Gender Bias in French Literature

Laurine Vianne[1,2,*], Yoann Dupont[1,3,*] and Jean Barré[1,4]

[1] *Lattice (Langues, Textes, Traitements informatiques, Cognition), 1 rue Maurice Arnoux, Montrouge, 92049, France*
[2] *Université Paris Cité, 85 boulevard Saint-Germain, Paris, 75006, France*
[3] *Université Sorbonne Nouvelle, 17 rue de la Sorbonne, Paris, 75230, France*
[4] *École nationale supérieure - Université PSL, 45 rue d'Ulm, Paris, 75005, France*

Abstract

This study delves into the representation of gender in French literature from 1800 to the present, aiming to assess the prevalence of gender stereotypes in the description of fictional characters. By employing an annotated corpus and statistical modeling techniques, the research explores how authors perpetuated gender biases while shaping characters and narratives. The findings reveal significant linguistic patterns that reinforce gender norms, with women being characterized by emotional and physical attributes, while men are associated with action and agency.

Keywords

french literature, bias, gender, automatic classification,

1. Introduction

Gender, as a social construct, has long played a significant role in shaping literary works. In recent years, the exploration of gender biases and representations in literature has become an increasingly important subject of study in literary scholarship. This paper delves into the gender bias prevalent in French literature from 1800 to now, exploring how authors perpetuated gender stereotypes while shaping their characters and narratives.

This work falls in the field from the Computational Literary Studies (CLS), which offers the ability to process extensive digitized texts in a matter of hours. Researchers can now engage in "distant reading", as proposed by Moretti [14], experimenting with the textual content of literary works. This approach allows scholars to zoom in and out of the literary past, gaining a deeper understanding of the general trends that describe the evolution of literature over time. In the realm of CLS, the practical use of the dichotomy of masculine and feminine is prevalent. Despite its somewhat reductive nature, scholars such as Koolen [11] argue that this approach facilitates the examination of gender roles and representations on a larger scale.

CHR 2023: Computational Humanities Research Conference, December 6 – 8, 2023, Paris, France

*Corresponding author.
laurine.vianne@gmail.com (L. Vianne); yoann.dupont@sorbonne-nouvelle.fr (Y. Dupont); jean.barre@ens.psl.eu (J. Barré)
https://www.lattice.cnrs.fr/membres/chercheurs-ou-enseignants-chercheurs/yoann-dupont/ (Y. Dupont); https://crazyjeannot.github.io/ (J. Barré)
0000-0001-8997-6189 (Y. Dupont); 0000-0002-1579-0610 (J. Barré)

Consequently, this research employs the masculine/feminine dichotomy to capture prevalent gender constructs in French literature.

Gender bias in fictional characterization refers to the systemic tendency to portray female characters in a derogatory or disadvantageous manner within works of fiction. This bias can manifest through various means, such as reinforcing stereotypes, limiting agency or complexity, or presenting characters in a demeaning or objectifying light based on their gender. In the seminal study conducted by Jockers and Kirilloff [10], the authors explored these biases by investigating the association between verbs and gendered pronouns and nouns in characterizations. They concluded that gender serves as a confounding factor in character agency and behavioral representations. Based on these findings, Underwood, Bamman, and Lee [17] revealed a disproportionate representation of male characters by male authors, whereas female authors tended to provide a more equal representation of both genders. The key finding indicated a decline in gender-based distinctiveness of fictional characters over time, suggesting a blurring of gender boundaries in contemporary fiction.

Following this line of research, the aim of this study is to identify consistent patterns of negative portrayals of female characters or the perpetuation of stereotypes, which may indicate gender bias. In alignment with Naguib, Delaborde, Andrault, Bekolo, and Seminck [15], we will concentrate on character space detection [18] through coreference resolution [5]. Our focus will be on evaluating how adjectives and verbs are utilized differently based on the gender of the characters. Building upon Antoniak, Field, Mun, Walsh, Klein, and Sap [1], which demonstrates that detecting power relations, and agency in characterization can unveil gender bias, we aim to elucidate how gender modifies portrayals in novels. This approach is anticipated to offer a more precise and insightful understanding of how gender representation unfolds within the context of literary works.

The first part of this paper is about the creation of a large annotated corpus using the predictions of a classifier. On this large annotated corpus, three main experiments were made to determine whether characters are described differently depending on their gender. First, we focused on agency using linguistic features: we tried to find if women were more often grammatical objects of the sentence than men. The results are presented in section 3.2. The second experiment, detailed in section 3.1, was to use lemmatized words linked to a character to predict its gender: if it works, it would mean that there are 'gendered' words that help the classification. Finally, we explored the vocabulary in more details to find out which words were more associated to women or to men in section 3.3.

2. Materials and methods

2.1. Corpus

We used the Chapitres[1] corpus for our analysis. It is comprised of 2942 French novels in XML-TEI[2] (Text Encoding Initiative) encoding from 1811 to 2020, put together in the context of the

[1] *Chapters* in French.

[2] TEI Consortium, eds. TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 1.0. TEI Consortium. http://www.tei-c.org/Guidelines/P5/.

french ANR[3] project "Chapitres".[4] An overview of the "Chapitres" corpus is given in table 1. Sentences and tokens were computed using the Spacy [8] fr_dep_news_trf model.[5].

Table 1
The number of documents, sentences and tokens (using Spacy) in the "Chapitres" corpus.

century	# documents	# sentences	# tokens
19	1519	1.98M	51M
20	1138	1.56M	34M
21	292	401K	9.3M

2.2. Annotation

A first part of this work was to manually annotate the gender of the 10 main characters from 100 novels picked randomly in the corpus Chapitres. The four classes are:

- 'm' for masculine characters
- 'f' for feminine characters
- 'p' for plural characters
- 'u' (undefined) when there were not enough clues

A 1000 lines (one for each character) tsv file was created. To each character, the most frequent mentions, and related adjectives and verbs were associated. (see section 2.3 for the extraction of related adjectives and verbs). A mention of a character is a linguistic item referring to this given character, notably proper names and pronouns.

Annotation was performed by three annotators who are native speakers. The gender was attributed depending on the mentions (pronouns, names, ...), and the gender of the adjectives. The latter are useful for first person narrators, as mentions do not carry gender marks. The dataset was split evenly across all annotators to ensure that every character would be annotated twice. This allowed us to compute an inter annotator agreement (IAA) using NLTK's [13] implementation of Cohen's κ [6]. Overall κ on the corpus was 0.7256. The annotations were afterwards adjudicated to provide a gold standard. While considered "substantial" by some [12], others recommend a higher κ of 0.8 [2], using 0.67 as lowest acceptable value. We then consider our IAA as acceptable. In table 2 are presented four examples of annotations, one for each class. In table 3 are presented the results of the annotation.

2.3. Textual features

For each character, the goal was to obtain all its mentions in the book, and all adjectives and verbs having a dependency link with a mention of this character. A first step was to associate

[3]FNA, *French National Agency*
[4]Website of the the project: https://chapitres.hypotheses.org
[5]The model is described at the following URL: https://spacy.io/models/fr#fr_dep_news_trf

Table 2
Examples of gender annotations. "g" is for "gender". Word translations are given in appendix A.1

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

Table 3
Distribution of the classes over the 1000 manually annotated characters. The meaning of tags are given in section 2.2.

measure	m	f	p	u
count	531	322	55	92
percentage	53%	32%	6%	9%

Table 4
Example of a parsed sentence in our format. Some headings had to be abbreviated to fit the paper. "Idx" means "index", "ChId" means "character id", "G" means "gender". The sentence translates to "Renaud caresses me with one of these intelligent glances that bring me back to him."

Idx	Word	Lemma	POS	Head	Dep	ChId	G
0	Renaud	Renaud	PROPN	2	nsubj	4	m
1	me	me	PRON	2	obj	1	f
2	caresse	caresse	VERB	2	root	0	u
3	d'	de	ADP	4	case	0	u
4	un	un	PRON	2	obl:mod	0	u
5	de	de	ADP	7	case	0	u
6	ces	ce	DET	7	det	0	u
7	regards	regard	NOUN	4	nmod	0	u
8	intelligents	intelligent	ADJ	7	amod	0	u
9	qui	qui	PRON	11	nsubj	0	u
10	me	me	PRON	11	obj	1	f
11	ramènent	ramener	VERB	7	acl:relcl	0	u
12	à	à	ADP	13	case	0	u
13	lui	lui	PRON	11	obl:arg	4	m

each token with 2 informations: is this token a character, and what gender is this character. To store these informations, we created a tsv format with the columns presented in table 4.

By using an NLP pipeline specifically tuned for novels, (fr-BookNLP, part of the multilingual BookNLP project [4, 3]), we extracted literary characters along with all their mentions, thus resolving their coreference.[6]

This NLP pipeline allowed us to associate column 7 (character id) to the first column. The 10 most important characters of each book were considered, thus column 7 contains character ids from 1 to 10 when the token is a mention of a character. Then, all sentences from our corpus were parsed with Spacy [8, 9] to obtain columns 2 to 6. For the last column, we had to manually annotate the gender (as described in section 2.2), or we used the predictions (see section 2.4) for the unannotated part.

[6]see appendix A.2 for the evaluation of Fr-BookNLP

In table 4, a sentence from *Claudine en ménage* by Colette is presented. The character with id 1 was annotated as a female and the one with id 4 as a male, as can be seen in the last two columns. 'Renaud' is the subject of the sentence. Character one's first mention is 'me', which is the object of the sentence (see 6th column).

In a second step, the information given by Spacy was used to define the lists of related adjectives and verbs. For each character, we retrieve all tokens that are linked to a mention of this character (each token whose index correspond to the head that governs the mention), if these tokens are adjectives or verbs.

To represent the features (the adjectives, verbs or mentions related to characters) as vectors, we decided to use a Bag of Words (BoW) representation. Thus, each line corresponds to a character, and each column represents a word. In each cell is found the number of occurrences of the word corresponding to the column for the character given by the line, divided by the total number of words associated to this character. This allows to have the relative frequency for each word. Because of the over-representation of masculine characters in our corpus (see table 3), we decided to balance the classes.

2.4. Building a classifier for gender predictions to annotate the corpus

We trained a classifier to predict the gender of characters and used these predictions as annotations for our experiments described below. We employed a Random Forest classifier [7], implemented in scikit-learn [16] to predict character gender. We ran our model in a basic 5-fold cross validation set up. Classes 'u' (undefined) and 'p' (plural) were removed to make the predictions easier.

Using a Bag of Words representation of the most frequent character mentions as input, the classifier proved effective in predicting gender based on crucial words such as 'il' (he) and 'elle' (she), as depicted in Figure 1.

We then inferred gender across all our corpora, which encompasses a vast array of 29,490 characters, holds significant importance in unveiling broader patterns and trends in gender representations within the literary landscape. By examining a wide range of characters from various genres, time periods, and cultural backgrounds, we aim to discern recurrent patterns and subtle nuances in character depictions, shedding light on the implicit gender constructs prevalent throughout literary history.

2.5. Predicting gender without grammatical gender marks

Following Underwood, Bamman, and Lee [17], we tried to reproduce on our corpus of French novels some of their findings. For this experiment, we use a similar method to the one used to annotate the corpus, described in 2.4. We adapted the features used to predict gender: instead of using mentions, we used adjectives and verbs.

A major difference between French and English that we had to take into account for this experiment is the gender mark on certain adjectives and participles. For example, in the French sentence below there are two signs that the speaker is feminine, while there are none in the English equivalent:

> "Je suis heureu**se** d'y être allé**e**"

"I'm glad I went there"

If a classifier is able to correctly predict whether a character is feminine or masculine due to these marks, there is no evidence that gender stereotypes helped the classification. Therefore, we chose to use lemmas instead of word forms. Thus, in the example above, the gendered words would become 'heureux' and 'aller'. This way, if the classifier achieves an accuracy which is significantly higher than the baseline, it will be because it used sociological features instead of grammatical ones.

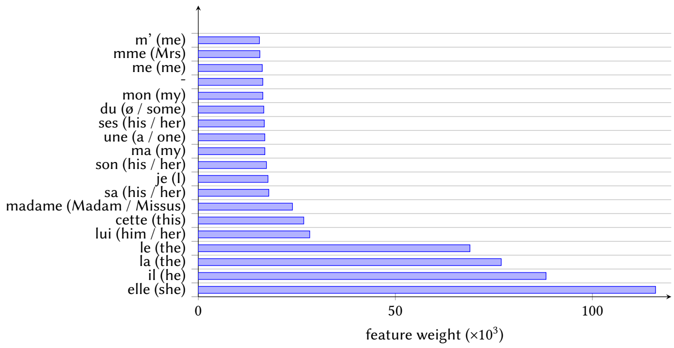

Figure 1: Most important features for gender prediction. In parentheses are their English translations.

2.6. Agency

In a literary setting, character agency refers to the capacity of a character to take intentional and autonomous actions within the narrative. It is the degree to which a character is portrayed as an active agent who drives the events and makes choices that impact the story's progression. Characters with low agency may be more passive or reactive, influenced by external forces or events rather than actively driving the story. They might have limited control over their circumstances or be more prone to being acted upon rather than taking independent action. To evaluate a character's agency, we compared the number of mentions of this character that are subject of the sentence, and the ones that are object. To retrieve this information, we used Spacy as dependency parser.

To compare the degree of agency of different characters, we created a metric. The *agency score*, between -1 and 1, corresponds to:

$$\frac{nsubj - obj}{nsubj + obj}$$

With *nsubj* the number of occurrences of a character or group of characters as subject of the sentence, and *obj* the number of occurrences of a character or group of characters as object of the sentence.

The higher the score, the more often the character (or group of characters) holds the position of subject of the phrase, and thus has an active role.

Table 5
Accuracy of gender prediction using random forests

features	accuracy
mentions (lemmatized)	0.9130
adjectives describing character	0.7827
adjectives (lemmatized)	0.6025
+ verbs (lemmatized)	0.6118

If we consider the tiny example of table 4 described in section 2.3, character one's agency score would be:

$$\frac{nsubj - obj}{nsubj + obj} = \frac{0 - 2}{0 + 2} = -1$$

3. Results

3.1. Character gender prediction

3.1.1. For the corpus annotation

Resulting accuracies range from 60.25% to 91.3%, as described in Table 5. The classifier demonstrated a good accuracy rate when using the mentions as features, achieving 91.3%. The errors are mostly due to inaccuracies in the identification of characters by BookNLP: it happens sometimes that a character has both feminine and masculine mentions because two different characters were mixed up. However, we considered that 91.3% accuracy was enough to annotate the whole corpus with this method.

3.1.2. For the bias analysis

In our first experiment (second line in the table), adjectives are used to predict the gender of the characters they describe. We ran the same classifier using lemmas instead of words for comparison (third line of the table). The accuracy using words is 0.782, which is higher than the 0.602 accuracy obtained with lemmas This may indicate that the classifier relied on grammatical gender marks when adjectives were not lemmatized.

It is also interesting to notice that the results are better when using only adjectives when word forms are used (0.782 with adjectives against 0.74 with adjectives and verbs), whereas it is the contrary when using lemmas (0.602 with adjectives against 0.612 with adjectives and verbs). A possible explanation is that adjectives bear a gender mark more than verbs. This would mean that the classifier relies a lot on these gender marks when it can.

Now, the results of the classifier using lemmas are interpretable. In table 6 is presented the confusion matrix obtained with the BoW representation containing lemmatized adjectives and verbs as input to a Random Forest. Because the two classes *male* and *female* were balanced, the baseline is 0.5. The classifier is able to predict the gender of a character given the adjectives and verbs related to it with an accuracy of 0.612. The features that were most useful for the classifier

Table 6
Confusion matrix for Random Forest with lemmatized adjectives and verbs related to the characters

		predicted labels	
		f	m
true labels	f	203	119
	m	131	191

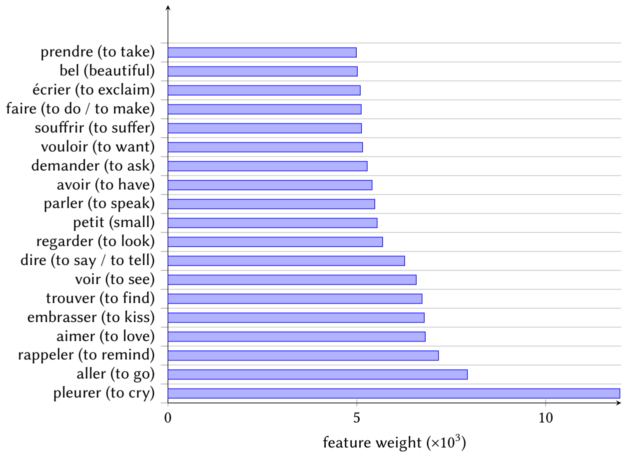

Figure 2: Most important lemmatized verbs and adjectives for gender prediction. In parentheses are their English translations.

Table 7
Dependency proportions

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

to classify the characters are presented in figure 2, and will be discussed later. Eventually, this result indicates that male and female characters are not characterized in the same way.

3.2. Agency

The results of our computations show that masculine characters tend to be a little bit more 'agent' than feminine ones, but the difference is not striking. Indeed, we can see in table 7 that men are more often subjects than women (85% of the time against 82%) while they are less often objects, but their agency scores are close. However, the small difference in the agency scores might imply that the descriptions of male characters exhibit slightly higher agency than those of female characters.

3.3. Gender biases in the vocabulary

Using the method described in section 2.3, we were able to obtain the verbs and adjectives having a dependency link with a given character. From this vocabulary, our goal was to determine whether some words were associated to one gender more than to the other.

As seen in figure 4, we can see that a common stereotype is to associate women to love and passion, and men to action. We wanted to investigate about such well-known stereotypes, to see if they also appear through words in the corpus. A simple and significant example is the verb 'aimer' (to love): we found that this verb occurs at least once for 44% of the female characters, against 25% for the male characters.

To visualize the results, plots were made with, for each verb, whether it is more associated to women or to men in the corpus. More precisely, the values correspond to:

$$\frac{masc - fem}{masc + fem}$$

With *masc* the proportion of masculine characters to which the given verb is related in the corpus, and *fem* the proportion of feminine characters. Thus, the more positive the values are, the more the word is associated to men, and negative with women. To avoid imprecision due to lack of data, verbs occurring for less than 1% of the characters were not considered.

3.3.1. Gendered differentiation through action verbs

This experiment was, among others, used to have another approach of agency, more focused on the lexicon than on the dependencies. Two kinds of verbs were used: action verbs (figure 3), and emotion verbs (figure 4). The lists are quite arbitrary: the categories 'action verb' and 'emotion verb' are not fixed. However, the list of action verbs contains verbs that show an influence on the narration. We considered that the more a character or a gender was related to these verbs, the more he is an agent of the narration and plays an active role in it. Conversely, emotion verbs are more static and focus on the mind.

When comparing figures 3 and 4, it appears that most of the 'action verbs' are considered as more masculine, and most of the 'emotion verbs' occur more for feminine characters than for masculine ones. Among these verbs conveying an emotion, those related to love are clearly more feminine: 'aimer' (to love), 'adorer' (to really like, to adore).

The mean of the proportion of masculine and feminine characters linked to a certain category of verbs or adjectives was computed to compare more precisely the two genders. The results, presented in table 8, correspond (for each category of adjectives or verbs) to:

$$\frac{\sum_{n=1}^{\#of verbs} \% \; of \; men \; or \; women \; linked \; to \; verb \; n}{\# \; of \; verbs \; or \; adjectives}$$

Action verbs occur in average for 3.66% of masculine characters, against 3.45% of feminine characters. The difference is slight; however, it is bigger when looking at the emotion verbs: 9.11% of men are, in average, linked to an emotion verb, against 12.26% of women.

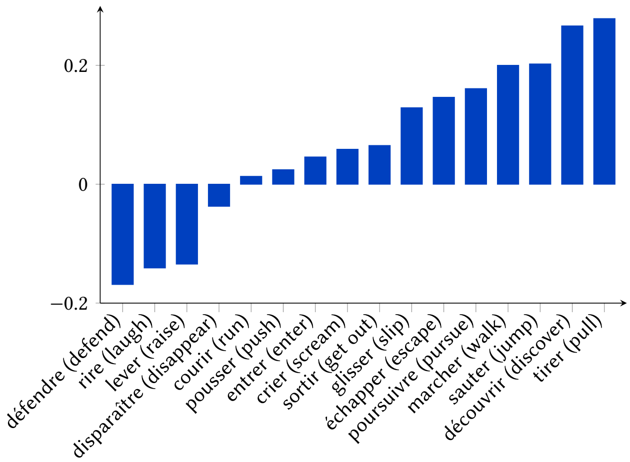

Figure 3: Comparison of the masculinity or femininity of action verbs in the corpus. A positive value indicates a stronger association with male characters, while a negative value indicates a stronger association with female characters. In parentheses are their English translations.

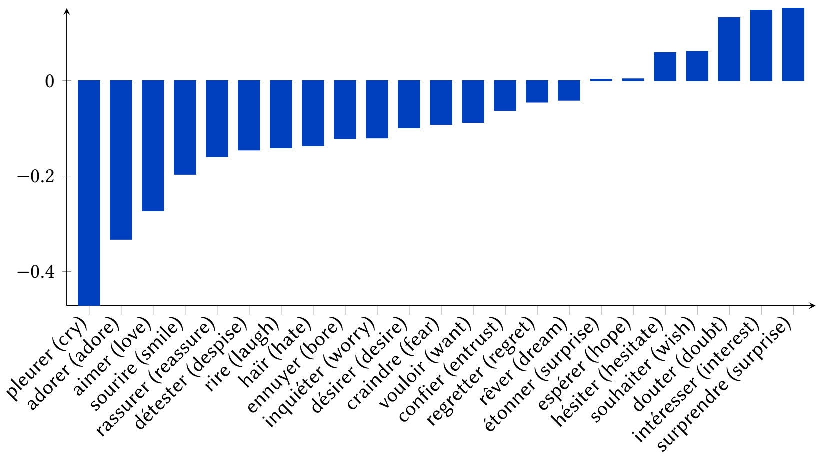

Figure 4: Comparison of the masculinity or femininity of emotion verbs in the corpus. A positive value indicates a stronger association with male characters, while a negative value indicates a stronger association with female characters. In parentheses are their English translations.

3.3.2. Other comparisons

We created a non-exhaustive speech verbs list to see which speech verbs are more associated to a gender than to the other. Figure 5 presents the results. Some interesting observations can be made: first, the verb 'crier' (to shout) is more masculine, as opposed to 'murmurer' (to whisper), which is more feminine. This seems to refer to the stereotype associating women to quietness and men to loudness.

Similarly, gender-related adjectives also contribute to reinforcing stereotypes in French lit-

Table 8
Mean proportion of feminine and masculine characters related to different categories of adjectives and verbs

category	men	women
action verbs	3.67%	3.45%
emotion verbs	9.11%	12.26%
speech verbs	12.01%	13.49%
physical adjectives	6.11%	9.94%

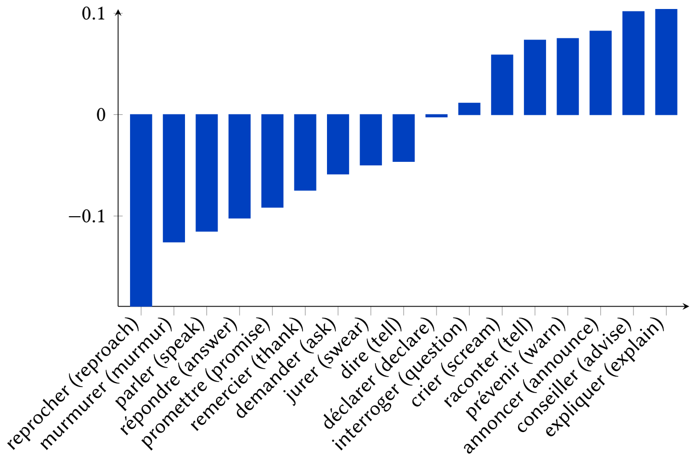

Figure 5: Comparison of the masculinity or femininity of speech verbs in the corpus. A positive value indicates a stronger association with male characters, while a negative value indicates a stronger association with female characters. In parentheses are their English translations.

erature. In figure 6, adjectives such as ”joli” (pretty) and ”doux” (gentle) are commonly used to describe female characters, emphasizing their appearance and nurturing qualities. While adjectives such as ”vieux” (old) and ”gros” (fat) are more frequently applied to male characters, their antonyms ”jeune” (young) and ”fin” (thin) are more associated to women. This opposition seems to refer to the valorization of youth and thinness for women.

4. Discussion

4.1. Analysis of the discriminant features

Considering the model that uses lemmas to predict the gender (third model in table 5) so it cannot use gender evidence from word, we want to discuss here the possibility that these features show gender biases. The discriminant features of our model, presented in figure 2, are the features that helped the most our classifier to predict the gender of a character.

Some of these lemmas are particularly associated to women or to men, as showed in section 3.3. On the first hand, the verb ’aimer’ (to love) is much more associated to feminine characters,

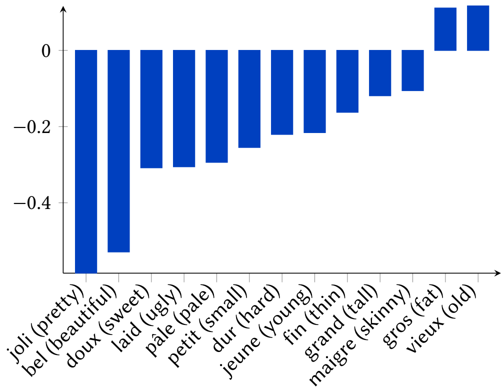

Figure 6: Comparison of the masculinity or femininity of physical adjectives in the corpus. A positive value indicates a stronger association with male characters, while a negative value indicates a stronger association with female characters. In parentheses are their English translations.

and is also the fourth most useful feature to determine if a character is feminine or masculine (see figure 2). Verbs commonly associated with female characters include ”pleurer” (to cry), ”reprocher” (to reproach), ”murmurer” (to murmur), and ”rire” (to laugh). These verbs tend to emphasize emotional expressions and portray women as more emotive and sensitive.

On the other hand, verbs associated with male characters include ”tirer” (to pull), ”découvrir” (to discover), and ”expliquer” (to explain). These verbs often depict actions and intellectual pursuits, implying that male characters are more active and analytical in their roles.

4.2. Limitations of the approach

Our approach is subject to the inherent accuracy limitations of many NLP and machine learning algorithms used, including fr-BookNLP, Spacy, and our Random Forest, all of which are prone to making errors.

As mentioned, the classifier used to predict the gender of the 29,490 characters achieves a 0.91 accuracy on the set of 1000 annotated characters. This means that approximately 9% of the characters on which the experiments were made do not have the correct label.

Most of the time, the errors are certainly due to the imprecision of the retrieved characters. Indeed, when doing the manual annotation, 9% of the characters were classified as ’undefined’ because the mentions referred in fact to different characters, as showed in table 2.

A limitation of this study lies in the fact that the temporal drift of gender bias has not been explored. Our intention was to develop methods to delve into this question using a large corpus and identify the associated challenges. Future research will specifically focus on this aspect, aiming to assess the resistance of these stereotypes to temporal shifts and explore their evolution within fiction.

5. Conclusion

In conclusion, this article focuses on the representation of gender in French fiction, from 1800 to the present day. By examining gender bias prevalent in French literature, the study sheds light on how authors perpetuated gender stereotypes while shaping their characters and narratives. The study highlights the importance of critically examining such representations to foster a more inclusive and diverse literary landscape.

Our statistical modeling reach from 61% to 91% accuracy, depending on the features given to the classifier (with or without obvious gender marks). We employed it to infer the gender of 29490 characters, on which we can run some experiments. This allowed us to study how each gender is portrayed in French literature.

The analysis of French literature revealed certain gender stereotypes that are reflected in the language used to describe female and male characters. These linguistic patterns in the portrayal of characters in French literature reflect underlying gender bias and perpetuate traditional gender norms and roles. For example, we found that women get characterized with far more physical and emotional adjectives and verbs than men. These last were more linked with action verbs, indicating a stronger agency for male characters.

Acknowledgments

This work was funded in part by the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA0001 (PRAIRIE 3IA Institute). Jean Barré's PhD is supported by the EUR (Ecole Universitaire de Recherche) Translitteræ (programme "Investissements d'avenir" ANR-10- IDEX-0001-02 PSL and ANR-17-EURE-0025).

References

[1] M. Antoniak, A. Field, J. Mun, M. Walsh, L. Klein, and M. Sap. "Riveter: Measuring Power and Social Dynamics Between Entities". In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*. Toronto, Canada: Association for Computational Linguistics, 2023, pp. 377–388. DOI: 10.18653/v1/2023.acl-demo.36.

[2] R. Artstein and M. Poesio. "Inter-coder agreement for computational linguistics". In: *Computational linguistics* 34.4 (2008), pp. 555–596.

[3] D. Bamman. *BookNLP*. 2021. URL: https://github.com/booknlp/booknlp.

[4] D. Bamman, T. Underwood, and N. A. Smith. "A Bayesian Mixed Effects Model of Literary Character". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Acl 2014. Baltimore, Maryland: Association for Computational Linguistics, 2014, pp. 370–379. DOI: 10.3115/v1/P14-1035.

[5] J. Barré, P. Cabrera Ramírez, F. Mélanie, and I. Galleron. "Pour une détection automatique de l'espace textuel des personnages romanesques". In: *Humanistica 2023*. Corpus. Association francophone des humanités numériques. Genève, Switzerland, 2023, pp. 56–61. URL: https://hal.science/hal-04105537.

[6] J. Cohen. "A coefficient of agreement for nominal scales". In: *Educational and psychological measurement* 20.1 (1960), pp. 37–46.

[7] T. K. Ho. "Random decision forests". In: *Proceedings of 3rd international conference on document analysis and recognition*. Vol. 1. Ieee. 1995, pp. 278–282.

[8] M. Honnibal and M. Johnson. "An improved non-monotonic transition system for dependency parsing". In: *Proceedings of the 2015 conference on empirical methods in natural language processing*. 2015, pp. 1373–1378.

[9] M. Honnibal, I. Montani, S. Van Landeghem, and A. Boyd. *spaCy: Industrial-strength Natural Language Processing in Python*. 2020. DOI: 10.5281/zenodo.1212303.

[10] M. Jockers and G. Kirilloff. "Understanding Gender and Character Agency in the 19th Century Novel". In: *Journal of Cultural Analytics* (2016). DOI: 10.22148/16.010.

[11] C. Koolen. "Reading beyond the female: the relationship between perception of author gender and literary quality". PhD thesis. University of Amsterdam - Universiteit van Amsterdam, 2018.

[12] J. R. Landis and G. G. Koch. "The measurement of observer agreement for categorical data". In: *biometrics* (1977), pp. 159–174.

[13] E. Loper and S. Bird. *Nltk: The natural language toolkit*. 2002. arXiv: cs/0205028.

[14] F. Moretti. "Conjectures on world literature". In: *New left review* (2000), pp. 54–68.

[15] M. Naguib, M. Delaborde, B. Andrault, A. Bekolo, and O. Seminck. "Romanciers et romancières du XIXème siècle : une étude automatique du genre sur le corpus GIRLS (Male and female novelists : an automatic study of gender of authors and their characters)". In: *Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Atelier TAL et Humanités Numériques (TAL-HN)*. Avignon, France: Atala, 2022, pp. 66–77. URL: https://aclanthology.org/2022.jeptalnrecital-humanum.8.

[16] 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". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[17] T. Underwood, D. Bamman, and S. Lee. "The Transformation of Gender in English-Language Fiction". In: *Journal of Cultural Analytics* (2018). DOI: 10.22148/16.019.

[18] A. Woloch. *The One vs. the Many*. Princeton University Press, 2003.

Table 9
translation of table 2

mentions	adjectives	g
he, gringoire, him/her, the, [...]	discontent, empty, tranquil, [...]	m
she, his/her, jeanne, the, [...]	free, tall, clear, ready, slow, [...]	f
we/us, our (plural), our (singular), we, [...]	happy, all, "those who came", decent, [...]	p
he, I, she, [...]	poor, kind, alone, beautiful, [...]	u

A. Appendix

A.1. Translation of table 2

A.2. Corpus distribution

Figure 7 shows the distribution of novels over time broken out by author's gender. The *Chapitres* corpus spans from 1810 to 2020, with a notable concentration of novels in the latter half of the 19th century. The proportion of female writers remains between 15 and 20% of the literary production, with a slight increase towards the end of the corpus, reaching about 30%. The corpus includes 714 authors, with 149 (21%) being female writers, contributing to a total of 626 novels (27%).

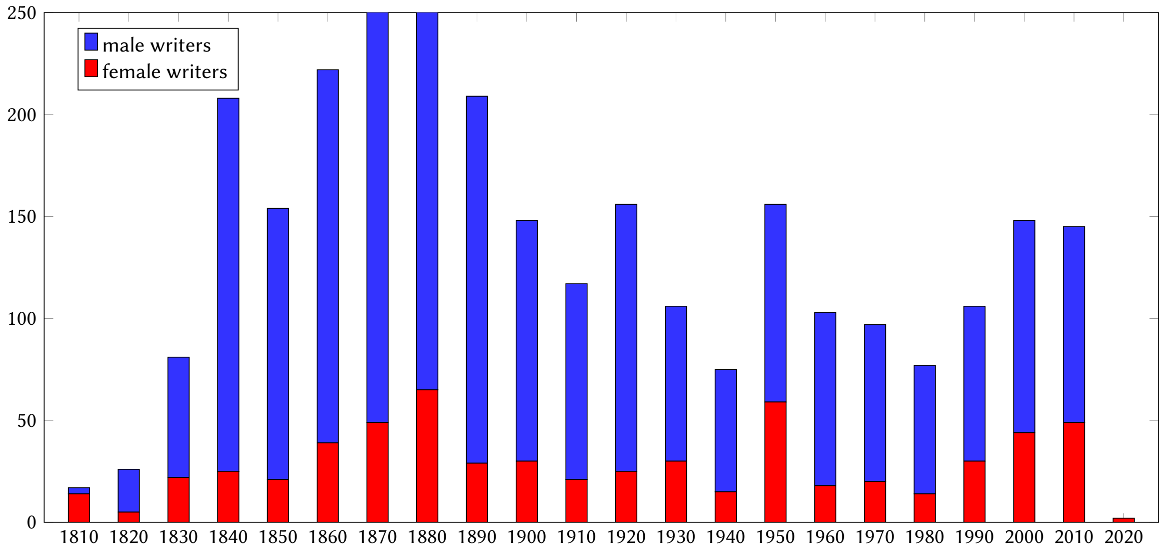

Figure 7: Distribution over time of the proportion of novels written by women writers

A.3. Fr-BookNLP evaluation

We evaluated Fr-BookNLP coreference pipeline using three metrics used in state of the art coreference evaluation. All three evaluate the correctness of coreference chains and how well they match the ground truth, at different levels (mentions, clusters, alignment).

None of them are really relevant for coreference at the novel's scale, so we followed [3] coreference evaluation, averaging those three metrics. This version of Fr-BookNLP is ten points behind English evaluation.

Table 10
Evaluation of the fr-BookNLP pipeline, scores for character mentions detection and coreference resolution

Metrics	F_1	
PER_mentions	89.3	
MUC	85.1	
B^3	67.1	*Average* 65.2
$CEAF_e$	43,5	

Coreference resolution is still hard for long texts. An issue is that characters can get duplicated, that is, the same character is detected in multiple coreference chains. While problematic at first glance, our primary objective is not to retrieve unique and distinct characters but a proxy for characterization as a whole. The focus is more on character's gender representation and the overall impact of these characters on the literary landscape, rather than identifying separate and non-repeating characters.