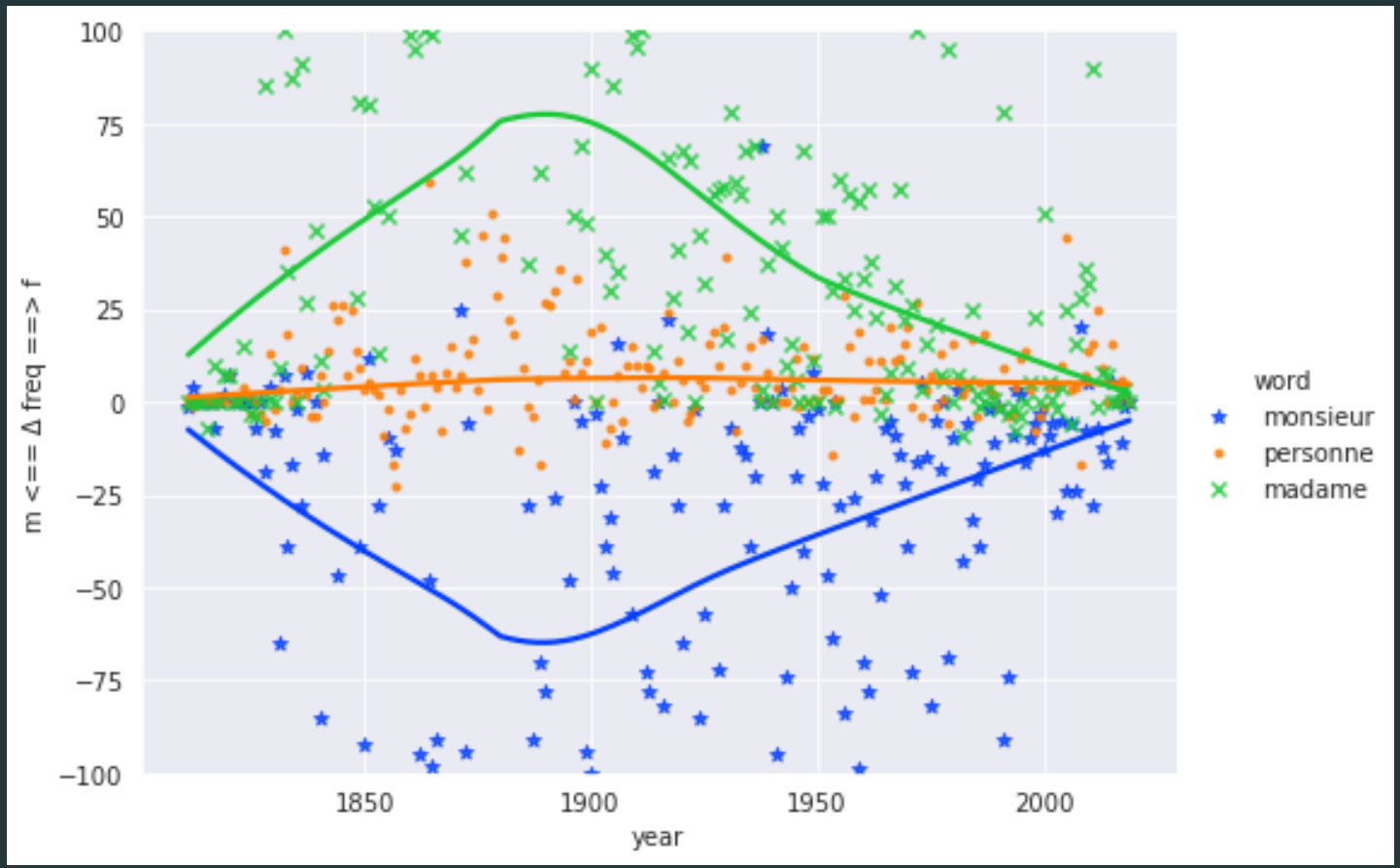
Family/Relations :

♀ : Madame, sentiment, bonheur, bijou, éprouver, monsieur, mère, coeur, aimer, affection, âme, douleur, lettre, Mademoiselle, charme, heureux, devoir, croire, inspirer, cousin, pensée, enfant, bonté, cruel, vie, amie, malheureux, regret, amour, marquise, généreux, tort, malheur, doux

Mrs, Mister, mother, Miss, cousin, child, friend

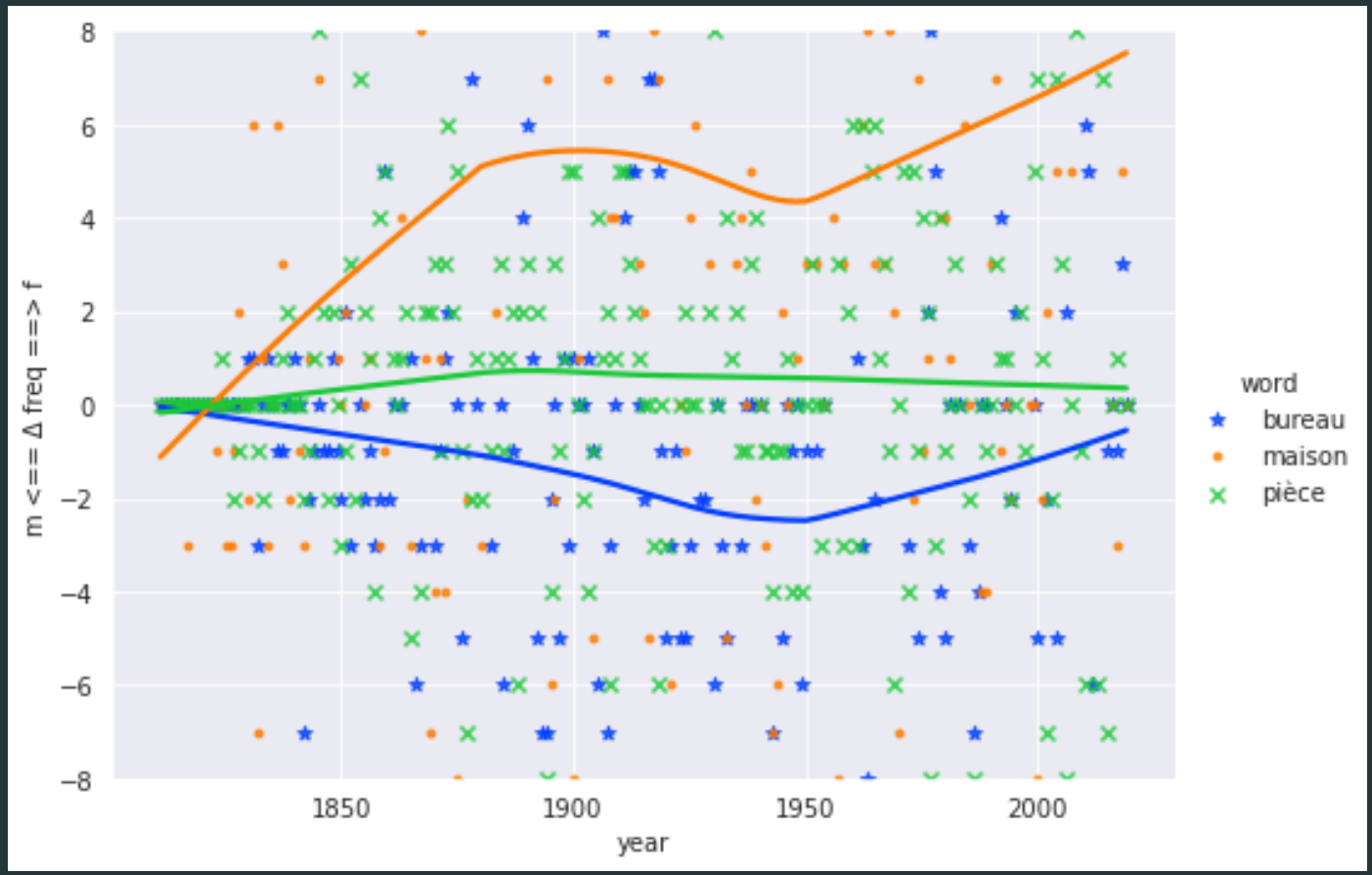
Evolution of Stereotypes over Time

The gendered connotation of a given word comparison of Monsieur, Madame & Personne (DHAI Intensive Week 2022)...



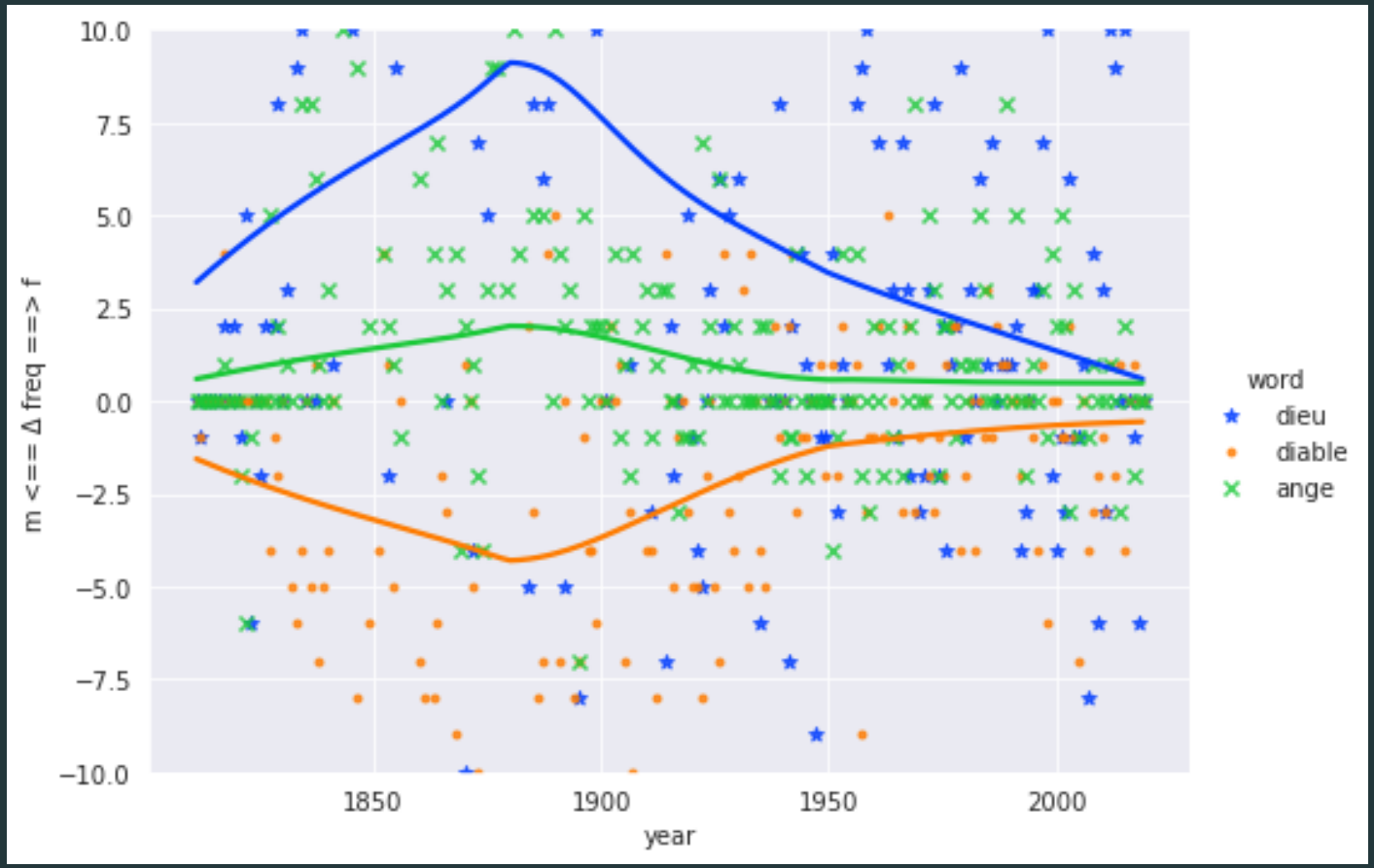
Evolution of Stereotypes over Time

The gendered connotation of a given word comparison of Bureau (desk), Maison (house) & Pièce (Room) (DHAI Intensive Week 2022)...



Evolution of Stereotypes over Time

The gendered connotation of a given word comparison of Dieu (God), Diable (devil) & Ange (angle) (DHAI Intensive Week 2022)...



Future Work : longitudinal studies on the use of stereotypes

- Develop a set of ‘features’ to measure stereotypicality of characters
  - # physical traits
  - # mentions of a job or function of a character
  - # different places where the characters goes
- Being able to detect a number of male and female stereotypes in novels
  - femme fatale
  - princes
  - mother
  - cowboy
  - sensitive poet
  - old wise man
- Measure changes in stereotypicality of characters through the career of writers and between time periods

Takeaways

- Fr-BooKNLP helps us to evaluate of how literary characterization is linked to gender biases.
- There exist individual words and lexical fields that are strongly related to gender stereotypes.
- The proportion of characterization of female characters strongly depends on the author's gender.
- Male authors write half as much about female characters as female authors.

Using LLMs & Fr-BookNLP for Subgenre Analysis : First Experiments

Subgenre Analysis : Adventure Novels as a Case Study

Adventure Novels vs Adventure Passages

- « Adventure is the essence of fiction » (Tadié, 1996) :
- Beyond subgenre labels : Passage level

LLMs as Annotation Power

- Can LLMs detect stereotypicality in adventure novels ?
- Can we detect the number of adventure scenes in a big corpus ?

Stereotypicality in Adventure Novels

Definition :
« Adventure is characterized by the importance of change of scenery (historical/geographical/fantastical or social) and of violent action putting the hero in mortal danger or physical peril » (Letourneux, 2010)

stereotypicality detection

- Working at the passage level (4-5 pages)
- What type of information is relevant ? Suspense, Spatiotemporal framework, Characterization ?

Method

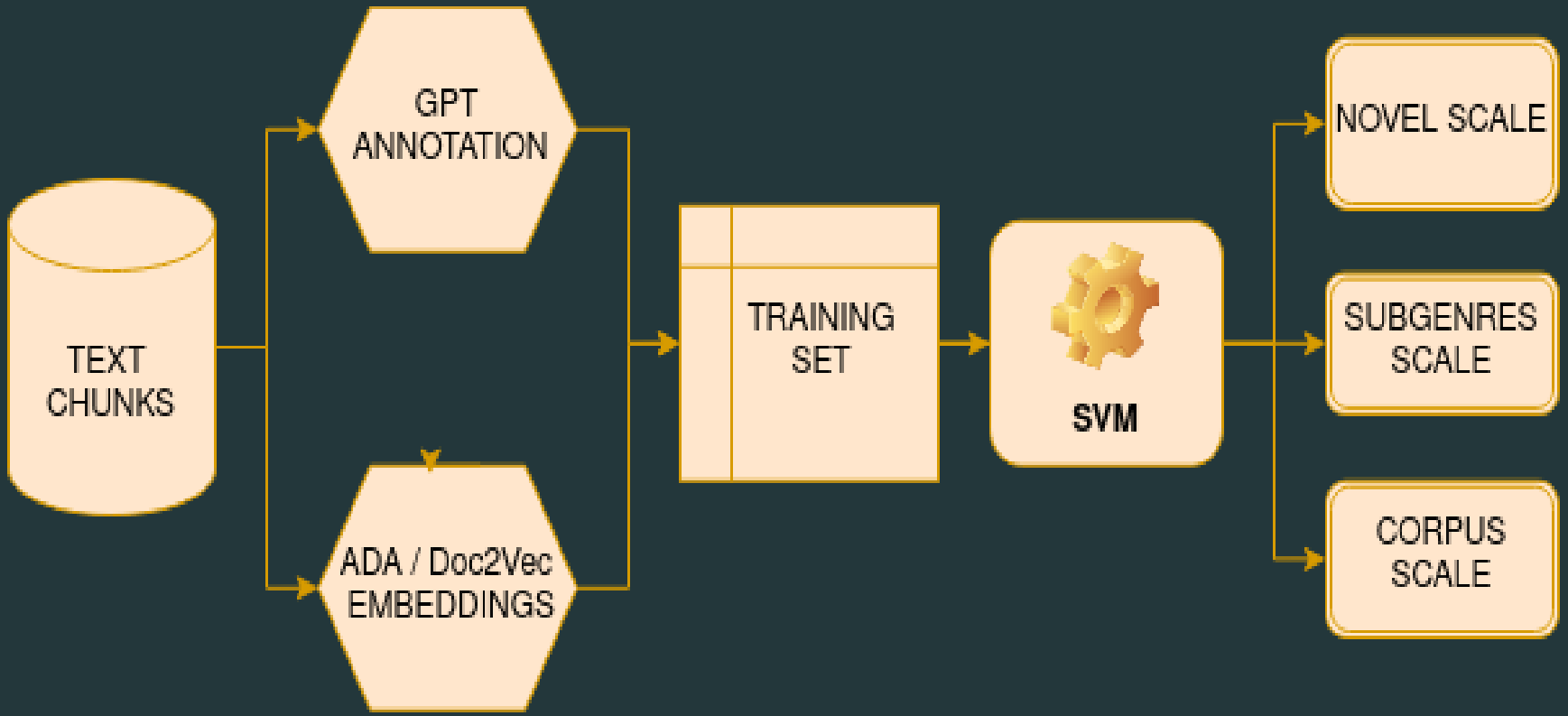


Figure 1 : Workflow

Annotation : Prompt Engineering

Prompt : Prompt : Give me a one-word output : ADVENTURE, if this text is typical of the adventure genre, else write NON_ADVENTURE. Prefer the output NON_ADVENTURE when unsure. HELP : adventure novels are characterized by the importance of changes of scenery (historical/geographical/fantastical or social) and of violent actions putting the hero in mortal danger or physical peril. A typical adventure scene would consist of someone (described as brave/heroic) doing something dangerous in a heroic manner and in a wild setting)

GPT 3.5 turbo annotation as ground truth

- Annotation of 1000 examples : ADVENTURE vs NON_ADVENTURE
- Evaluation of a few examples

Training Pipeline

Textual Features - 3 levels

- All passages' tokens
- Fr-BookNLP tokens (characterization + chronotope)
- Random tokens

Embeddings

- OpenAI Embeddings
- Paragraph Vectors

Statistical Modeling

- SVM - SOTA for text classification

SVM Results

	All tokens	BookNLP tokens	Random tokens
ADA embeddings	**0.86**	0.77	0.72
DBoW	0.78	0.69	0.63

Table 1 : Benchmark evaluation

Multiscale Analysis : Novel Scale

Vingt mille lieues sous les mers (Verne, 1869)

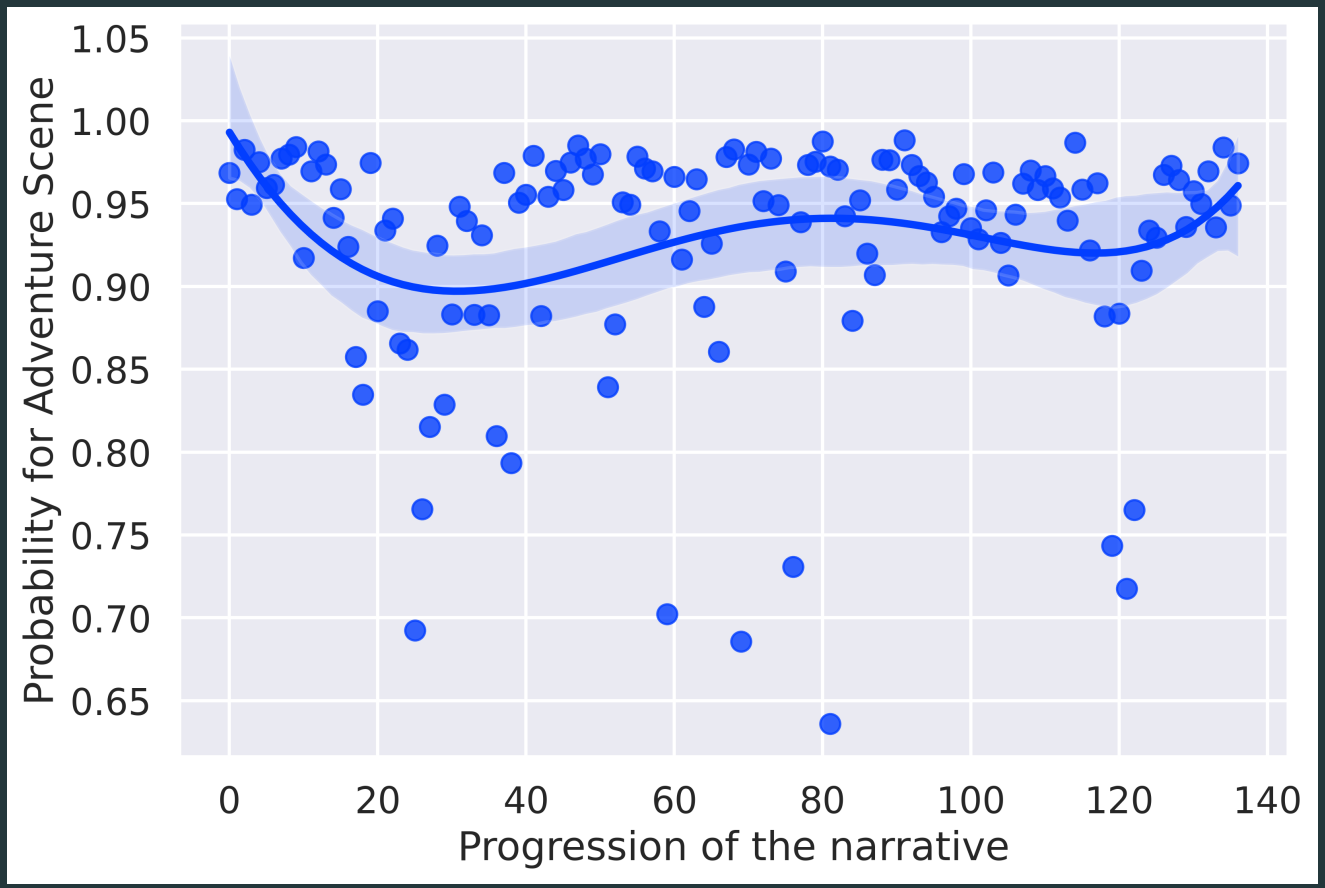


Figure 2 : Adventures Scenes in *Vingt mille lieues sous les mers*

Multiscale Analysis : Novel Scale

l'Éducation Sentimentale (Flaubert, 1869)

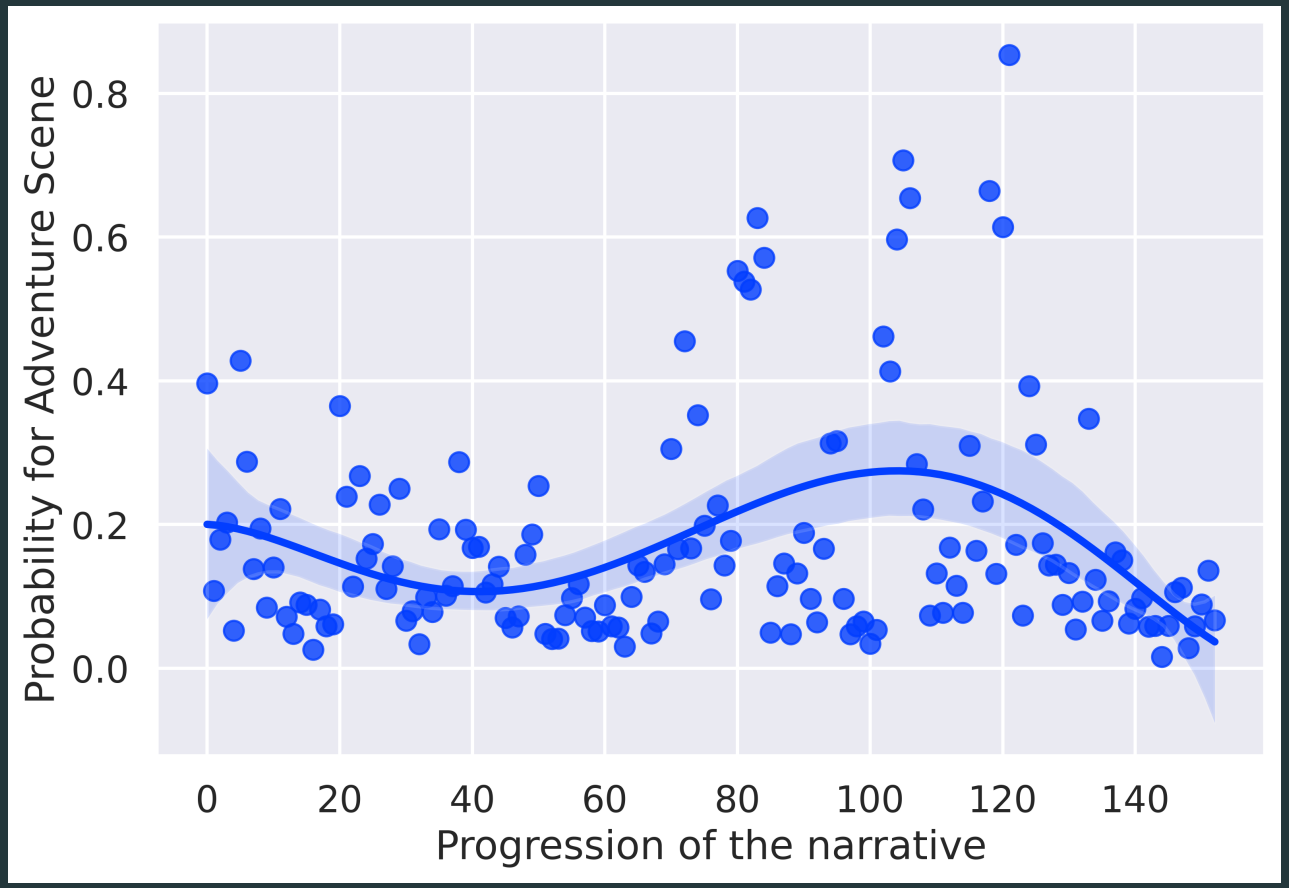


Figure 3 : Adventures Scenes in l'*Éducation Sentimentale*

Multi-scale Analysis : Subgenre Scale

Subgenre labels : Chapitres corpus

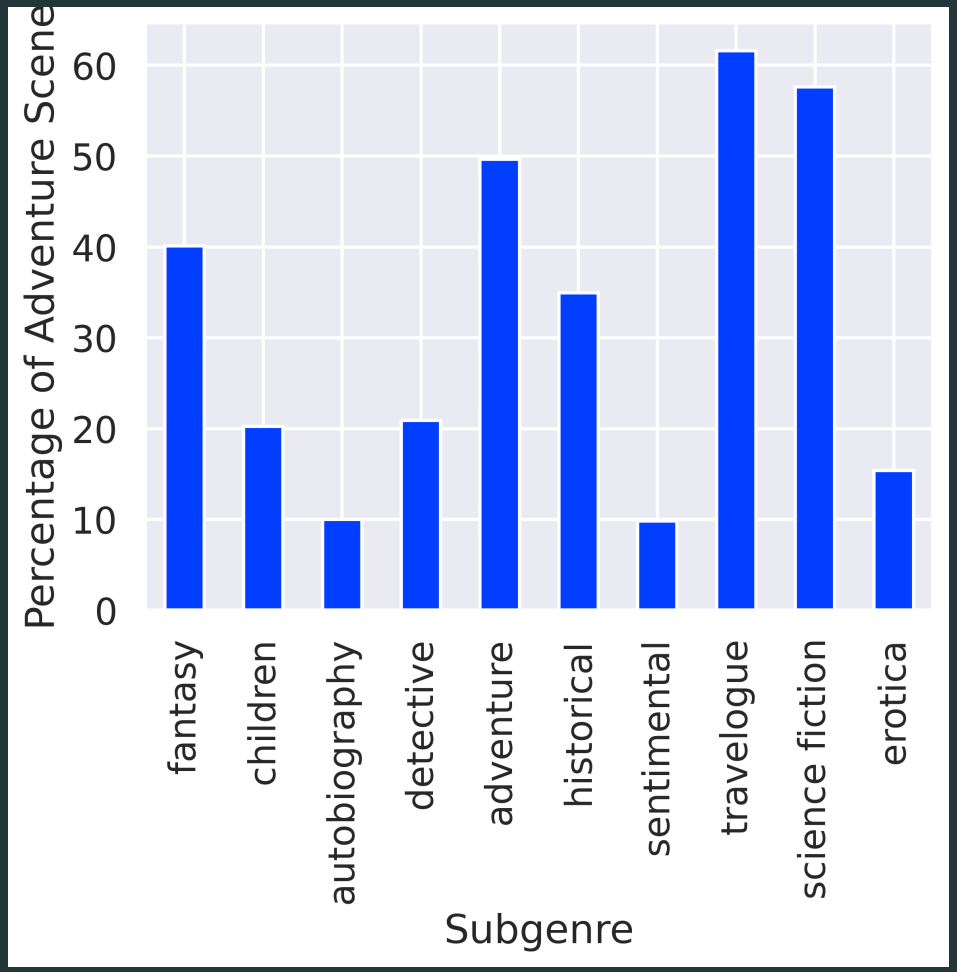


Figure 4 : Percentage of Adventures Scenes by subgenres

Multi-scale Analysis : Corpus Scale

Whole corpus - 10.000 annotated chunks from a corpus of 3000 novels

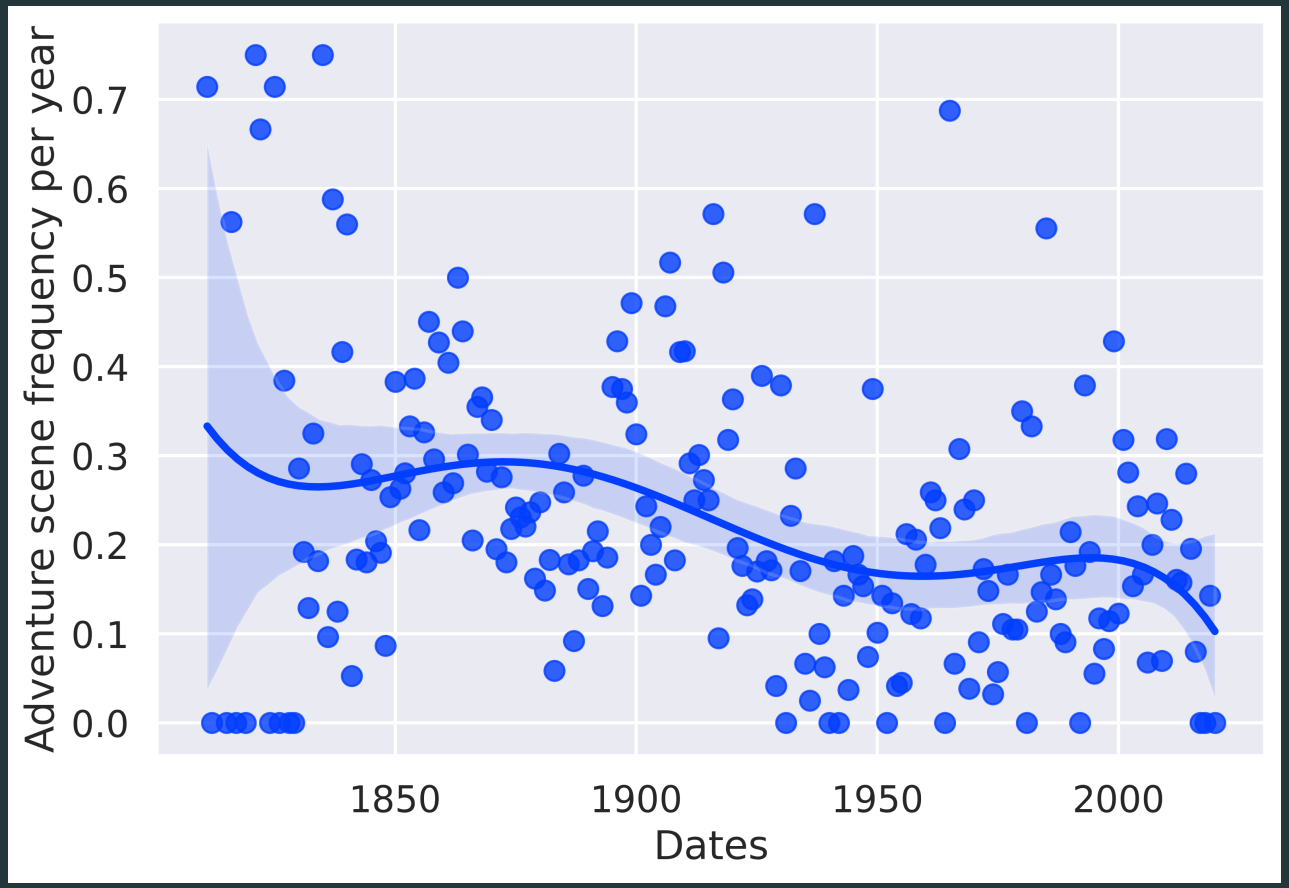


Figure 5 : Adventures Scenes in the whole corpus

Takeaways & Future Work

- LLMs can be useful for literary scholars
- Annotation Power : Automatic close reading ?
- Human evaluation of synthetic annotation
- Finetune an ”open” LLM (Llama2/3 ? Mistral/Hermes ?)
- Detect stereotypical adventure scene might be too broad : fine grained our labels : Combat Scene, Love Scene, Discovery Scene..

Conclusion

- Fr-BookNLP online deployment - Coming soon (Christmas ?).
- Can contribute significantly to literary studies - Scaling up, trends over centuries of literature.
- Gender bias - Literary subgenres - Character networks
- LLMs : New tool for DH research. They can be useful - but can we trust them ? And more broadly, do we need them ?

Thank you!

Questions ?

Feel free to reach us!

olga.seminck@cnrs.fr
jean.barre@ens.psl.eu
frederique.melanie@ens.psl.eu
thierry.poibeau@ens.psl.eu

Useful links - BookNLP

BookNLP Berkeley

- *https://github.com/dbamman/litbank*
- *https://github.com/booknlp/booknlp*

Fr-BookNLP

- *https://www.lattice.cnrs.fr/projets/booknlp/*
- *https://github.com/lattice-8094/fr-litbank/*