



UNIVERSITÉ  
LIBRE  
DE BRUXELLES

TP4 - STIC B545

« Artistes femmes » dans la presse  
belge : Observation de l'environnement  
linguistique et historique autour  
d'elles.

MOKAM NEGUEM Jauline - Matricule : 000558632

Année académique 2023 - 2024

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Méthodologie</b>	<b>3</b>
<b>3</b>	<b>Analyse</b>	<b>4</b>
3.1	Etape 1 : Exploration préliminaire du corpus . .	4
3.1.1	Parcours des fichiers . . . . .	4
3.1.2	Distribution du vocabulaire . . . . .	4
3.2	Etape 2 : Nettoyage du corpus . . . . .	5
3.3	Etape 3 : Enrichissement du corpus . . . . .	5
3.3.1	Extraction des mots clés . . . . .	5
3.3.2	Nuage de mots . . . . .	6
3.3.3	Reconnaissance et extraction d'entités nom- mées . . . . .	7
3.3.4	Analyse de sentiment . . . . .	8
<b>4</b>	<b>Observations</b>	<b>8</b>
<b>5</b>	<b>Conclusion</b>	<b>10</b>
<b>6</b>	<b>Bibliographie</b>	<b>10</b>

# 1 Introduction

Le domaine de l'Art est un secteur qui était en plein éveil dans les années 1900, spécialement pour les femmes qui cherchaient à se distinguer et à influencer les mouvements culturels et artistique de cette époque. Dans ce travail, nous nous intéressons à la thématique des « artistes femmes » en Belgique du point de vue de la presse belge. Nous souhaitons à l'aide du corpus CAMille avoir un aperçu de l'environnement et l'évolution historique de celles - ci. Nous nous servirons de différentes techniques vues en cours pour extraire le maximum d'informations pertinentes à notre analyse.

## 2 Méthodologie

Notre intérêt est de savoir qu'elle place les artistes femmes occupaient dans la société belge. Nous nous sommes axés sur l'analyse de ces trois points :

- Des noms d'artistes femmes et leur domaine d'art
- Leurs œuvres ou des événements marquants (évolution historique)
- Les discours journalistiques utilisés : tournures lexicales, vocabulaire, les sentiments autour des articles et textes écrits à leur sujet (perception)

Afin d'extraire des articles pertinents pour la thématique qui nous intéresse, nous avons utilisé la requête : "artistes,femmes"-4, qui se traduit littéralement par : liste de tous les documents contenant le mot artiste et femme séparés de maximum 4 mots . Nous n'avons pas mis de délimitation de date ou de journaux. Notre requête nous a fourni 1304 résultats. Parmi lesquels nous avons exporté les 1000 premiers. Dans la suite de notre travail, nous allons effectuer tout d'abord une exploration préliminaire des documents du corpus, faire un nettoyage du corpus et enfin appliquer des techniques d'enrichissement.

## **3 Analyse**

### **3.1 Etape 1 : Exploration préliminaire du corpus**

#### **3.1.1 Parcours des fichiers**

Cette partie concerne la manipulation des données visuelles et quantitatives du corpus. Notre corpus contient des articles allant de (1835 à 1969). Nous avons écrit un script qui nous a permis de recenser le nombre total de document inclus, le classement des articles selon les journaux qui les ont publiés, le nombre d'exemplaires par journaux, la liste des années où nous n'avons eu aucun article publié portant sur le sujet.

Il y a 18 exemplaires du journal L'Avenir du Luxembourg, 36 exemplaires de La Libre Belgique, 179 exemplaires de L'Indépendance belge, 6 exemplaires de L'Indépendance belge (édité en Angleterre), 40 exemplaires du Journal de Bruxelles, 86 exemplaires du Journal de Charleroi, 223 exemplaires de La Meuse, 29 exemplaires du journal Le petit Bleu, 0 exemplaires du journal Le Vingtième Siècle, 24 exemplaire de Vers l'Avenir, 68 exemplaires du journal Le Peuple, 229 exemplaire du journal Le Soir et 8 exemplaires du journal Le Drapeau Rouge

#### **3.1.2 Distribution du vocabulaire**

Nous avons créé une liste de Stopwords qui permet de filtrer les mots vides de notre corpus. Cette liste a été enrichie tout au long du travail. Nous avons eu 229 Stopwords pour notre corpus général. Par la suite nous avons converti l'ensemble de nos 1000 fichiers txt en un seul fichier allcaf.txt. Cette conversion nous a permis de procéder à l'analyse de la distribution du vocabulaire. Après tokenisation du corpus, le nombre de mots contenu dans l'ensemble du vocabulaire de 8155964 mots. Ce diagramme représente les 20 mots les plus fréquents de ce corpus

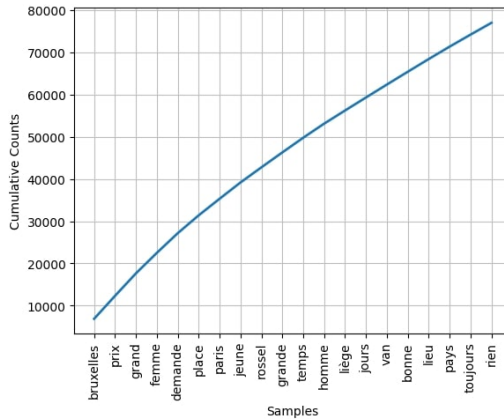


FIGURE 1 – Plot des 20 mots les plus fréquent du corpus

## 3.2 Etape 2 : Nettoyage du corpus

La première étape de notre nettoyage a été de filtrer le corpus, en extrayant les phrases qui comportent les termes de notre recherche. Ce filtrage a éliminé les parties des journaux qui n'abordaient pas le thème. C'est sur ce nouveau corpus que nous avons en étape 2 appliqué la fonction de nettoyage qui consiste à l'élimination des déchets du texte.

## 3.3 Etape 3 : Enrichissement du corpus

### 3.3.1 Extraction des mots clés

Nous avons extrait les mots-clés en forme de bigrammes. Nous avons eu en résultat des mots comme (jeune femme, artiste femme, artiste peintre, grande artiste, etc.) Les bigrammes ont été les résultats les plus pertinents sur l'ensemble des documents, car dans certains cas l'extraction en trigramme nous conduisaient à des associations de mots qui perdent le sens de l'expression.

[('femme d'un artiste', 7.784618554785209e-06),  
 ('femme artiste', 1.490291815594245e-05),  
 ('artiste', 2.0160730625965632e-05),  
 ('d'une femme artiste', 3.101233162821727e-05),  
 ('femme', 4.566545210628855e-05),  
 ('jeune artiste', 8.761273344482263e-05),  
 ('artiste femme', 8.858956903810237e-05),  
 ('artiste peintre', 9.229394615835761e-05),  
 ('d'un artiste', 9.62996238831671e-05),  
 ('jeune femme', 9.805073057224135e-05),  
 ('jeune femme artiste', 0.00010582588283643202),  
 ('ARTISTE PEINTRE FEMME', 0.0001082962875975644),  
 ('grand artiste', 0.00011414361877846362),  
 ('femme d'un', 0.00017127363712333022),  
 ('d'un grand artiste', 0.000214548004938389038),  
 ('d'une femme', 0.00024990114314278264),  
 ('artiste peintre dem', 0.0002562023503192437),  
 ('FEMME ARTISTE PEINTRE', 0.00032488866278729333),  
 ('qu'une artiste femme', 0.00033422558403053325),  
 ('ARTISTE cherche Femme', 0.0003433098082976935)]

FIGURE 2 – Mots clés du corpus filtré

### 3.3.2 Nuage de mots



FIGURE 3 – Nuage de mots

La génération du nuage de mot nous donne un aperçu visuel du thème du corpus, on peut observer les mots les plus fréquents comme « artistes, femmes, talent ... » On y trouve plusieurs adjectifs qualificatifs utilisés dans les articles pour décrire les artistes, des noms de métiers d'art ( Peintre, lyrique, concert, exposition, etc. »)

### 3.3.3 Reconnaissance et extraction d'entités nommées

Ici notre objectif était de pouvoir créer une liste des noms de femmes citées dans les articles, et de pouvoir plus tard analyser leur relation avec le monde de l'art. Pouvoir extraire les différents organisations mentionnées dans notre corpus et les lieux marquants.

- Extraction des entités personnes : Cette liste de nom de personnes extraite est composée des 10 premiers noms triés.

```
moïna apparaît 1 fois dans le corpus
Camille Dela- apparaît 1 fois dans le corpus
Simone May apparaît 1 fois dans le corpus
Marguerite Gobât apparaît 1 fois dans le corpus
K. Maycr apparaît 1 fois dans le corpus
Autoi- liette Dunois apparaît 1 fois dans le corpus
Marguerite Durand apparaît 1 fois dans le corpus
Recettes apparaît 1 fois dans le corpus
Podlesnaia apparaît 1 fois dans le corpus
Upton Sinclair apparaît 1 fois dans le corpus
C'harl-e Chaplin apparaît 1 fois dans le corpus
dfat apparaît 1 fois dans le corpus
ciasse apparaît 1 fois dans le corpus
Kay FRANCIS apparaît 1 fois dans le corpus
```

FIGURE 4 – Quelques exemples d'extraction de noms de personnes

- Extraction des entités lieux : nous constatons ici que Bruxelles apparaît moins de fois que lorsque l'extraction est faite sur le corpus original.

```
Paris apparaît 13 fois dans le corpus
Rome apparaît 5 fois dans le corpus
Bruxelles apparaît 5 fois dans le corpus
Allemand apparaît 3 fois dans le corpus
Anvers apparaît 3 fois dans le corpus
Hollywood apparaît 2 fois dans le corpus
talenl apparaît 2 fois dans le corpus
Opéra apparaît 2 fois dans le corpus
Hollandaise apparaît 2 fois dans le corpus
Berlin apparaît 2 fois dans le corpus
Venise apparaît 2 fois dans le corpus
Gainsborough apparaît 2 fois dans le corpus
Canari apparaît 1 fois dans le corpus
Cinéma OLYMPIA Saint-Nicolas apparaît 1 fois dans le corpus
```

FIGURE 5 – Quelques exemples d'extraction de lieux

- Extraction des entités organisations

```

Académie apparaît 3 fois dans le corpus
BREENDONCK apparaît 1 fois dans le corpus
Breendonck apparaît 1 fois dans le corpus
Pierlingnan apparaît 1 fois dans le corpus
Sûreté générale La bonne foi de la femme apparaît 1 fois dans le corpus
Cambre apparaît 1 fois dans le corpus
Oratoriens de Londres apparaît 1 fois dans le corpus
Officier de la Lésion d'honneur apparaît 1 fois dans le corpus
Académie royale d'Anvers apparaît 1 fois dans le corpus
Société des gens de lettres apparaît 1 fois dans le corpus
Conservatoire royal apparaît 1 fois dans le corpus
société apparaît 1 fois dans le corpus
nois apparaît 1 fois dans le corpus

```

FIGURE 6 – Quelques exemples d'extraction d'organisations

### 3.3.4 Analyse de sentiment

Nous avons sélectionné manuellement 10 phrases du corpus, pour effectuer l'analyse de sentiment. Nous avons principalement choisi celles dont l'OCR était de meilleure qualité, et où un nom d'artiste femme était mentionné. Notre objectif était de percevoir la coloration du discours à l'endroit de ces personnes. Les résultats sont en majorité positifs et non subjectifs.

```

get_sentiment("Mégane Broylant est une jeune artiste qui, après un an et demi /e travail, vient d'ouvrir une exposition dans une modeste salle d
[132] Python
--- This text is 2% positive and 0.20000000000000004% subjective.

get_sentiment("Madeleine Deltoire n'en reste pas moins une véritable artiste vibrante et sincère dont la personnalité ne tardera pas, nous en so
[133] Python
--- This text is 60% positive and 8.7% subjective.

get_sentiment("Françoise Nussy est une femme intelligente, une artiste de talent et l'excellente maman de trois grands garçons.")
[136] ✓ 0.0% Python
--- This text is 50% positive and 0.33749999999999997% subjective.

get_sentiment("Mlle de a Bruchellerie m'ille fois mieux qu'une virtuose : simplement et grandement une artiste «.")
[140] ✓ 0.0% Python
--- This text is 20% positive and 0.35% subjective.

```

FIGURE 7 – Exemples de phrases analysées

REMARQUE : nous avons constaté que lorsque les données textuelles sont nettoyées, il peut y avoir perte de certaines informations, qui sont parfois utiles au traitement.

## 4 Observations

L'objectif de notre travail était de pouvoir relever des informations pertinentes sur la thématique de l'art et des femmes d'Art à travers un corpus de journaux. De cette analyse, nous



avons pu relever des informations utiles. Le discours journalistique renvoie beaucoup de coloration positive autour de ces artistes, cela s'est illustré d'une part par le nuage de mot où nous avons déjà noté la présence de plusieurs adjectif qualificatifs positifs, par la suite par notre analyse de sentiment de quelques phrases extraites aléatoirement du corpus.

De nombreuses artistes femmes se distinguaient à cette époque. En belgique l'éveil et la distinction artistique étaient déjà en marche.

Nous notons également que dans l'environnement de ces femmes à cet époque le combat n'était plus seulement à la mise en lumière et la distinction de leurs oeuvres, car celles ci étaient déjà très ventées par les journaux, mais le combat se trouvait au niveau idéologique où les artistes femmes souhaitaient être plus valorisé en tant que artistes dotées de talents comme tout autres et non que l'attention soit porté à leur genre.

## 5 Conclusion

En conclusion, notre travail portait sur l'analyse du sous corpus de la base de données CAMille orienté sur le thème "artistes femmes" » dans la presse belge, à l'aide de techniques de traitement automatique des langues. Nous avons exploré notre sous corpus, en appliquant des techniques d'extraction de mots clés, de reconnaissance d'entités nommées, d'analyse de sentiments etc. Notre objectif était de pouvoir observer l'environnement historique autour de ces femmes d'Arts à cette époque. Nous avons donc tout au long de notre travail collectés des informations pertinentes sur elles, leurs noms, les perceptions à leurs égard, les problématiques qu'elles rencontraient encore etc. Ce travail nous a permis de mieux s'exercer dans la manipulation de corpus automatiquement et la démarche de recherche et d'exploration d'une thématique. Toutes les techniques de traitement automatique des corpus n'ont pas été utilisé dans cette analyse, mais selon l'approche peuvent permettre d'explorer encore plus les données d'un corpus de texte.

## 6 Bibliographie

- "Femmes artistes en Belgique (1880-1914)" par Joke Spruyt Cet ouvrage explore le rôle des femmes artistes en Belgique à la fin du XIXe et au début du XXe siècle.
- "Femmes artistes en Belgique (de 1830 à nos jours)" par Xavier Canonne Cet ouvrage pourrait offrir une vue d'ensemble de la contribution des femmes artistes en Belgique sur une période plus étendue.