



| PSL★

Lattice  
Frigg

PR[AI]RIE  
Practical Reasoning in Logic and Knowledge Representation

# Aperçu des Travaux sur le Genre des Personnages au Lattice

---

Jean Barré  
21 mars 2024

Lattice : ENS-PSL-CNRS

# Étude sur le Genre : Questions de Recherche

- Quelle est la répartition entre les personnages masculins et féminins ?
- Quels stéréotypes sont associés aux personnages masculins et féminins ?
- Comment les stéréotypes de genre évoluent-ils avec le temps ?
- Comment le genre de l'auteur influence-t-il les résultats ?

Ted Underwood, 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>

## Principaux résultats :

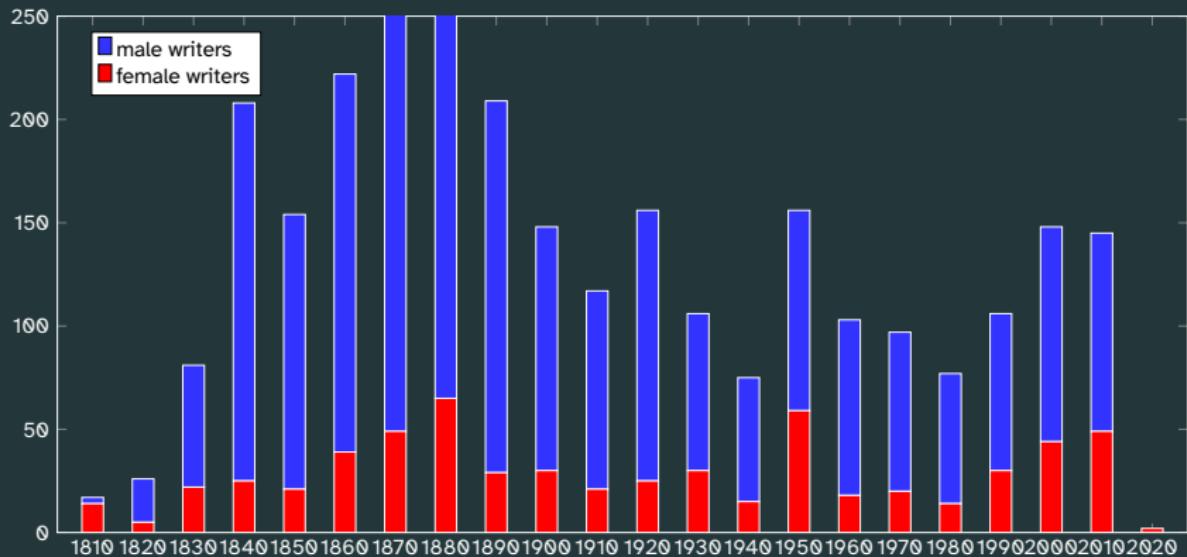
1. Un modèle prédictif entraîné avec des mots comme caractéristiques et des étiquettes féminines et masculines perd en précision entre les années 1980 et aujourd'hui
2. Le temps accordé aux personnages féminins est 3 fois moins important dans le cas d'un auteur masculin
3. Suivre les mots individuels liés au genre

# Études sur le genre : Publications du 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.  
Jean Barré.  
**DHAI Intensive Week 2022.**
- Gender bias in French literature.  
Laurine Vianne, Yoann Dupont & Jean Barré.  
**CHR2023.**

# Corpus Chapitres

- 2961 novels
- 79.301 : mean number of tokens per novel



**Figure 1 :** Distribution dans les temps

French BookNLP nous donne la chaîne de coreférence :

- “M. Barton; lui; l’; l’; M. Barton; Excellent homme; il; M. Barton; sa; il; M. Barton; lui; il; ...”
- “Juliette; Juliette; lui; son; son;...”

Accuracy entre 0.60 and 0.91 :

- using mentions : 0.91 ([CHR2023](#)), 0.64 ([TALN2022](#))

## Méthode

- Classifieur : RandomForest
- Annotation manuelle de 10 personnages  $\times$  100 livres
- 2 objectifs :
  - Utiliser les prédictions pour étendre l'annotation de genre
  - Le classifieur peut-il prédire le genre uniquement à partir du vocabulaire ?

## Caractéristiques textuelles

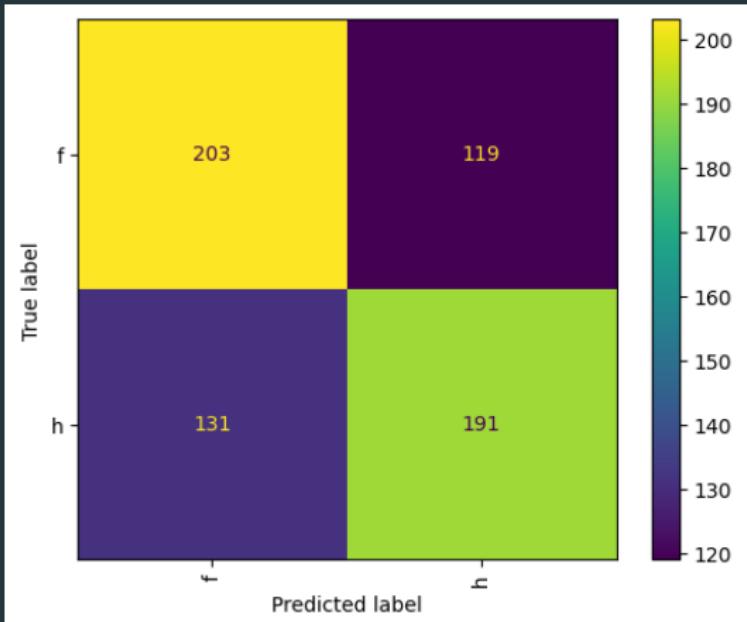
- Sac de mots
- Adjectifs et/ou verbes liés aux personnages
  - Analyse de dépendance avec le modèle SpaCy fr\_dep\_news\_trf
  - Lemmatisation pour supprimer les marques de genre évidentes en français
- Mentions (she, Mrs. ..., the woman...)

## Prédire le genre sans les marques grammaticales

- Utilisation uniquement des adjectifs et des verbes (pas de mentions)
- Suppression des marques de genre évidentes :  
*"Je suis heureuse d'y être allée"*  
*"I'm glad I went there"*
- Lemmatisation

# Efficacité des modèles de prédiction du genre

1. Mentions : 92%
  - Modèle utilisé pour généraliser
2. Lemmatized  
Adjectives+Verbs : 61%



**Figure 2 :** Matrice de confusion (avec adjectifs+verbes lemmatisés)

# Distribution des personnages hommes et femmes

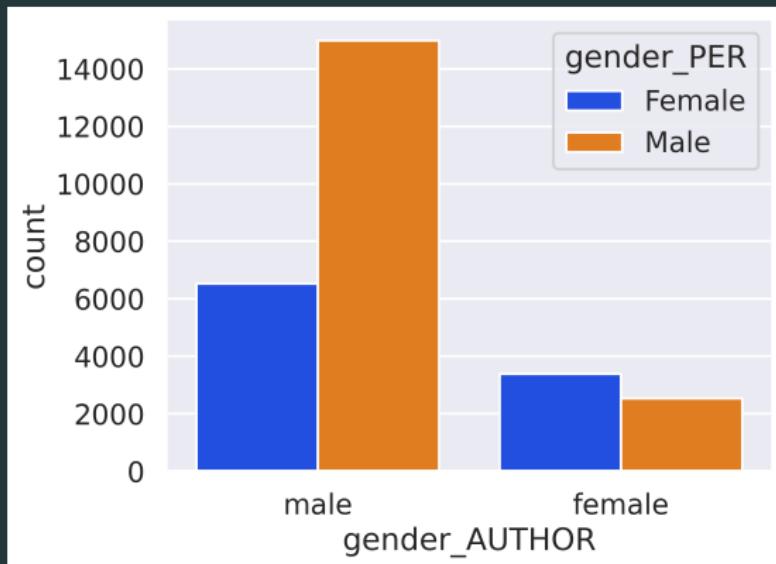
En fonction du genre de l'auteurice ([TALN2022](#))...

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

- Les auteurs écrivent plus sur les personnages masculins que sur les personnages féminins.
- Les autrices ont des personnages féminins avec des chaînes plus longues (personnages principaux).

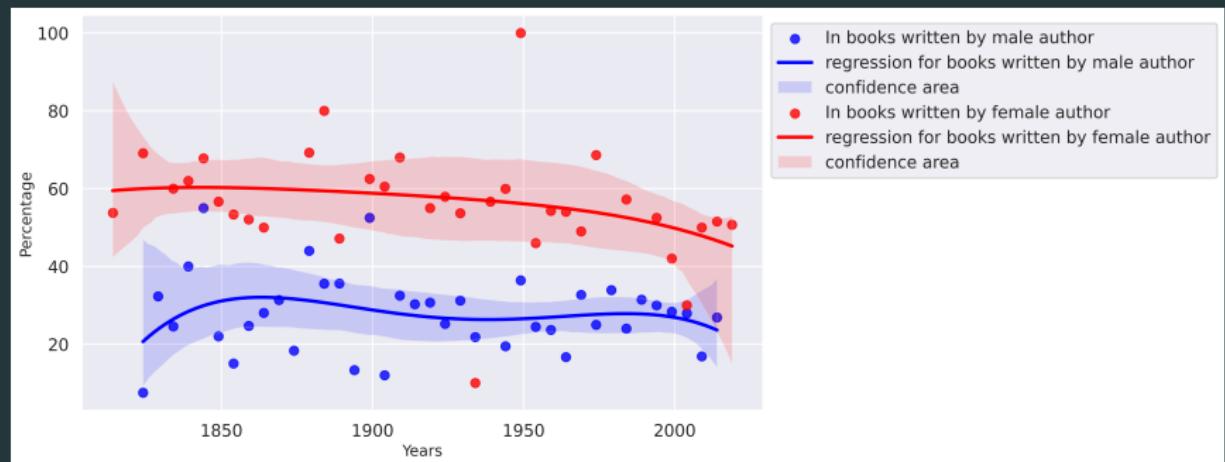
# Distribution des personnages Hommes et Femmes

En fonction du genre de l'auteurice (Dhai Intensive Week 2022)



# Distribution des personnages Hommes et Femmes dans le temps

En fonction du genre de l'auteurice (Dhai Intensive Week 2022)

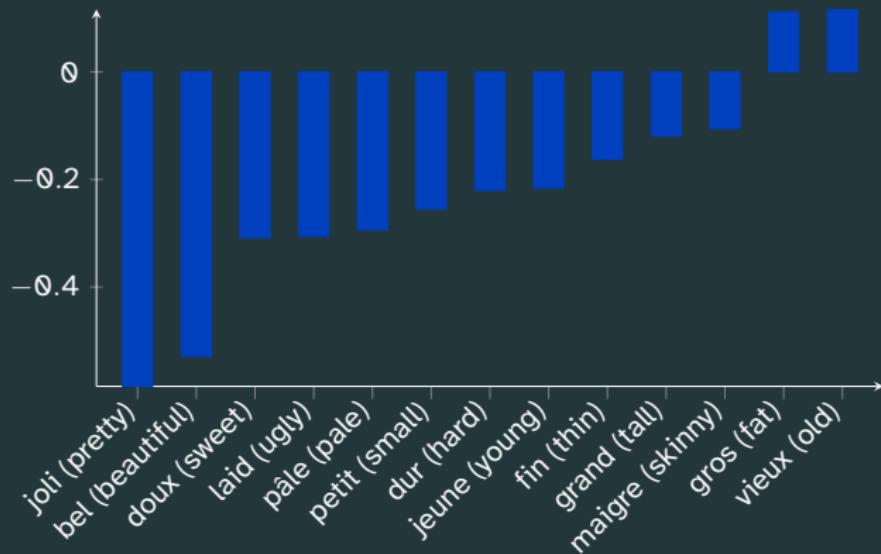


## Distribution of Male and Female Characters

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

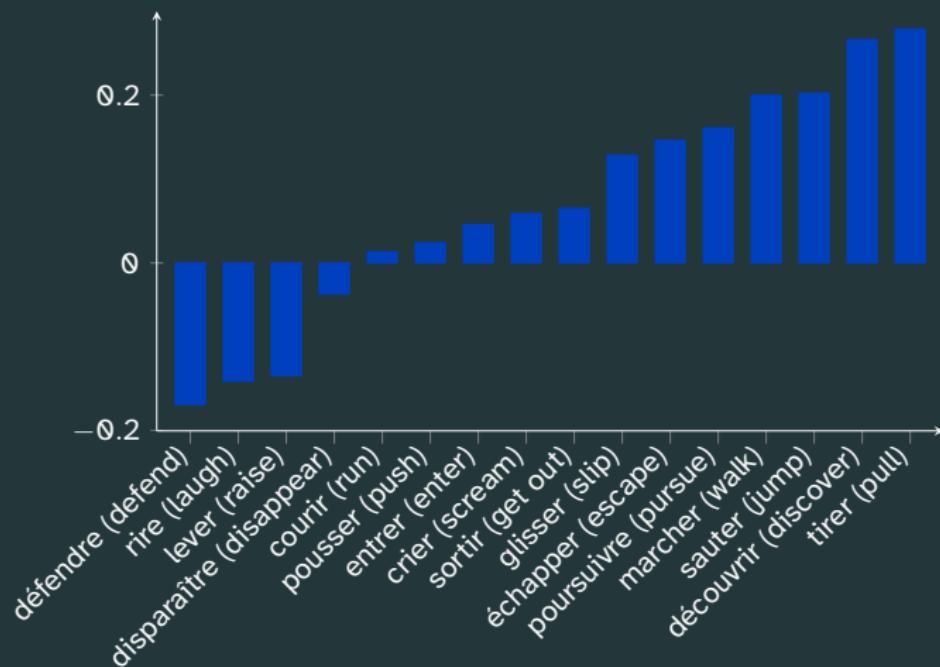
# Stéréotypes de Genre **CHR2023** :

## Adjectifs Associés avec les Personnages Féminins et Masculins



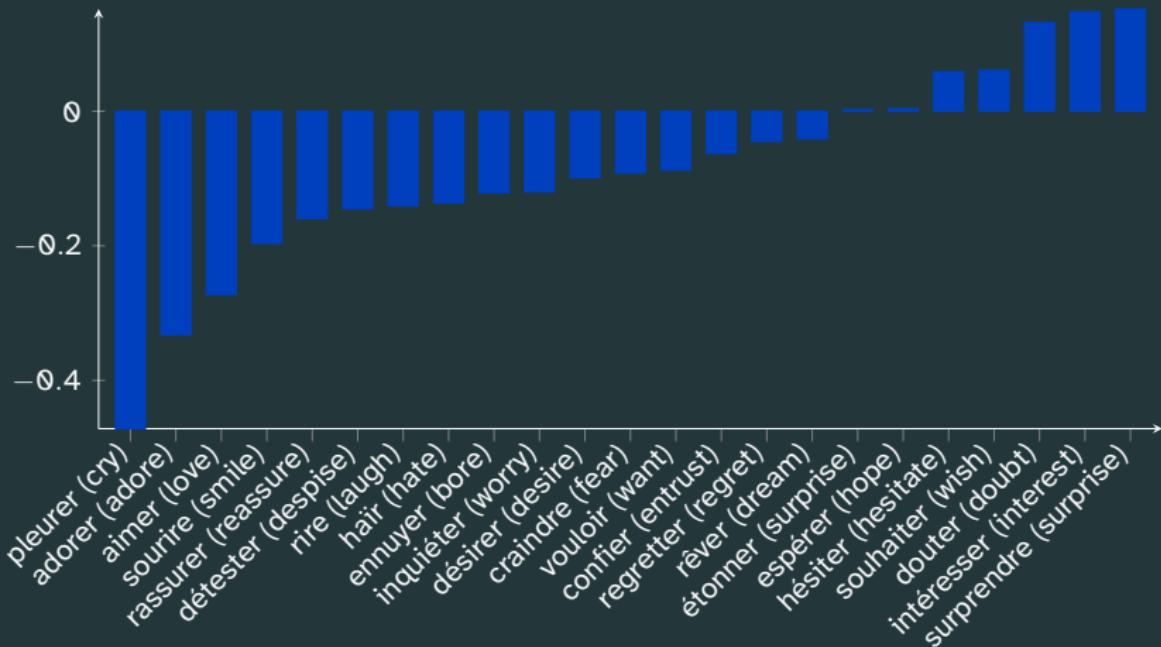
# Stéréotypes de Genre **CHR2023** :

## Verbes d'Action Associés avec les Personnages Féminins et Masculins



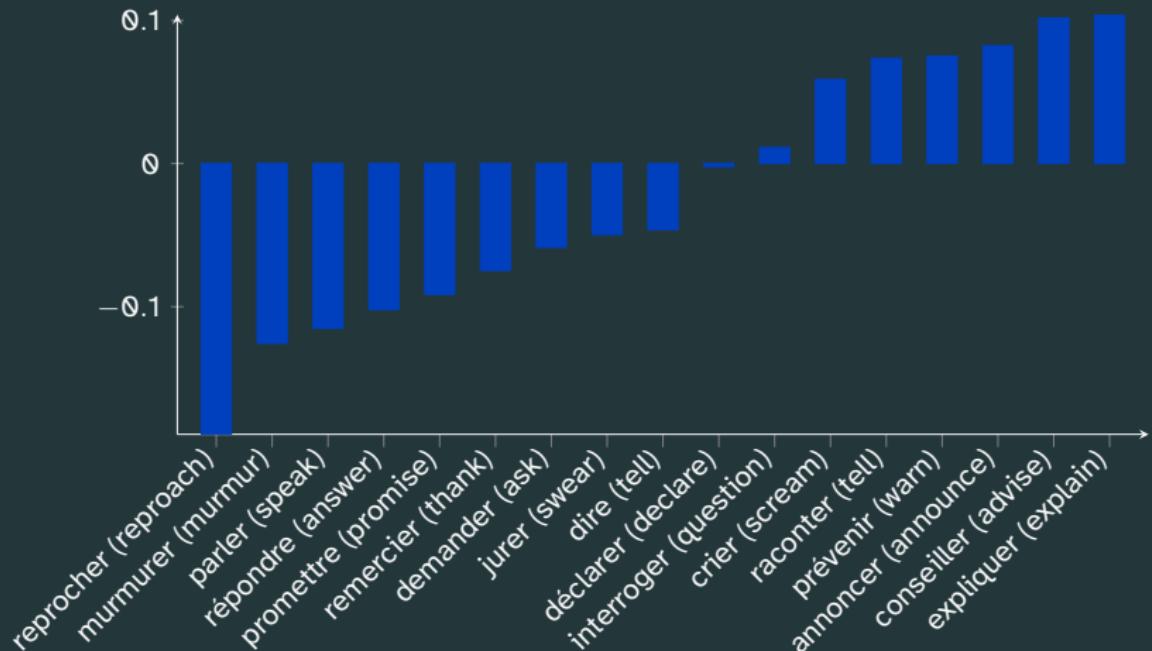
# Stéréotypes de Genre **CHR2023** :

## Verbes d'Émotions Associés avec les Personnages Féminins et Masculins



# Stéréotypes de Genre **CHR2023** :

## Speech Verbs Associated with Male and Female Characters



## Stéréotypes de Genre **TALN2022** :

### Vocabulaire le plus spécifique pour les auteurs et les autrices

---

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

## Stéréotypes de Genre **TALN2022** :

### Vocabulaire le plus spécifique pour les auteurs et les autrices

---

Male jobs :

O' : 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

## Stéréotypes de Genre **TALN2022** :

### Vocabulaire le plus spécifique pour les auteurs et les auteures

---

War/violence :

O' : 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

## Stéréotypes de Genre **TALN2022** :

### Vocabulaire le plus spécifique pour les auteurs et les auteures

---

Hierarchy/Titles :

O' : **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

## Stéréotypes de Genre **TALN2022** :

### Vocabulaire le plus spécifique pour les auteurs et les auteures

---

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

## Stéréotypes de Genre **TALN2022** :

### Vocabulaire le plus spécifique pour les auteurs et les auteures

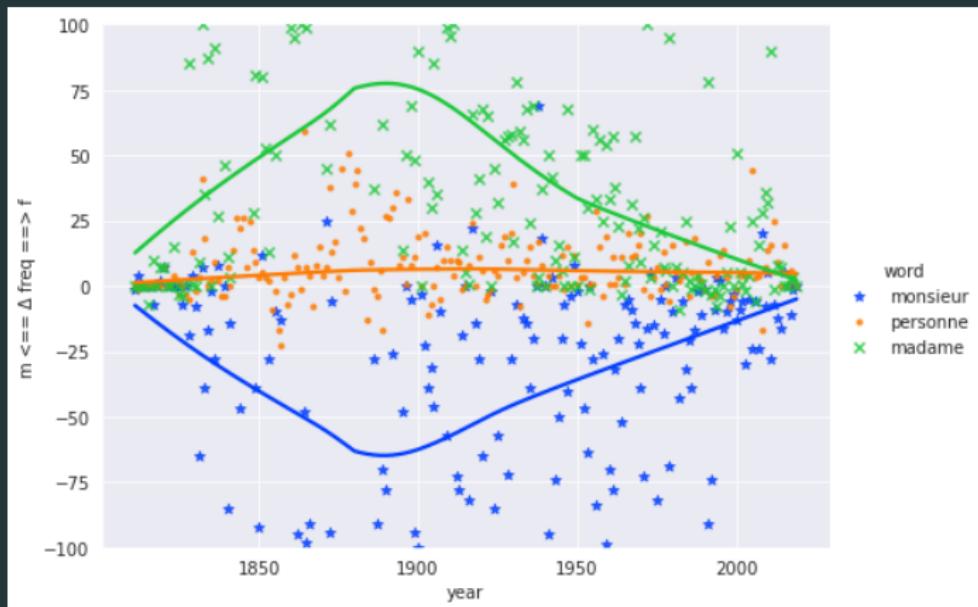
Family/Relations :

♀ : **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

# Évolution des stéréotypes sur le temps

La connotation générée d'un mot donné : Monsieur, Madame & Personne (Dhai Intensive Week 2022)...



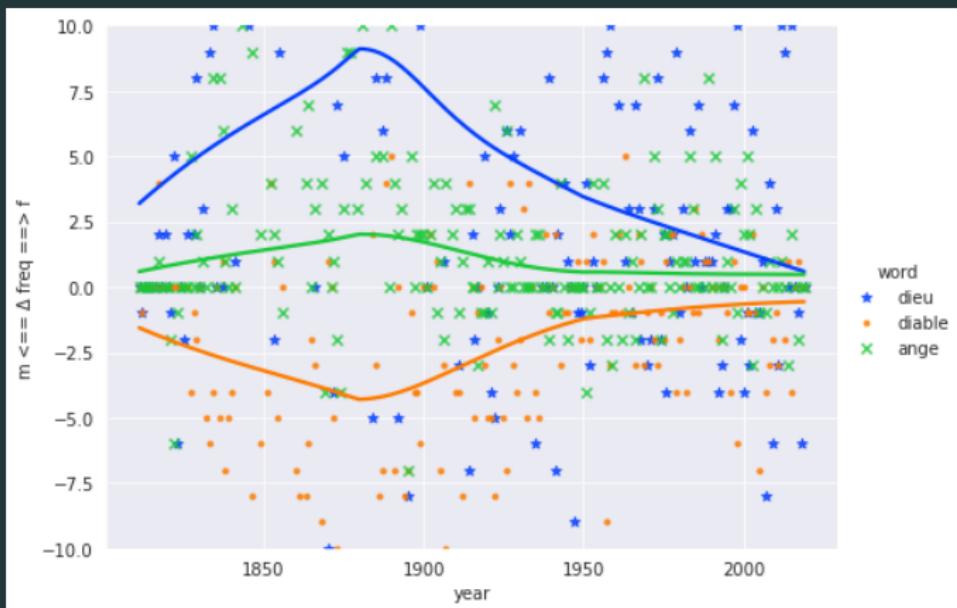
# Évolution des stéréotypes sur le temps

La connotation générée d'un mot donné : Bureau, Maison & Pièce  
(Dhai Intensive Week 2022)...



# Évolution des stéréotypes sur le temps

La connotation générée d'un mot donné : Dieu, Diable & Ange (Dhai Intensive Week 2022)...



## Score d'Association

---

Nous utilisons une différence normalisée pour mesurer l'association entre une population A ou une population B :

$$\frac{A - B}{A + B} \quad (1)$$

## Score d'Association

---

Nous utilisons une différence normalisée pour mesurer l'association entre une population A ou une population B :

$$\frac{A - B}{A + B} \quad (1)$$

Cela varie de -1 à 1, où -1 signifie "associé uniquement à B" et 1 signifie "associé uniquement à A".

## Score d'Association

---

Nous utilisons une différence normalisée pour mesurer l'association entre une population A ou une population B :

$$\frac{A - B}{A + B} \quad (1)$$

Cela varie de -1 à 1, où -1 signifie "associé uniquement à B" et 1 signifie "associé uniquement à A".

Nous utilisons ce score pour mesurer comment les mots ou les relations syntaxiques sont associés au genre.

# Agentivité

---

Estimation de l'agentivité des personnages :

- Capacité des personnages à diriger l'action, prendre des décisions...
- Question : Un genre est-il représenté comme plus 'agent' ?
- Comparaison des fonctions grammaticales dans la phrase
- Analyse syntaxique avec Spacy pour savoir si une mention est *sujet* ou *objet* de la phrase

# Agentivité

---

Estimation de l'agentivité des personnages :

- Capacité des personnages à diriger l'action, prendre des décisions...
- Question : Un genre est-il représenté comme plus 'agent' ?
- Comparaison des fonctions grammaticales dans la phrase
- Analyse syntaxique avec Spacy pour savoir si une mention est *sujet* ou *objet* de la phrase

Score d'agentivité : score d'association avec les dépendances *sujet* ou *objet*.

- **Hommes** : 0,70
- **Femmes** : 0,64

# Travaux Futurs : études longitudinales sur l'utilisation des stéréotypes

- Développer un ensemble de « caractéristiques » pour mesurer la stéréotypie des personnages
  - # traits physiques
  - # mentions d'un emploi ou d'une fonction d'un personnage
  - # différents endroits où les personnages vont
- Être capable de détecter un certain nombre de stéréotypes masculins et féminins dans les romans
  - femme fatale
  - prince
  - mère
  - cow-boy
  - poète sensible
  - vieil homme sage
- Mesurer les changements de stéréotypie des personnages tout au long de la carrière des écrivains et entre les périodes temporelles

# Conclusions

- Fr-BooKNLP nous aide à évaluer comment la caractérisation littéraire est liée aux préjugés de genre.
- Il existe des mots individuels et des champs lexicaux qui sont fortement liés aux stéréotypes de genre.
- La proportion de caractérisation des personnages féminins dépend fortement du genre de l'auteur.
- Les auteurs masculins écrivent moitié moins sur les personnages féminins que les auteures féminines.

# Questions ?

# Méthodes

## Prédiction du genre sans marques de genre grammaticales

- Utilisation uniquement des adjectifs et des verbes (pas de mentions)
- Suppression des marques de genre évidentes :  
*"Je suis heureuse d'y être allée"*  
*"I'm glad I went there"*
- Lématisation

# Annotation

**Table 1** : Examples of gender annotations.

| mentions                      | adjectives                                | g |
|-------------------------------|---|---|
| il, gringoire, lui, la, [...] | mécontent, vide, tranquille, [...]        | m |
| elle, son, jeanne, la, [...]  | libre, grande, nette, prête, lente, [...] | f |
| nous, nos, notre, on, [...]   | heureux, tous, venus, décents, [...]      | p |
| il, je, elle, [...]           | pauvre, aimable, seule, beau, [...]       | u |

## Dependency proportions for agency

**Table 2 : Dependency proportions**

|              | men  | women |
|--------------|------|-------|
| subject      | 0.85 | 0.82  |
| object       | 0.15 | 0.18  |
| agency score | 0.70 | 0.64  |

# corpus preprocessing

