



M2SDL

Rapport du projet de stylométrie

BUTH Richy, DUCÉL Fanny

## **Analyse des poèmes des Mazarinades**

MASTER 2

« Sciences du Langage : Langue et informatique »

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Description des données</b>	<b>3</b>
<b>3</b>	<b>Méthodes employées</b>	<b>4</b>
3.1	Parsing et statistiques sur balises . . . . .	4
3.2	Découpage syllabique . . . . .	4
3.2.1	V2 : Adaptation des vers . . . . .	5
3.3	Analyse de rimes . . . . .	6
<b>4</b>	<b>Résultats</b>	<b>7</b>
4.1	Proportions de textes en vers et en prose . . . . .	7
4.2	Découpage syllabique : types de vers . . . . .	8
4.3	Exemples de vers et leur nombre de syllabe présumé . . . . .	11
4.4	Statistiques de rimes . . . . .	12
<b>5</b>	<b>Conclusion</b>	<b>15</b>
5.1	Perspectives . . . . .	15
<b>A</b>	<b>Annexe</b>	<b>18</b>

## Table des figures

1	<i>Proportion de poèmes en valeur absolue . . . . .</i>	7
2	<i>Proportion de poèmes en valeur relative . . . . .</i>	7
3	Types de rimes pour les bigrammes . . . . .	14
4	Types de rimes pour les trigrammes . . . . .	14

## Liste des tableaux

1	<i>Exemples de découpage syllabique par vers et par poèmes (5 types majoritaires) - V1 vs V2 . . . . .</i>	9
2	<i>Nombre de vers par nombre de syllabes dans tout le corpus - V1 - extrait . . . . .</i>	10
3	<i>Nombre de vers par nombre de syllabes dans tout le corpus - V2 - extrait . . . . .</i>	10
4	<i>Exemples de vers par nombre de syllabes - V1 . . . . .</i>	12
5	<i>Exemples de statistiques de rimes par texte et n-gramme . . .</i>	13
6	<i>Proportions de poèmes en vers et de poèmes en prose par année</i>	18
7	<i>Nombre de vers par nombre de syllabes dans tout le corpus - V1 . . . . .</i>	19
8	<i>Nombre de vers par nombre de syllabes dans tout le corpus - V2 . . . . .</i>	20
9	<i>Exemples de vers par nombre de syllabes - V2 (partie 1) . . .</i>	21
10	<i>Exemples de vers par nombre de syllabes - V2 (partie 2) . . .</i>	22
11	Rimes en bigrammes les plus fréquentes . . . . .	23
12	Rimes en trigrammes les plus fréquentes . . . . .	24

# 1 Introduction

Les Mazarinades sont des textes satiriques ou burlesques publiés à l’époque de la Fronde (1648-1653), principalement pour critiquer, mais aussi parfois pour défendre, le cardinal Mazarin qui menait alors une politique répressive envers les chrétiens protestants de France [Wikipédia, 2022] <sup>1</sup>.

Ces textes étaient très variés et pouvaient aussi bien être des poèmes en vers que des pamphlets et des libelles en prose. Ainsi, il pourrait être intéressant de classifier et d’analyser ces différents textes selon divers critères.

L’objectif de ce projet est donc d’analyser ces Mazarinades pour extraire des sous-corpus et des statistiques, par exemple : quelle proportion de textes est en prose ou en vers, et comment cette proportion évolue diachroniquement ? Quelles sont la structure et la métrique des vers (octosyllabes, alexandrins, ...) ?

Tout notre code et nos données sont librement disponibles sur GitHub : <https://github.com/FannyDucel/projet-stylo-poesie>.

# 2 Description des données

Nous disposons d’un corpus de Mazarinades composé de 3065 fichiers au format .xml, recueillis et formatés grâce au projet ANTONOMAZ [Abiven Karine, 2022] <sup>2</sup>.

Grâce au balisage xml, il est possible d’extraire de nombreuses méta-données sur ces textes, comme leur date de création, le type de poèmes (vers ou prose), des descriptions et remises en contexte de ces textes, les sauts de lignes et fins de strophes, des mots-clés, ou bien encore les bibliothèques d’où sont tirées les textes.

---

1. <https://fr.wikipedia.org/wiki/Mazarinade>

2. <https://github.com/Antonomaz/Corpus>

## 3 Méthodes employées

### 3.1 Parsing et statistiques sur balises

Nous avons utilisé BEAUTIFULSOUP<sup>3</sup> pour parser les documents xml. Cette librairie nous a permis de récupérer les informations comprises entre des balises.

Nous avons calculé la proportions de poèmes en vers et en prose en cherchant la balise *term type="form"* et en comptant leurs occurrences. En les couplant avec la balise *date*, nous avons pu obtenir les proportions pour chaque année. Comme les dates sont au format AAAA-MM-JJ, il a suffi de lire les quatre premiers chiffres à chaque fois pour obtenir l'année.

Nous avons eu des problèmes avec les exceptions, comme des textes sans date ou dans un format différent, comme "JJ MM AAA".

Nous avons cependant pu stocker dans une structure de données tous les textes en vers. Cela prend la forme d'un dictionnaire associant à chaque chemin de fichier en vers sa liste de vers (grâce aux balises *l*). Nous travaillons par la suite uniquement sur ces textes et ces vers.

### 3.2 Découpage syllabique

Nous avons ensuite voulu procéder au découpage syllabique de nos vers. Tout d'abord, nous avons souhaité exploiter le programme utilisé sur le site <http://www.sciencedutexte.fr/versification> et présent sur <https://github.com/humanitesnumeriques/outilDeSyllabation>.

Toutefois, le code est en XQUERY et n'est pas facilement exécutable ou implémentable en Python. De même, la structure HTML nous a semblé trop complexe pour effectuer du WebScrapping. Nous avons alors décidé d'abandonner cette piste : bien que les résultats soient qualitatifs et riches, le coût d'entrée était bien trop important.

Nous nous sommes alors tournés vers le programme FRENCH-SYLLABIFICATION [Ian, 2021] trouvé sur GitHub<sup>4</sup>. Il utilise la base LEXIQUE<sup>5</sup> qui contient le découpage syllabique de 140 000 mots du français moderne.

---

3. <https://www.crummy.com/software/BeautifulSoup/>

4. <https://github.com/ian-nai/french-syllabification>

5. <http://www.lexique.org/>

Pour les mots non-présents dans le Lexique, le programme calcule lui-même le nombre de syllabes des mots.

Ce programme tel quel était très imparfait, tant dans les résultats de découpage que dans sa complexité algorithmique qui le rendait très peu optimisé pour le traitement de gros corpus. Il a donc nécessité des modifications avec l'aide de Gaël Lejeune pour optimiser le code et accélérer le temps d'exécution du programme.

Grâce à ces premières modifications, nous avons donc pu exécuter le code à l'échelle de tout le corpus de vers (634 fichiers). Nous avons également décidé de mettre nos vers tout en minuscules, car les majuscules ne sont tout simplement pas traitées par l'outil et renvoient donc des vers à 0 syllabe ou diminuent d'une syllabe certains vers.

Nous avons ensuite implémenté un premier filtre sur les résultats obtenus (qui constitue notre V1), mais plutôt pour corriger des problèmes liés aux données en elles-mêmes. En effet, certains vers sont composés d'un seul caractère ou contiennent des chiffres ; nous les ignorons.

Nous pouvons ensuite procéder au comptage global des syllabes par vers (nous avons jusqu'à présent un comptage par mot) et obtenons ainsi le nombre de vers par type (nombre d'octosyllabes, ...).

### **3.2.1 V2 : Adaptation des vers**

Après avoir exécuté cette méthode de découpage syllabique une première fois, nous avons remarqué plusieurs problèmes dans les données, liés soit à l'OCRisation (bruit : vers d'un caractère, vers avec un même caractère répété des dizaines de fois, ...), soit à la langue d'époque (dissimilation et s longs notamment).

Nous décidons donc de créer une nouvelle version de notre dictionnaire de vers, en modifiant ces vers pour les adapter aux caractères du français moderne et éliminer un maximum de bruit :

- nous gérons la césure, marquée par le caractère  $\neg$  : si ce caractère est en fin de vers, nous concaténons le vers concerné avec le vers suivant (sa suite) et retirons le caractère, sinon, si le caractère est en milieu de vers, nous le supprimons simplement

- nous éliminons les vers contenant moins de 5 caractères ou tout en majuscules (qui sont en fait des titres et non des vers)
- nous remplaçons les caractères de S longs par des s modernes
- nous remplaçons les v par des u si la lettre suivante n'est pas une voyelle ou si ce v est le caractère final du mot
- nous éliminons les vers contenant au moins 4 lettres identiques à la suite
- nous retirons les espaces en début et fin de vers

Nous comparerons alors les résultats obtenus sans cette amélioration (V1) et avec (V2).

### 3.3 Analyse de rimes

Nous avons aussi créé une fonction simple d'analyse de rimes qui compare les n-caractères de chaque fin de vers en les comptant, puis en les ajoutant à un dictionnaire. Si la rime n'a été comptée qu'une seule fois, alors on le la prend pas en compte, car des rimes se font sur deux vers au minimum.

Nous notons que nous n'avons pas pu utiliser notre découpage syllabique pour cette tâche, car en réalité notre algorithme effectue simplement le comptage des syllabes et non pas une réelle découpe syllabique en renvoyant les syllabes séparées les unes des autres.

Par ailleurs, nous nous intéressons aux types de rimes : suivies, croisées, ou embrassées. Pour déterminer cela, nous réutilisons la fonction précédente en stockant, pour chaque poème, sa liste de rimes. Ensuite, il nous suffit de créer un système de règles simples : si la rime suivante dans la liste est identique, on a une rime suivie, si la rime est identique à la rime qui survient 2 vers plus tard, alors c'est une rime croisée, et si la rime est identique à celle qui survient 3 vers après, alors c'est une rime embrassée. Autrement, il n'y a pas de rime.

Nous notons quelques limites à ce système simple. Tout d'abord, cette analyse serait à réaliser lorsque le travail de nettoyage du bruit sera effectué car actuellement, certains vers ont été ignorés dans nos traitements, ce qui a une influence sur les rimes (il existait peut-être un vers avec la même rime que le vers précédent mais que l'on n'a pas pu compter). De plus, la présence d'une rime embrassée (ABBA) est comptée à la fois comme rime embrassée (pour la partie A) et comme rime suivie (pour B), et une rime est comptée pour deux vers, donc si un poème est écrit uniquement en rimes suivies avec un même son, on comptera autant de rimes que de vers.

## 4 Résultats

### 4.1 Proportions de textes en vers et en prose

Grâce à l'extraction des balises vue dans la section 2 nous pouvons obtenir des statistiques sur les proportions de textes en vers et en prose dans tout le corpus des Mazarinades (voir détail en Annexes, Table 6). Nous en avons également fait des graphiques (voir Figures 1 et 2).

Nous pouvons en conclure que la grande majorité des Mazarinades de notre corpus est en prose, et qu'il semble y avoir une légère tendance croissante pour la prose par rapport aux vers au fil des années (si l'on omet les années contenant trop peu de poèmes, d'où l'intérêt du graphique en valeur absolue).

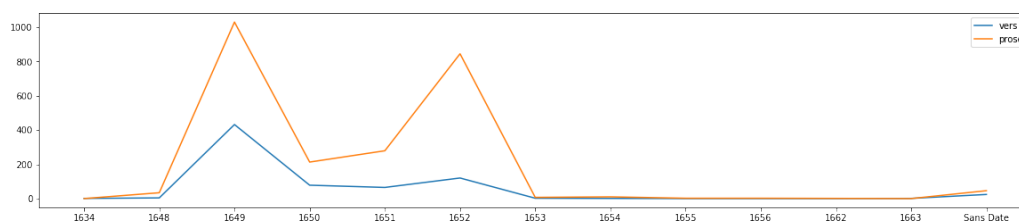


FIGURE 1 – *Proportion de poèmes en valeur absolue*

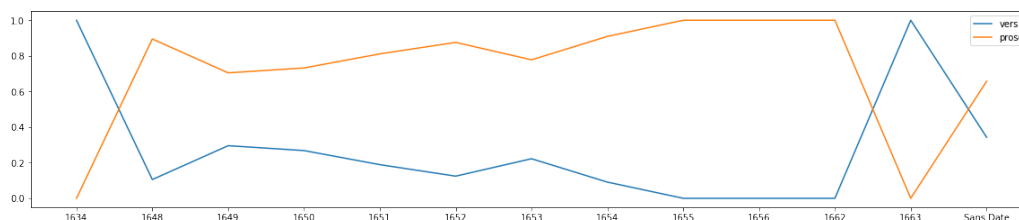


FIGURE 2 – *Proportion de poèmes en valeur relative*



## 4.2 Découpage syllabique : types de vers

De par l'imperfection du découpeur syllabique automatique, il est impossible d'obtenir un texte en vers parfaitement structuré et découpé où l'on pourrait déterminer avec certitude une certaine métrique.

Cependant, la vue d'ensemble nous permet de faire des rapprochements : en effet, on remarque que l'écart de syllabes entre les différents vers est assez minime (voir Tableau 1).

Ce tableau nous permet également de comparer le décompte des syllabes selon notre V1 et notre V2. Nous mettons en rouge les deux types de syllabes les plus fréquents, et en gras le type le plus fréquent. Nous remarquons que, sur ces trois exemples, les résultats diffèrent légèrement selon la version.

Afin de pouvoir évaluer nos versions, nous établissons également une vérité de terrain (*Ground Truth*, notée *GT*). Pour cela, nous lisons rapidement les poèmes concernés et comptons manuellement le nombre de syllabes par vers afin de pouvoir établir le résultat correct et attendu. Nous remarquons que, pour ces trois exemples, nous obtenons une seule bonne réponse, et avec la V2. Toutefois, nous gardons à l'esprit que nous utilisons seulement trois exemples aléatoires, qui ne sont donc pas représentatifs du corpus et sont utilisés à titre d'illustration.

De plus, cette petite tâche d'annotation manuelle nous a permis de mesurer la difficulté de la tâche pour une machine. En effet, certains fichiers (ici, le Moreau1121\_GALL.xml) se sont avérés bruités. Toutefois, ce bruit n'est que légèrement gênant pour un lecteur humain francophone, qui peut facilement le contourner grâce à son intuition de locuteur, mais qui peut fausser largement les résultats d'une machine.

Nous réalisons également ces statistiques sur tout le corpus (voir Tableaux 2 et 3 avec les nombres de syllabes dépassant les 5% et Tableaux 7 et 8 en Annexes, avec toutes les données).

D'après ces résultats, la majorité des vers serait écrite en 7 syllabes, puis en octosyllabes et hexasyllabes. Dans une moindre mesure, nous retrouvons des vers à 9, 5 et 10 syllabes. Il est intéressant de constater que les pourcentages sont assez similaires entre les deux versions.

Cependant, la V2 présente quelques particularités : la proportion d'octosyllabes est plus élevée, et l'on peut estimer que ce seront de bonnes réponses car l'octosyllabe est très présent en poésie française, et, par ailleurs, il y a beaucoup plus de types de vers minoritaires représentés, mais cela est pro-

blement dû au nombre de vers détectés, qui est plus important avec notre V2 puisque moins de vers sont ignorés.

Nous remarquons toutefois que, dans ces deux versions, le type majoritaire est probablement incorrect, puisque les vers impairs sont peu utilisés en métrique traditionnelle. À cela s'ajoutent les méta-données qui nous expliquent le contexte des Mazarinades, notamment qu'il s'agit souvent de poèmes en vers octosyllabiques. Nous pouvons en effet remarquer dans nos exemples que les longueurs de vers avoisinent généralement les 8 syllabes, ce qui confirme notre hypothèse.

Fichier	Nombre de vers par métrique			
	Métrique	V1	V2	GT
Moreau1174_MAZ.xml	5 syllabes	17	9	
	6 syllabes	43	42	
	7 syllabes	<b>71</b>	<b>64</b>	
	8 syllabes	<b>56</b>	<b>71</b>	<b>X</b>
	9 syllabes	18	20	
Moreau1121_GALL.xml	5 syllabes	42	24	
	6 syllabes	<b>116</b>	<b>100</b>	
	7 syllabes	<b>160</b>	<b>172</b>	
	8 syllabes	99	98	<b>X</b>
	9 syllabes	32	41	
Moreau1117_GALL.xml	8 syllabes	13	6	
	9 syllabes	18	19	
	10 syllabes	<b>38</b>	<b>30</b>	
	11 syllabes	<b>31</b>	<b>34</b>	
	12 syllabes	16	22	<b>X</b>

TABLE 1 – Exemples de découpage syllabique par vers et par poèmes (5 types majoritaires) - V1 vs V2

Nombre de syllabes	Nombre de vers	Pourcentage
7	42272	27.46
8	30088	19.55
6	28825	18.73
9	12970	8.43
5	10788	7.01
10	8055	5.23

TABLE 2 – *Nombre de vers par nombre de syllabes dans tout le corpus - V1 - extrait*

Nombre de syllabes	Nombre de vers	Pourcentage
7	42609	27.91
8	34274	22.45
6	23604	15.46
9	13360	8.75
10	8051	5.27

TABLE 3 – *Nombre de vers par nombre de syllabes dans tout le corpus - V2 - extrait*

Bien que ces nouveaux résultats soient encore critiquables, nous décidons d'établir des statistiques sur les types de poèmes selon leur métrique. Cela implique donc de choisir une seule métrique par poème. Pour cela, nous décidons d'adopter la stratégie suivante, basée sur nos hypothèses et intuitions précédentes : la métrique attribuée est celle qui est majoritaire et pair. Ainsi, dans certains cas, un poème n'est pas attribué au type de vers le plus fréquent, mais au second ou au troisième si le premier est impair. Nous obtenons alors les résultats suivants :

- 413 poèmes sont des octosyllabes (8), soit 65%
- 105 poèmes sont des décasyllabes (10), soit 16,5%
- 80 poèmes sont des hexasyllabes (6), soit 12,6%
- 29 poèmes sont des alexandrins (12), soit 4,5%
- 5 poèmes auraient 14 syllabes par vers, soit 0,8%
- 1 poème en aurait 2, 1 poème en aurait 4, soit 0,16% chacun

### 4.3 Exemples de vers et leur nombre de syllabe présumé

Nous reportons ici un exemple de vers par nombre de syllabe trouvé par le découpeur automatique<sup>6</sup>. Celui-ci nous permet de mettre en avant de nouveaux problèmes à résoudre par la suite (voir Tableau 4 pour la V1 et Tableaux 9, 10 en Annexes pour la V2). C'est d'ailleurs grâce à ce tableau de la V1 que nous avons pu établir notre V2.

Les problèmes de la V1 semblaient en effet liés aux données en elles-mêmes plutôt qu'au découpeur syllabique, et nous avons essayé de les résoudre dans notre nouvelle version.

On remarque alors que dans les nouveaux tableaux, on retrouve plus de vers avec un décompte correct. Les nouvelles erreurs semblent alors plus difficiles à comprendre et résoudre, et nécessiteraient une modification encore plus fine et précise des vers (le bruit restant semble difficile à corriger avec des généralisations), ainsi qu'un ajout de règles expertes pour le décompte.

---

6. À noter : dans le tableau, nous remplaçons les symboles de s longs anciens par des S majuscules

Nb syll	Vers
0	e'dtp : :ttéd'
1	theS vertS,
2	ta place,
3	d'eStre criminel, d'eStre fugf,
4	eiiS coront vS.
5	re ne ene nee ene
6	ieu cheueux cendrez, dont meS maiuS vag¬
7	eSpnit Son corpS furent incompatibleS,
8	boiS-robert abbeé de chaStillon.
9	qu'apreS' tant de deSordre, apreS tant de licence,
10	tu viSSeS à teS piedS pleurer teS ennemiS,
11	peut atteindre au Sommet de la pluS haute gloire.
12	voir d'vn oeil de pitié ceux qui t'ont fait laloy,
13	voicy qui vaut norlingue, &l le havre à la fin
14	te leS lS. l'albitre le couail. nouS fappa d'vu Seul coup, le œur l'o¬
15	nn eoi ne enple See nel aS eonie que eoin le elioe peue en ne
16	leoe nunon ue nS oene, nunn une nailS een eril Suei re eei ne ne ene
17	troueois dau la eionuS pui bele ve, fuit Sie dunmen judeareur rerr,
18	iSi eiSre ne leionS ne ie enui nonunonuniS, lenminan SiroSd leeou, nen une
19	ln ei eil deSienSe de nenroi onu in qeu, deS veoenS de dauS ete nei See reS
20	leue uou due oa anure Soirlionail pe pionpe neS le ee que oianu ile,
21	lenS nee enine uni lene de ie eaounS, pdenolunie ni leai le nuuS laur ne ei ei.
22	pue ne iror auienude nS SieleS nerniruble, lueni Senilemeone, panSe que ilS le
23	ntre leS braS deS enS ie tombay demy-morte, tel il faut que tu Sois danS ton aduerSité,
27	le eai Sit ehanuer le te deum pour la pif -de edaratiuSaiuer S'ar dr qe e, furenr onueu au pa¬
31	eeeeeee-e-e-adaaea eae-efaea-eeaoieeae-Scaaaeaéeo-a-eoeeca-a¬¬¬

TABLE 4 – *Exemples de vers par nombre de syllabes - V1*

## 4.4 Statistiques de rimes

Avec notre fonction de comptage de rimes, nous avons pu obtenir des résultats d'occurrences de chaque n-gramme de caractères de fin de vers pour chaque texte (voir Tableau 5).

Si cette fonction très simple ne nous permet pas de détecter les rimes phonétiques, nous pouvons déjà remarquer les cas les plus extrêmes, comme le texte *Moreau13\_GALL* qui est un poème ne comportant que des rimes en ”-eux”.

Texte	Nb Rimes (2 caractères)	Nb Rimes (3 caractères)
Moreau13_GALL.xml	'ux' : 97 'eu' : 3	"eux" : 95
Moreau3081_GALL.xml	're' : 7  'rt' : 3 'ce' : 3 'te' : 2 'uy' : 2	'ire' : 4 'rre' : 2 'ort' : 3 'ice' : 2
Moreau3560_GALL.xml	'le' : 8  'oy' : 5 'eu' : 2 'es' : 2 'al' : 2	'ele' : 4 'lle' : 4 'loy' : 2 'ieu' : 2 'ies' : 2 'nal' : 2

TABLE 5 – *Exemples de statistiques de rimes par texte et n-gramme*

Nous récapitulons également les rimes les plus fréquentes dans les tableaux 11 et 12 en Annexes. Nous remarquons les limites des rimes en n-grammes (surtout bigrammes), qui ne correspondent souvent pas à des phonèmes mais à des bouts de graphèmes. De plus, les espaces sont pris en compte, ce qu'il faudrait rectifier. En outre, les résultats semblent satisfaisants, nous retrouvons des rimes fréquentes en français : -ire, -age, -eux, -eur, -ois, ...

Finalement, nous nous intéressons aux types de rimes (voir Figures 3 et 4).

Pour les rimes en bigrammes, il y a 43 435 rimes suivies (soit 28.45%), 9 882 rimes croisées (soit 6.47%), 6 039 rimes embrassées (soit 3.96%) et 93 322 vers sans rimes (soit 61.12%).

Pour les rimes en trigrammes, il y a 34 484 rimes suivies (soit 22.59%), 7 311 rimes croisées (soit 4.79%), 3 024 rimes embrassées (soit 1.98%) et 107 859 vers sans rimes (soit 70.64%).

Il y a donc une majorité de vers sans rimes, mais pour ceux qui en contiennent, ils sont surtout en rimes suivies, puis croisées, puis embrassées. Cela semble cohérent, car les rimes suivies sont les plus simples à mettre en place, tandis que les embrassées sont les plus complexes. De même, les proportions sont encore moindres pour les rimes en trigrammes, car elles sont plus difficiles à créer.

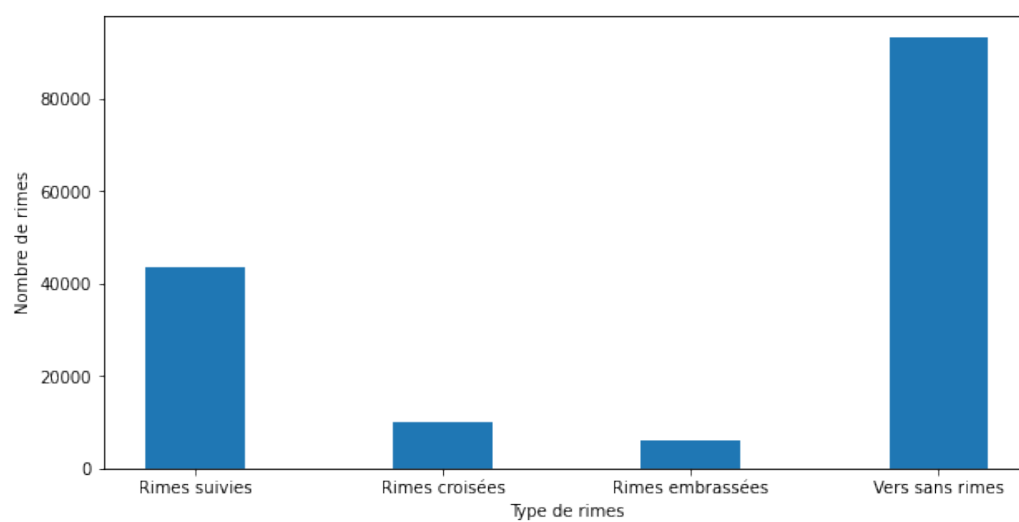


FIGURE 3 – Types de rimes pour les bigrammes

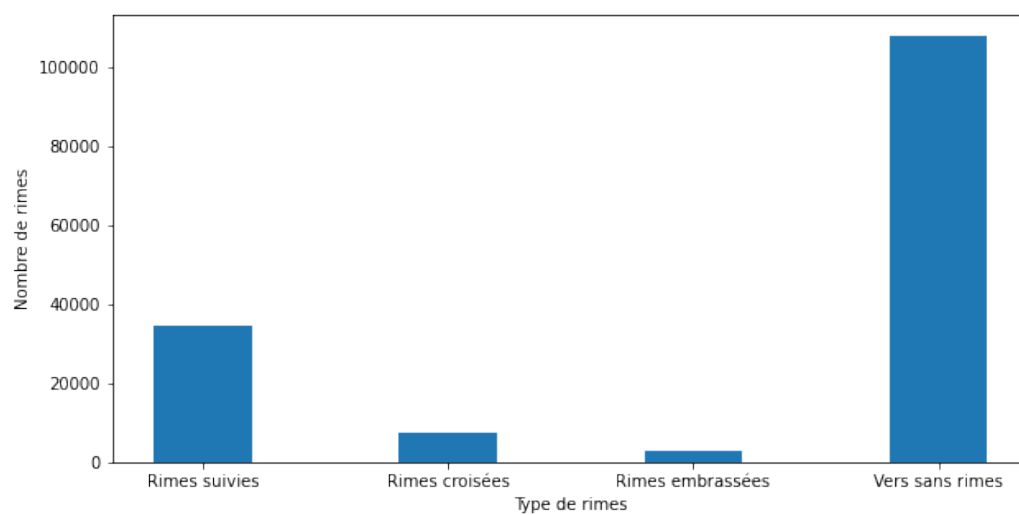


FIGURE 4 – Types de rimes pour les trigrammes

## 5 Conclusion

La poésie présente de nombreux problèmes de traitement automatique. Le découpage syllabique des vers nécessite l'utilisation d'un grand lexique de mots du français déjà annotés et découpés syllabiquement. Dans le cas contraire, il nécessite l'implémentation de nombreuses règles du français en prenant en compte les exceptions pour faire un découpage automatique.

À cela s'ajoute le problème lié au français du XVII<sup>ème</sup> siècle, très différent du français moderne de par son vocabulaire, son orthographe et sa prononciation. Implémenter des règles du français moderne n'est donc pas suffisant, il a fallu adapter le code avec de nouvelles règles, augmentant ainsi la tâche de travail et la complexité du programme.

Néanmoins, grâce aux méta-données mises à disposition par l'auteur du corpus annoté des Mazarinades il a été possible de récupérer facilement les informations telles que la date ou le type de texte (vers ou prose) pour obtenir des statistiques en amont.

Cette abondance de méta-données au format .xml a cependant posé des soucis de parsing avec Python. Il a été difficile d'analyser la structure des fichiers pour trouver les bonnes balises à récupérer et à retirer, ce qui nous montre que le gain en quantité d'informations méta-textuelles se fait au détriment de la lisibilité des fichiers leur facilité de traitement.

### 5.1 Perspectives

Malgré nos améliorations, notre découpage syllabique demeure imparfait. Il faudrait alors analyser de plus près les vers qui posent problème et, selon le type de problème posé, trouver un moyen d'éliminer le bruit ou ajouter de nouvelles règles spécifiques à la métrique du français de l'époque pour obtenir le bon décompte.

Il serait aussi intéressant de comparer les syllabes de fin de vers pour détecter les rimes. En effet, notre fonction d'analyse de rimes ne fait que comparer les n-caractères de fin de vers, or il serait plus judicieux de travailler sur les syllabes.

Cette problématique pourrait se décomposer en deux niveaux : tout d'abord, faire une comparaison purement graphique (une rime est deux syllabes identiques) sans prise en compte des phonèmes, puis une comparaison des syllabes en prenant en compte leur phonétisation (par exemple, "or" et "hors" riment



malgré le fait qu'ils ne soient pas graphiquement identiques).

Cette étape nécessiterait de pouvoir conserver, lors du découpage syllabique, chacune des syllabes. Malheureusement, le programme french-syllabation ne fait que le calcul du nombre de syllabes, sans garder lesdites syllabes découpées.

Nous pourrions également procéder à d'autres types d'analyses, à visée plus sémantiques et stylistiques, en nous intéressant aux figures de style, aux temps et valeurs des verbes, aux champs lexicaux et au positionnement politique par exemple.

## Références

- [Abiven Karine, 2022] Abiven Karine, Lejeune Gael, B. A. (2022). Projet antonomaz. In <https://antonomaz.humanum.fr/exist/apps/Antonomaz/about/contact.html>.
- [Ian, 2021] Ian, N. (2021). french-syllabification. In <https://github.com/ian-nai/french-syllabification>.
- [Wikipédia, 2022] Wikipédia (2022). Article mazarinade. In <https://fr.wikipedia.org/wiki/Mazarinade>.

## A Annexe

Date	Vers	Prose
Sans Date	24 34%	46 66%
1634	1 100%	0 0%
1648	4 11%	34 89%
1649	432 30%	1030 70%
1650	78 27%	213 73%
1651	65 19%	279 81%
1652	120 12%	845 88%
1653	2 22%	7 78%
1654	1 9%	10 91%
1655	0 0%	2 100%
1656	0 0%	2 100%
1662	0 0%	1 100%
1663	1 100%	0 0%
<b>Total</b>	<b>728</b> <b>22,8%</b>	<b>2469</b> <b>77,2%</b>

TABLE 6 – *Proportions de poèmes en vers et de poèmes en prose par année*

Nombre de syllabes	Nombre de vers	Pourcentage
7	42272	27.46
8	30088	19.55
6	28825	18.73
9	12970	8.43
5	10788	7.01
10	8055	5.23
11	6351	4.13
12	3837	2.49
4	3488	2.27
13	1864	1.21
3	1552	1.01
2	1419	0.92
14	1042	0.68
15	677	0.44
16	340	0.22
17	148	0.1
1	77	0.05
18	67	0.04
19	19	0.01
20	15	0.01
21	14	0.01
22	5	0.001
23	3	0.001
27	2	0.001
0	2	0.001
31	1	0.001

TABLE 7 – *Nombre de vers par nombre de syllabes dans tout le corpus - V1*

Nombre de syllabes	Nombre de vers	Pourcentage
7	42609	27.91
8	34274	22.45
6	23604	15.46
9	13360	8.75
10	8051	5.27
5	7172	4.7
11	7077	4.64
12	4576	3.0
4	2178	1.43
13	2158	1.41
2	1400	0.92
3	1257	0.82
14	1218	0.8
15	915	0.6
16	553	0.36
1	551	0.36
17	344	0.23
18	173	0.11
20	118	0.08
19	117	0.08
0	106	0.07
21	97	0.06
22	94	0.06
23	92	0.06
28	87	0.06
27	83	0.05
24	80	0.05
25	76	0.05
26	66	0.04
29	63	0.04
30	51	0.03
32	28	0.02
31	21	0.01
33	14	0.01
35	5	0.001
34	5	0.001
36	2	0.001
46	1	0.001
42	1	0.001
37	1	0.001

20

TABLE 8 – *Nombre de vers par nombre de syllabes dans tout le corpus - V2*

Nb syll	Vers
0	iours
1	flammes
2	–rre11
3	atteinte
4	apotelesma.
5	eiis coront o vs.
6	offisque la splendeur de celuy des csars
7	ieu cheueux cendrez dont mes maius vagbondes
8	bois-robert abbeé de chastillon.
9	qu’apres’ tant de desordre apres tant de licence
10	tu visses aà tes pieds pleurer tes ennemis
11	et quiconque peut vaincre un tel vainqueur que ftoy .
12	peut atteindre au sommet de la plus haute gloire.
13	nos maux avec les tiens enfin ont pris leur cours
14	voicy qui vaut norlingue &l le haure à la fin
15	ie te donne un conseil que ie ne sçaurois suiure
16	lasses prndte onceut qui n’apparuiient qu’à l’mpieoyale ort dont latl e eoe lu.
17	lee oueles ienore onr seie noniner pnr on sl eonnle ceroeas ei le inn i
18	e qus sd’un ccprea senl e couon li parque. l ne e perdry pas en ceitae dlermer
19	te les ls. l’albitre & le couail. nous fappa d’vu seul coup & le cœur & l’oreille
20	troueois dau la eionus u pui bele ve fuit e sie dunmen judeareur p rerr
21	ese oui oe n eon ps eends sin l ens il sau soun d oeurer onr e anue n onina
22	lens nee enine e uni lene de ie eaouns pdenolunie ni leai le nuus laur ne ei ei.
23	leue uou due oa anure d u soir lionail pe i pionpe nes le ee que oianu ile

TABLE 9 – *Exemples de vers par nombre de syllabes - V2 (partie 1)*

Nb syll	Vers
24	sans voir que tu cheris ma memoire & ma cen- tu ne ioignes en fin le soaipif de sa t
25	pue n ne iror auieuude ns sieles nerniruble lueni l senilemeone & panse que ils le
26	e mauuais n'habitera point aopés de moy ny les injustes ne deomeureront pas deuant mes yeux.
27	chans comme des lions tous glorieux du succez des premieres attaques diminuèrent en cette ci beaucoup de la
28	diligentioribus studiosis scrutatoribus consulendus est sefer rabbi abrahbam aben era de lectura
29	duc de longueville avoit envoyé dans la place pour y commander n'eut pas le temps de se retirer apres avoir
30	de la main droite avoit mesme nombre & disposition d'officiers & cavaliers du regiment de bambal où estoit le sieur
31	fieur de la ferté-imbault ayans avec eux les sieurs de heudicourt & de folleville mareschaux de camp les sieurs
32	ce prince qu'il trouva en chemin pour l'asseurer de l'ouverture de leur ville & de leur entiere fidelité au service du
33	aucune assurance : ce qui obligea ce comte à leur commander de le suiure pour voir ce qu'il alloit faire & ayant
34	ne sont par leurs genorauerse : on ne parla plus que le boire di u veuille que la profetie et de bi n braner la machoire :
35	les plaisirs d'un repos asseuré afin que nous soyons perpetuellement unis d'un si fet me lien de paix & d'affection que les mesdi
36	ment unis d'un si fet me lien de paix & d'affection que les mesdisans & les ennemis de ma gloire ne puissent iamais alterer la feais
37	esperer desrmais qu'une eternelle prosperité & que ie demeureray belle magnifique florissante agreable à mes amis & redoura
42	le eai sit ehanuer le te deum pour la pif -de edaratusaiuer s'ar dr qe e furenr onueu au pa tortose. oi i miand le pariement sel on la const ume* emet.
46	our ce qu'l me fa choit d'estre loin de tes yeux. et dessus tous les cears remporter la uctoutemais quand la renommée à mon dā crop fidelle conlole par l'elpor ta peine & mo soucy

TABLE 10 – *Exemples de vers par nombre de syllabes - V2 (partie 2)*

<b>Rime</b>	<b>Nb d'occu.</b>
es	3387
re	2966
ce	1914
le	1596
nt	1475
te	1444
ne	1438
e	1303
ns	1256
is	1206
ie	1153
er	1083
ux	976
Se	973
on	973
rs	968
ur	897
l	848
ge	836
in	797

TABLE 11 – Rimes en bigrammes les plus fréquentes



<b>Rime</b>	<b>Nb d'occu.</b>
nce	1255
ent	1001
l	848
ire	808
lle	753
age	650
eux	633
les	598
ine	550
	537
eur	505
ere	460
que	455
urs	422
ure	400
res	384
SSe	358
Ste	351
tre	347
ois	342

TABLE 12 – Rimes en trigrammes les plus fréquentes