



## RAPPORT DE PROJET DE FIN DE FORMATION

DST – MAR23



*Figure 1: Arbre blanc du Gondor version "data" réalisé par DALL-E*

## **L'UNIVERS DE J.R.R TOLKIEN A LA LOUPE DU NLP : UNE ANALYSE DATA DE SES ŒUVRES**

---

BENARD Jocelyn  
POINSEAUX Hadrien

Mentor : GRIMM Gaspard  
Cheffe de cohorte : LENET Sarah

Version 2.0 du 01/12/2023

---

1.	INTRODUCTION	3
1.1.	CONTEXTE DU PROJET	3
1.2.	PROBLEMATIQUE	3
2.	PRESENTATION DES DONNEES	4
2.1.	DONNEES BRUT	4
2.2.	PREPROCESSING DES DONNEES	6
3.	VECTORISATION DES ŒUVRES ET CLASSIFICATION	7
4.	ANALYSE DE TEXTE : FOCUS SUR LES ŒUVRES DU MONDE DE LA TERRE DU MILIEU	15
4.3.	ANALYSE GENERALE	15
4.4.	ANALYSE DE SENTIMENTS	19
4.5.	ANALYSE DES COMMUNAUTES	22
5.	CREATION D'UN CHATBOT	25
6.	CONCLUSION	26

## 1. INTRODUCTION

### 1.1. CONTEXTE DU PROJET

L'origine de notre projet s'ancre dans une fascination commune pour les œuvres fantastiques de J.R.R. Tolkien, un auteur dont l'univers complexe et riche a inspiré des générations. Une idée germa dans l'esprit de Jocelyn entreprendre une aventure audacieuse au sein de cet univers, mais à travers une lentille innovante - celle de la data science. Cette proposition ne tarda pas à captiver Hadrien, dont la passion pour le monde de Tolkien le conduisit naturellement à rejoindre cette quête.

Ce projet ne se contente pas de s'inscrire dans notre amour pour la littérature de Tolkien ; il est aussi le reflet de notre profond intérêt pour les développements récents en matière de traitement naturel du langage (NLP). Nous percevons dans l'union de ces deux domaines une opportunité unique de repousser les frontières de la technologie et de l'analyse littéraire. Notre objectif est d'explorer comment la data science peut enrichir notre compréhension et notre appréciation des textes de Tolkien, en générant des insights inédits et en créant de nouvelles façons d'interagir avec son œuvre.

L'envergure du projet est sans doute ambitieuse. Il implique de naviguer à travers divers modèles de NLP, chacun offrant des perspectives et des capacités distinctes. Cela nécessite une exploration méthodique des dernières avancées dans le domaine, ainsi qu'une adaptation et une interprétation créative de ces technologies.

Notre voyage, semblable à celui de la Fraternité de l'Anneau, est complexe et semé d'obstacles.

Bonne Lecture.

### 1.2. PROBLEMATIQUE

Comment l'application des technologies de traitement naturel du langage (NLP) peut-elle enrichir notre compréhension et notre interaction avec l'univers littéraire complexe de J.R.R. Tolkien ? Quelles en sont les limites

## 2. PRESENTATION DES DONNEES

### 2.1. DONNEES BRUT

L'ensemble des textes suivants sont à notre disposition au format .txt et dans leur langue originale (anglais) :

- The Hobbit
  - Genre : Fantasy
  - Publication : 1937
  - Univers : Terre du Milieu
  - Public : Jeunesse, tous âges
  - Résumé : Aventure de Bilbo Baggins qui trouve l'Anneau Unique.
- Farmer Giles of Ham
  - Genre : Fantasy
  - Publication : 1949
  - Univers : Indépendant
  - Public : Jeunesse, tous âges
  - Résumé : Histoire humoristique d'un fermier qui devient un héros.
- The Fellowship of the Ring (Le Seigneur des Anneaux, première partie)
  - Genre : Fantasy épique
  - Publication : 1954
  - Univers : Terre du Milieu
  - Public : Adultes, jeunes adultes
  - Résumé : Début de la quête pour détruire l'Anneau Unique.
- The Two Towers (Le Seigneur des Anneaux, deuxième partie)
  - Genre : Fantasy épique
  - Publication : 1954
  - Univers : Terre du Milieu
  - Public : Adultes, jeunes adultes
  - Résumé : Suite de la quête et diversification des histoires des personnages principaux.
- The Return of the King (Le Seigneur des Anneaux, troisième partie)
  - Genre : Fantasy épique
  - Publication : 1955
  - Univers : Terre du Milieu
  - Public : Adultes, jeunes adultes
  - Résumé : Conclusion de la quête et des conflits majeurs.

- The Adventures of Tom Bombadil
  - Genre : Poésie, fantasy
  - Publication : 1962
  - Univers : Terre du Milieu
  - Public : Tous âges
  - Résumé : Collection de poèmes sur Tom Bombadil et d'autres thèmes de la Terre du Milieu.
- Smith of Wootton Major
  - Genre : Fantasy
  - Publication : 1967
  - Univers : Indépendant
  - Public : Jeunesse, tous âges
  - Résumé : Histoire d'un village et d'un monde de fées.
- The Silmarillion
  - Genre : Fantasy épique, mythopoeia
  - Publication : 1977 (posthume)
  - Univers : Terre du Milieu
  - Public : Adultes, jeunes adultes
  - Résumé : Histoire de la création et des premiers âges de la Terre du Milieu.
- Unfinished Tales of Númenor and Middle-earth
  - Genre : Fantasy épique
  - Publication : 1980 (posthume)
  - Univers : Terre du Milieu
  - Public : Adultes, jeunes adultes
  - Résumé : Collection d'histoires inachevées se déroulant dans la Terre du Milieu.
- Mr. Bliss
  - Genre : Littérature enfantine
  - Publication : 1982 (posthume)
  - Univers : Indépendant
  - Public : Enfants
  - Résumé : Histoire illustrée d'un homme et de ses aventures avec des animaux et un véhicule magique.

- The Letters of J.R.R. Tolkien
  - Genre : Correspondance
  - Publication : 1981 (posthume)
  - Univers : Principalement Terre du Milieu
  - Public : Adultes, chercheurs
  - Résumé : Collection de lettres offrant un aperçu de la vie et des pensées de Tolkien.
- The Children of Húrin
  - Genre : Fantasy épique
  - Publication : 2007 (posthume)
  - Univers : Terre du Milieu
  - Public : Adultes, jeunes adultes
  - Résumé : Tragédie de Túrin Turambar, se déroulant dans le Premier Âge.
- The Legend of Sigurd and Gudrún
  - Genre : Poésie épique
  - Publication : 2009 (posthume)
  - Univers : Mythologie nordique
  - Public : Adultes, amateurs de mythologie
  - Résumé : Poèmes épiques basés sur la légende de Sigurd et Gudrún.
- The Fall of Gondolin
  - Genre : Fantasy épique
  - Publication : 2018 (posthume)
  - Univers : Terre du Milieu
  - Public : Adultes, jeunes adultes
  - Résumé : Histoire de la chute de la ville elfique de Gondolin.

## 2.2. PREPROCESSING DES DONNEES

Dans le cadre de notre projet, un processus rigoureux de prétraitement des données a été mis en œuvre pour garantir la qualité et la pertinence de l'information traitée. Ce processus est crucial pour l'analyse de texte et le traitement automatique du langage naturel. Voici les étapes clés que nous avons suivies :

- 1) Extraction des noms des livres
- 2) Construction du corpus
- 3) Division des textes en mots
- 4) Mise en minuscule
- 5) Suppression de la ponctuation

- 6) Filtrage des stop words : Les 'stop words', ou mots vides, sont des mots couramment utilisés dans une langue mais généralement sans signification spécifique pour l'analyse du texte, comme « and », « the », « in »...
- 7) Lemmatisation : La lemmatisation est un processus de réduction des mots à leur forme de base ou lemmes.

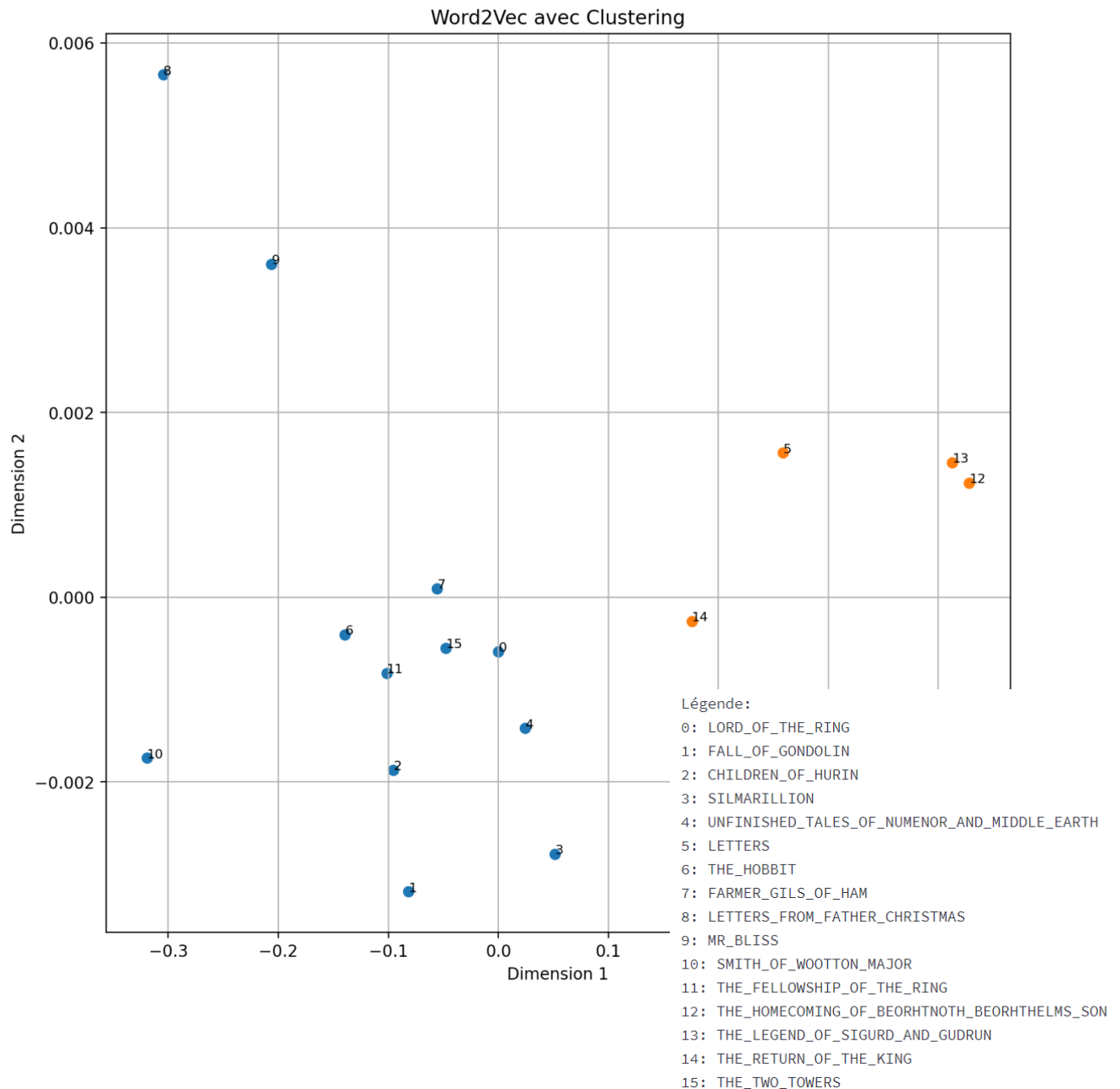
### 3. VECTORISATION DES ŒUVRES ET CLASSIFICATION

Dans cette section, nous avons mis en œuvre une vectorisation des œuvres de J.R.R. Tolkien, suivie d'une classification. L'objectif était de découvrir des modèles et des relations profondes au sein de son univers littéraire via la vectorisation.

Le modèle Word2Vec a été entraîné sur ces corpus prétraités, permettant la création de vecteurs représentatifs pour chaque œuvre. En parallèle, nous avons utilisé le modèle SentenceTransformer, basé sur BERT, pour encoder les textes dans un espace vectoriel.

Ces vecteurs ont ensuite été réduits à deux dimensions via PCA pour une visualisation intuitive.

L'application de la méthode de clustering KMeans, guidée par des scores de silhouette pour déterminer le nombre optimal de clusters, a révélé des groupes distincts au sein de ces vecteurs réduits.



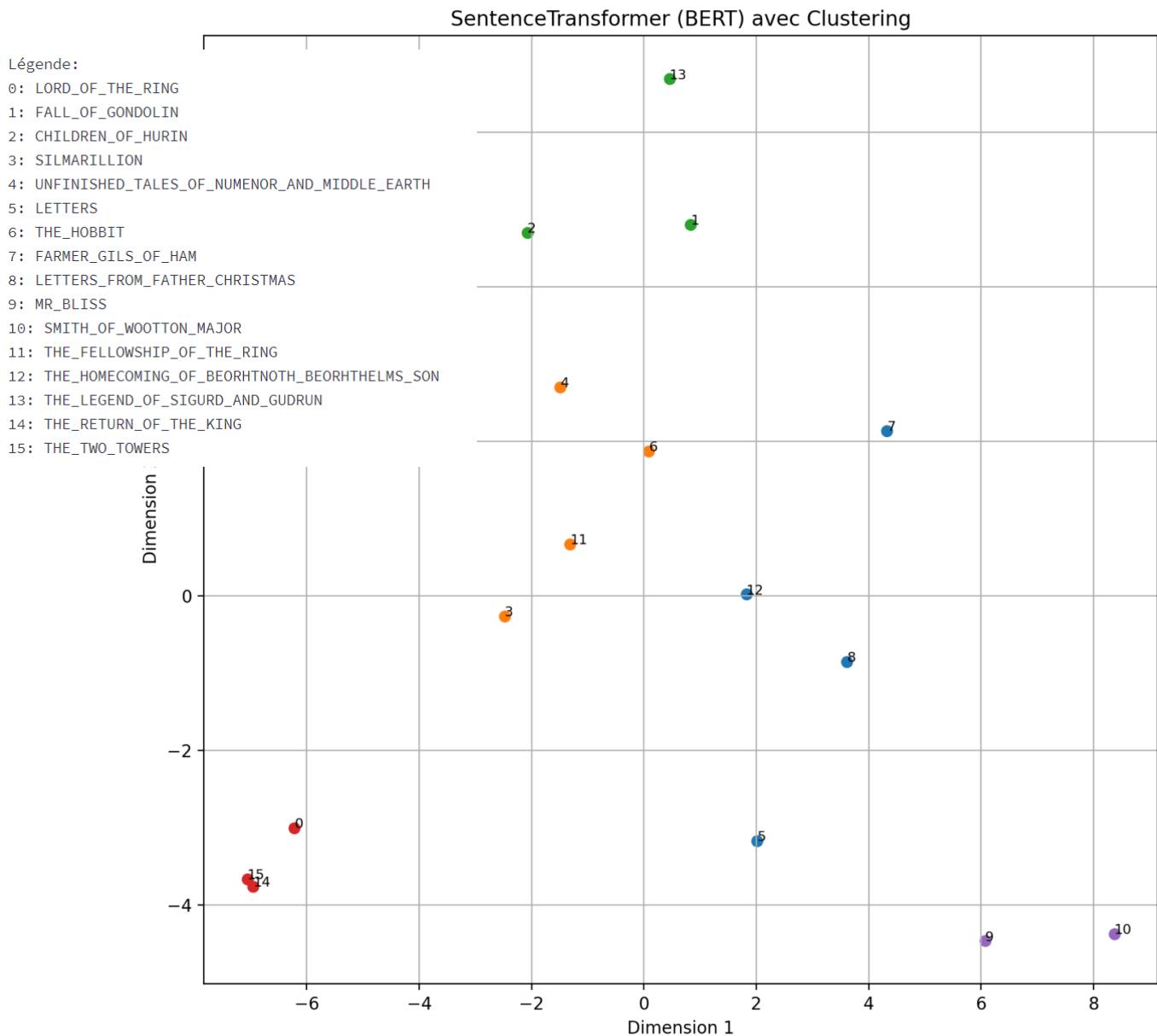
Les clusters générés par le modèle Word2Vec semblent avoir capturé des similitudes syntaxiques entre les textes. Ces similitudes pourraient inclure l'usage de la langue, la récurrence de noms de personnages et de lieux, ce qui a vraisemblablement contribué au regroupement des œuvres. Par exemple, l'utilisation répétée de structures de phrases ou de vocabulaire spécifique à l'univers de Tolkien aurait pu être un facteur de cette agrégation.

En se basant sur l'analyse des clusters, le modèle a distingué les œuvres en deux groupes principaux. Le groupe bleu semble contenir des œuvres caractérisées par un ton plus enfantin et plus accessible, qui serait plus approprié pour un jeune public. Cela pourrait refléter une langue plus simple, des thèmes plus légers et un style narratif engageant pour les enfants. D'autre part, le groupe orange regroupe des œuvres avec des tonalités plus



sombres ou plus historiques, qui pourraient s'adresser à un public plus mûr. Cela suggère une complexité thématique plus grande, une utilisation de la langue plus recherchée et la présence de motifs plus graves ou sérieux qui résonnent avec un lectorat adulte.

Cependant, il est important de noter que le modèle Word2Vec peut passer à côté de nuances de sens car il se base sur la moyenne des vecteurs de mots, ce qui peut noyer les subtilités dans une représentation générale. De plus, le modèle peut être sensible au bruit provenant de mots moins fréquents, qui peuvent altérer la précision des vecteurs moyens et, par conséquent, influencer la formation des clusters.



Les clusters obtenus par le modèle SentenceTransformer, qui repose sur BERT, semblent avoir capturé des similitudes sémantiques entre les textes de Tolkien. La proximité thématique, telle que des intrigues partagées ou des références mythologiques communes, a probablement joué un rôle clé dans le regroupement des œuvres. Par exemple, des histoires entrelacées ou des éléments mythologiques similaires présents dans différents livres pourraient expliquer pourquoi certains textes sont regroupés ensemble.

Les vecteurs générés par le SentenceTransformer reflètent de manière assez fidèle les livres de Tolkien. Par exemple, la proximité observée entre "La Communauté de l'Anneau" et "Le Hobbit" est cohérente, car "La Communauté de l'Anneau" se réfère fréquemment à des événements survenus dans "Le Hobbit". Deux groupes principaux associés au monde de la Terre du Milieu sont identifiés : orange et rouge. Trois autres groupes, vert, bleu et violet, semblent correspondre à des œuvres plus diverses et en dehors de la Terre du Milieu, avec le groupe violet incluant des nouvelles plus courtes, tandis que les autres groupes contiennent des ouvrages plus substantiels.

On note aussi une grande similarité entre "Le Retour du Roi" et "Les Deux Tours", ce qui est logique puisqu'à l'origine ces textes faisaient partie d'un seul manuscrit. "La Communauté de l'Anneau", quant à elle, a été retravaillée pour faire le lien avec "Le Hobbit". Cela est en phase avec l'histoire éditoriale de l'œuvre de Tolkien, où "Le Hobbit", bien qu'écrit en premier, a été suivi par "Le Seigneur des Anneaux" qui a élargi et approfondi l'univers et les histoires introduites. La classification des textes reflète donc non seulement le contenu narratif et thématique, mais aussi l'histoire de la création et de la publication des œuvres de Tolkien.

Cependant, en captant le sens sémantique, le SentenceTransformer peut parfois trop généraliser, omettant de distinguer les différences subtiles entre des textes qui partagent des thèmes similaires. Cela pourrait conduire à regrouper des textes qui, bien qu'ils traitent de sujets semblables, présentent des nuances et des détails distincts qui ne sont pas capturés par le modèle.

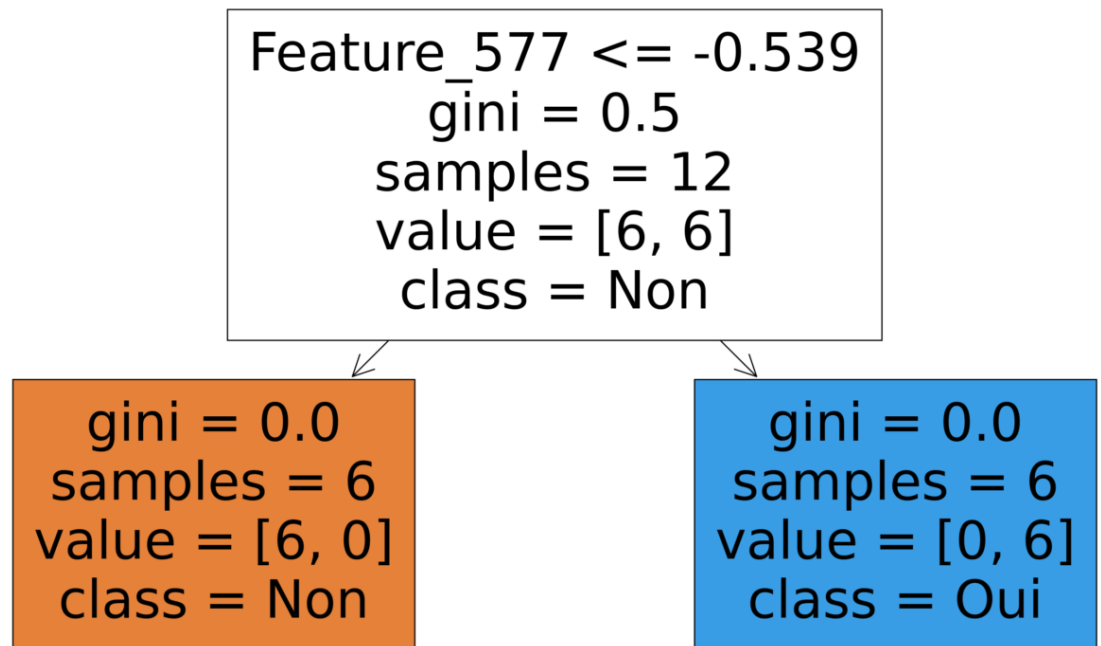
Enfin il faut garder en mémoire 2 problématiques :

- **Les distorsions dues à la PCA** : La méthode de l'ACP (Analyse en Composantes Principales), bien qu'utile pour réduire la dimensionnalité et visualiser les données, peut conduire à une perte d'informations. La représentation en 2D peut ne pas saisir entièrement la complexité de l'espace multidimensionnel original. Ainsi, certaines subtilités et liens entre les textes peuvent être masqués ou altérés dans le processus de réduction dimensionnelle.
- **La subjectivité dans le clustering** : Le choix du nombre de clusters repose sur le score de silhouette, qui ne correspond pas toujours à des regroupements intuitifs ou significatifs d'un point de vue littéraire. En effet, cette mesure est purement statistique et ne tient pas compte des nuances littéraires ou thématiques qui pourraient être évidentes pour un lecteur humain. Par conséquent, bien que les groupes proposés par le modèle offrent une certaine compréhension de la structure des données, ils nécessitent une interprétation prudente et un examen critique pour s'assurer qu'ils reflètent fidèlement les liens littéraires entre les œuvres de Tolkien.

En complément, nous avons chargé des données de labels associées à chaque œuvre pour une classification plus poussée. Ces labels ont été encodés et utilisés pour entraîner des modèles d'arbres de décision, permettant de prédire divers attributs des livres (Genre fantasy ou non, Publication à titre posthume ou non, Œuvre issue du monde la Terre du Milieu ou Non).

Arbre de décision pour publication posthume :

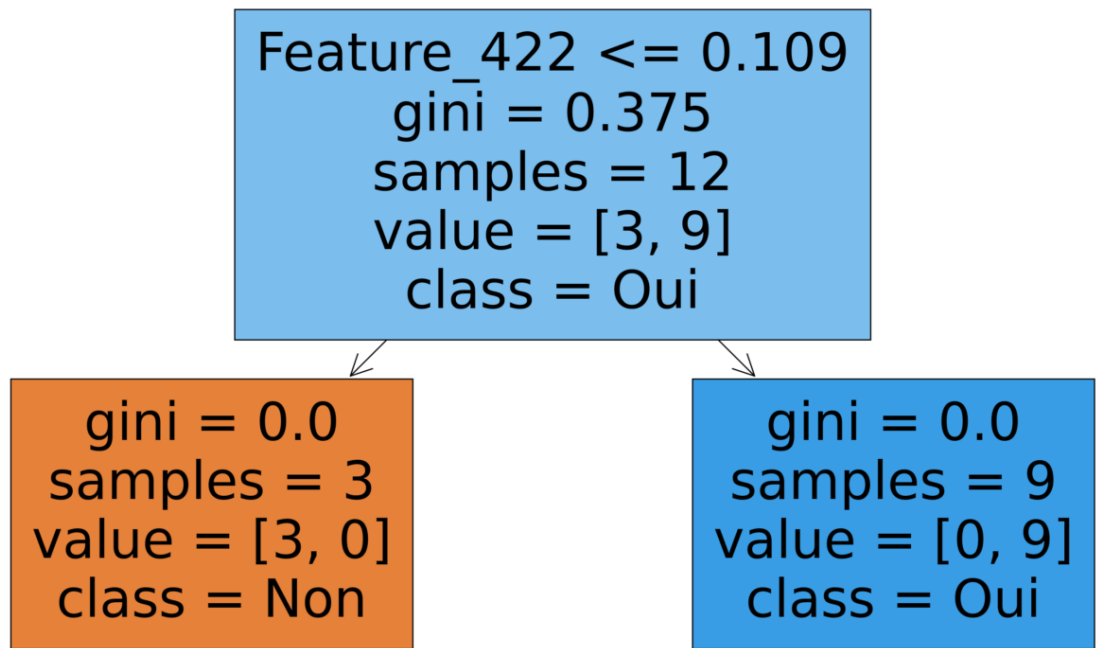
Score de précision : 0.75



Avec un score de précision de 0.75, l'arbre semble relativement bon pour déterminer si une œuvre a été publiée posthume. La séparation basée sur la caractéristique 577 montre une répartition équilibrée des échantillons (6 "Non", 6 "Oui"), mais il faut noter que la mesure de pureté Gini est de 0.5 à la racine, ce qui indique une séparation initiale assez aléatoire.

Arbre de décision pour genre fantasy :

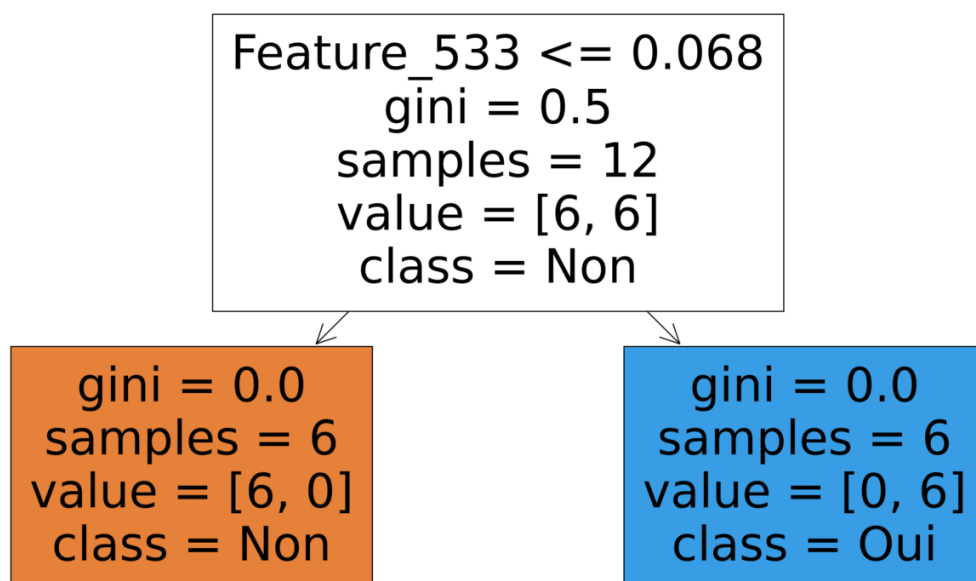
Score de précision : 0.50



Le score de précision plus bas, 0.50, suggère que classer une œuvre comme appartenant au genre fantasy est plus complexe. Cela pourrait indiquer que les frontières entre le genre fantasy et d'autres genres dans l'œuvre de Tolkien ne sont pas aussi claires et que des éléments de fantasy peuvent s'entremêler avec d'autres genres littéraires.

Arbre de décision pour œuvre issue de la Terre du milieu :

Score de précision : 0.75



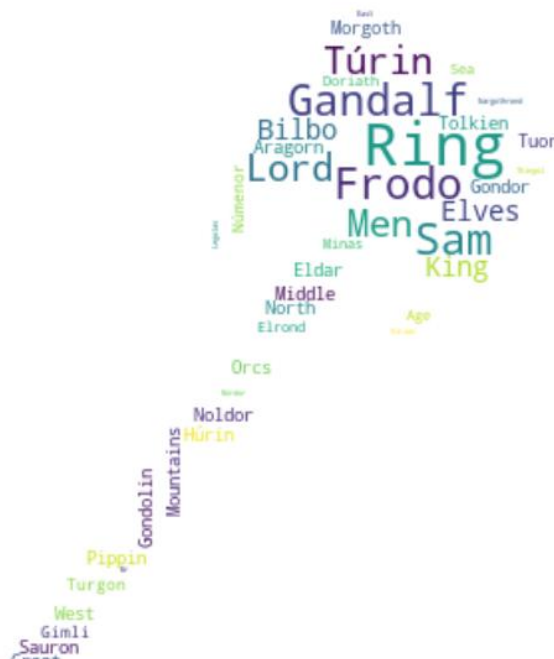
Un autre score de précision de 0.75 montre que les caractéristiques utilisées par l'arbre capturent bien les éléments qui distinguent les œuvres se déroulant en Terre du Milieu. Cependant, tout comme pour la publication posthume, la mesure de Gini à la racine est de 0.5, signifiant que la séparation initiale des classes "Non" et "Oui" est équitable, mais pas nécessairement significative

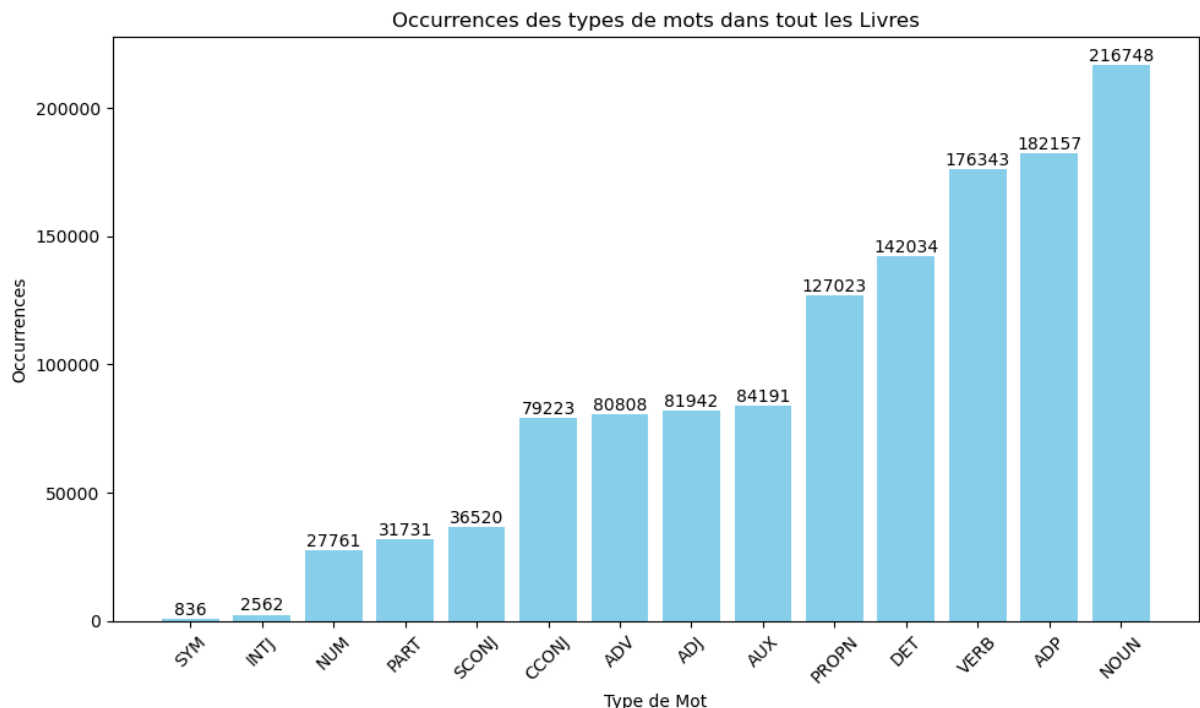
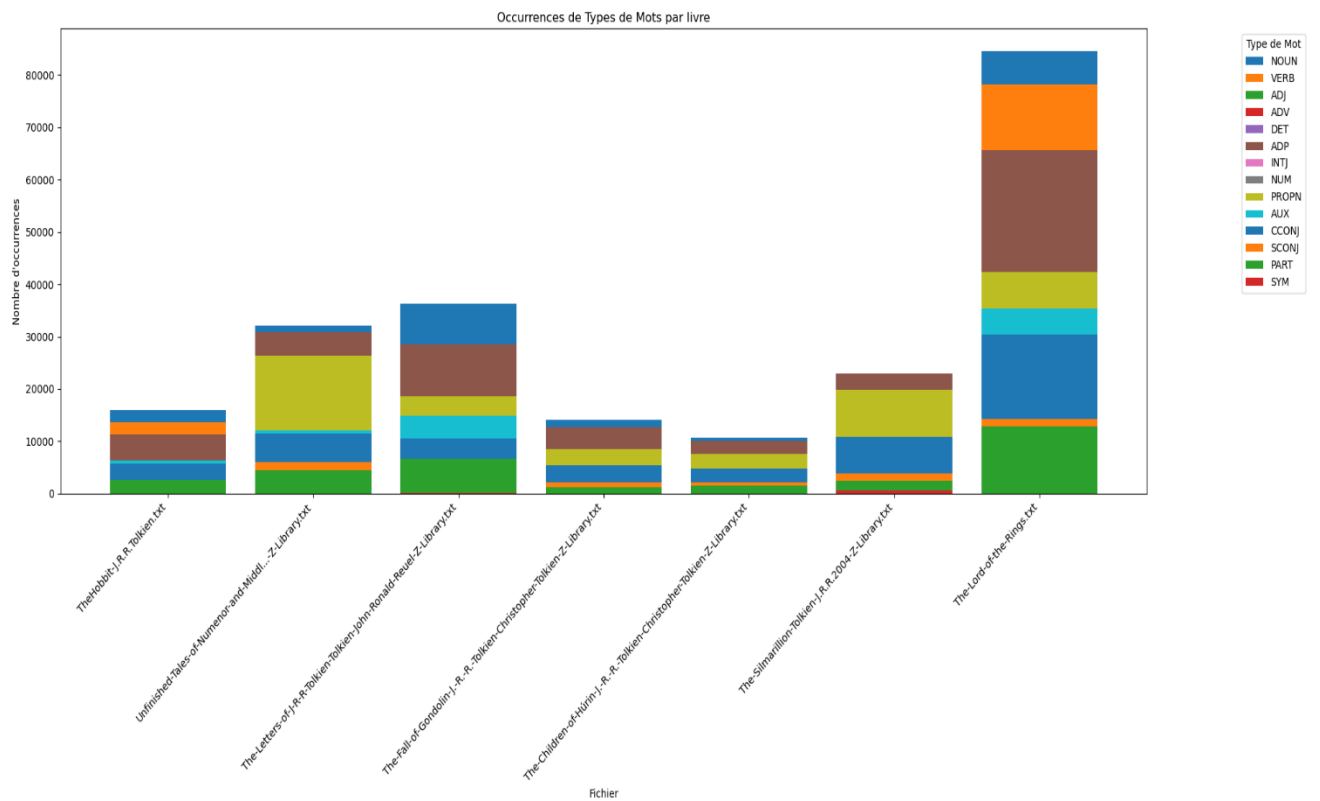
Les résultats laisseraient supposer qu'il est possible de trouver des caractéristiques dans les vecteurs de texte permettant de classer les œuvres selon leur genre, leur appartenance au monde de la Terre du Milieu, ou leur publication à titre posthume. Cependant, la réalité est plus nuancée. Le score de précision pour le genre fantasy est à la limite de l'aléatoire, ce qui peut indiquer que les critères utilisés ne sont pas suffisamment discriminants ou que le genre fantasy chez Tolkien est imbriqué avec d'autres éléments littéraires rendant la classification moins claire.

Pour la classification de la publication posthume et de l'appartenance à la Terre du Milieu, bien que les scores de précision soient identiques, il est important de considérer le déséquilibre potentiel des classes dans les données. Si une classe est prédominante, même un modèle peu performant pourrait atteindre un score de précision élevé simplement en prédisant toujours la classe majoritaire.

De plus, l'arbre de décision, par nature, peut surinterpréter les données d'entraînement (overfitting), capturant des régularités qui ne se généralisent pas bien à de nouvelles données. Cela peut conduire à des performances exagérément optimistes sur l'ensemble d'entraînement mais médiocres sur de nouvelles données non vues.

En résumé, bien que les arbres de décision aient fourni certains indices sur la structure des données, une analyse plus approfondie est nécessaire pour valider ces résultats et comprendre les nuances derrière les critères de classification utilisés. Il serait également bénéfique d'explorer d'autres méthodes de modélisation pour améliorer la précision et la robustesse des prédictions.







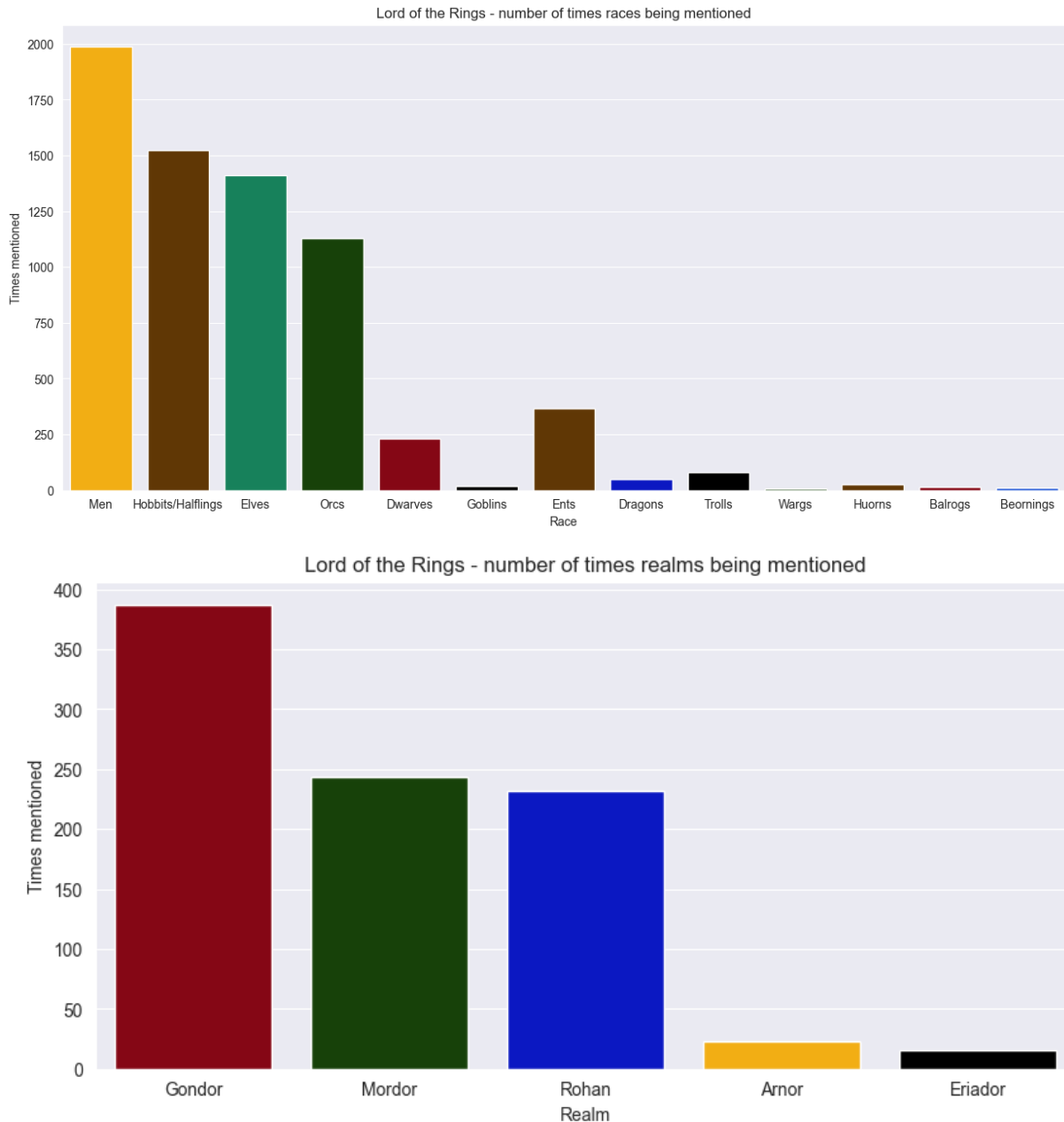
Notre étude a révélé une forte occurrence de noms propres, ce qui nous a incités à explorer une analyse graphique.

Nous avons donc décidé de nous concentrer sur une analyse des sentiments des personnages principalement issus du "Seigneur des Anneaux" et de "The Hobbit". Cette analyse vise à comprendre les fluctuations comportementales des personnages, notamment leurs choix parfois discutables et leur quête de rédemption sur une période donnée.

### Les graphiques issus de l'analyse du Seigneur des Anneaux :

one great like said time came many long road passed hobbit even well know thing away went name last stone frodo dark old men night far must behind year fire ring orc u way lay seemed tree day good made leave saw upon gate hand much may king gandalf stood think aragorn un mountain shadow went back may first say thought yet gender

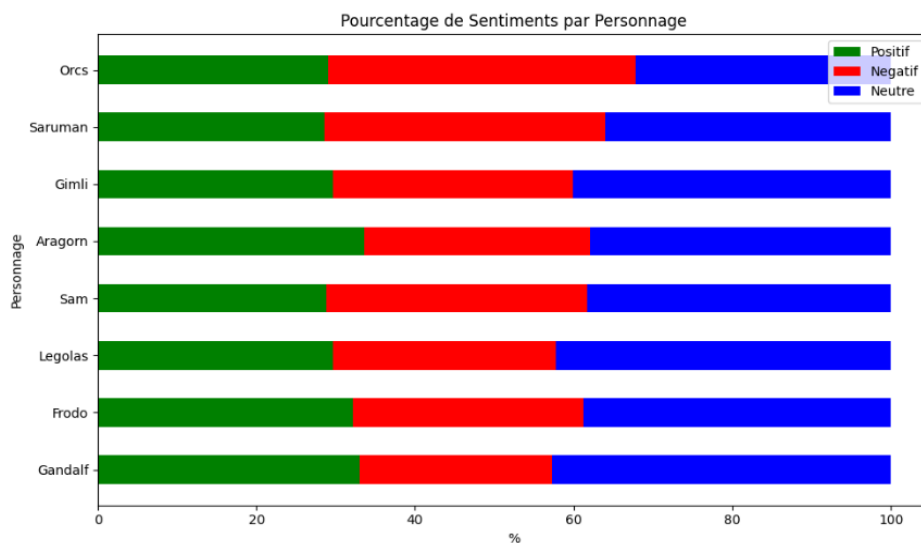
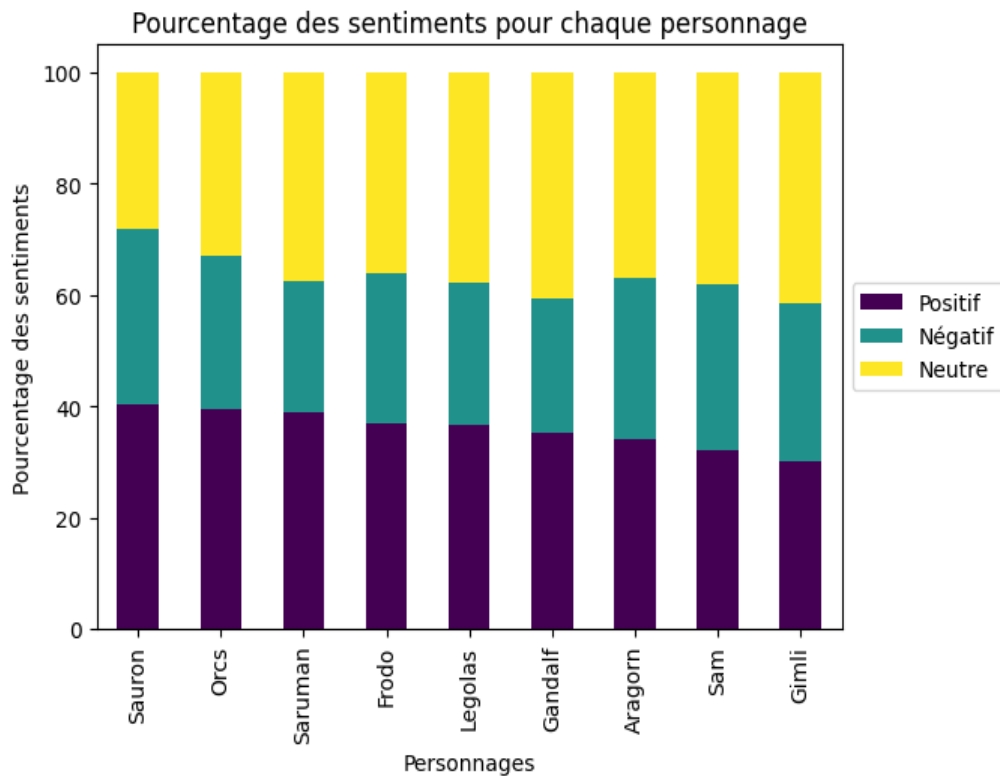




#### 4.4. ANALYSE DE SENTIMENTS

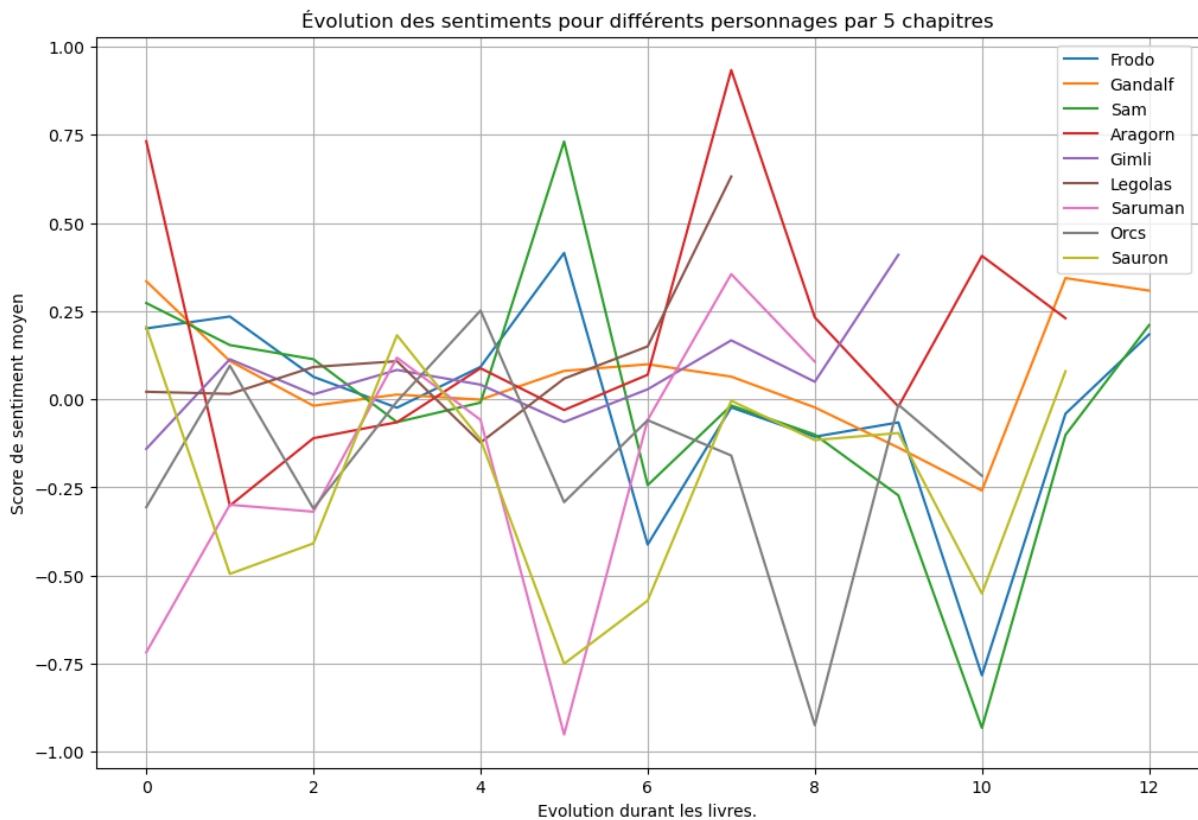
Dans cette section, nous nous focalisons sur le traitement du langage naturel (NLP) pour effectuer une analyse approfondie des sentiments associés aux personnages de "Le Seigneur des Anneaux" (LOTR) et de "The Hobbit". Notre expertise nous permet de comprendre que les personnages de ces œuvres traversent des phases variées, oscillant entre des états émotionnels négatifs et positifs, souvent dans un contexte de rédemption.

Initialement, notre objectif était de déterminer si l'analyse de sentiments est capable de distinguer si un personnage est perçu de manière positive ou négative à travers le texte.



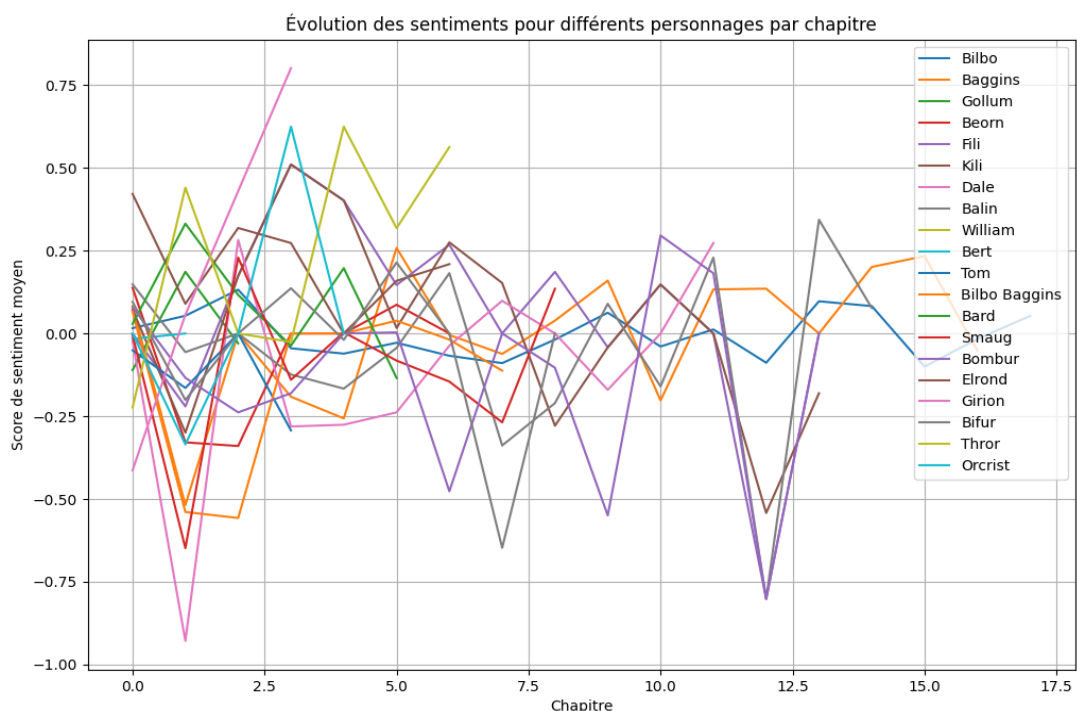
Les résultats préliminaires montrent que, tandis que Spacy tend à classer à tort des personnages intrinsèquement négatifs comme positifs (exemple : Sauron), nltk, en utilisant le dictionnaire 'vader\_lexicon', se révèle plus efficace pour identifier les nuances négatives. Nous observons également que les personnages présentent souvent des sentiments ambivalents, avec une marge de différence maximale d'environ 10% entre les sentiments positifs et négatifs, et une proportion notable de sentiments neutres (environ 40% par personnage). Cette neutralité suggère que les algorithmes éprouvent des difficultés à interpréter les sentiments dans le style d'écriture plus ancien de J.R.R. Tolkien.

Pour une analyse plus poussée, nous choisirons nltk avec 'vader\_lexicon' pour examiner l'évolution des sentiments des personnages à travers les trois tomes de LOTR.



Les résultats indiquent une oscillation entre sentiments positifs et négatifs pour la majorité des personnages, avec souvent une tendance à devenir plus positif vers la fin, illustrant le concept de rédemption mentionné. Les moyennes fluctuent entre -0.50 et 0.25, renforçant l'hypothèse que les sentiments sont généralement neutres et difficiles à catégoriser par l'algorithme. Cette neutralité pourrait être attribuée aux descriptions détaillées typiques de Tolkien, qui pourraient influencer la perception des sentiments dans les textes.

Une analyse similaire appliquée à "The Hobbit" révèle des tendances comparables en termes d'oscillation des sentiments et de rédemption des personnages.



Les sentiments fluctuent également entre -0.50 et 0.25, ce qui confirme des similitudes dans l'évolution des personnages entre les deux œuvres.

Pour compléter notre analyse, il est pertinent de considérer deux concepts centraux dans l'œuvre de Tolkien : la "**Longue Défaite**" issue des mythes anglo-saxon, germanique et chrétien et le procédé stylistique du "**Deus Ex Machina**".

La "Longue Défaite" fait référence à une lutte continue et souvent infructueuse contre le mal, symbolisant la persévérance face à l'inévitable déclin ou échec. Cela se reflète dans l'oscillation des sentiments des personnages, qui traversent des périodes de désespoir ou de négativité, illustrant leur lutte dans un monde marqué par l'imperfection et le conflit.

D'autre part, le "Deus Ex Machina", littéralement "Dieu issu de la machine", est un élément narratif où une solution inattendue ou un pouvoir extérieur intervient pour résoudre des situations apparemment désespérées. Dans le cadre du seigneur des anneaux : Gollum qui entraîna la chute de l'anneau, les aigles qui viennent sauver Sam et Frodon.

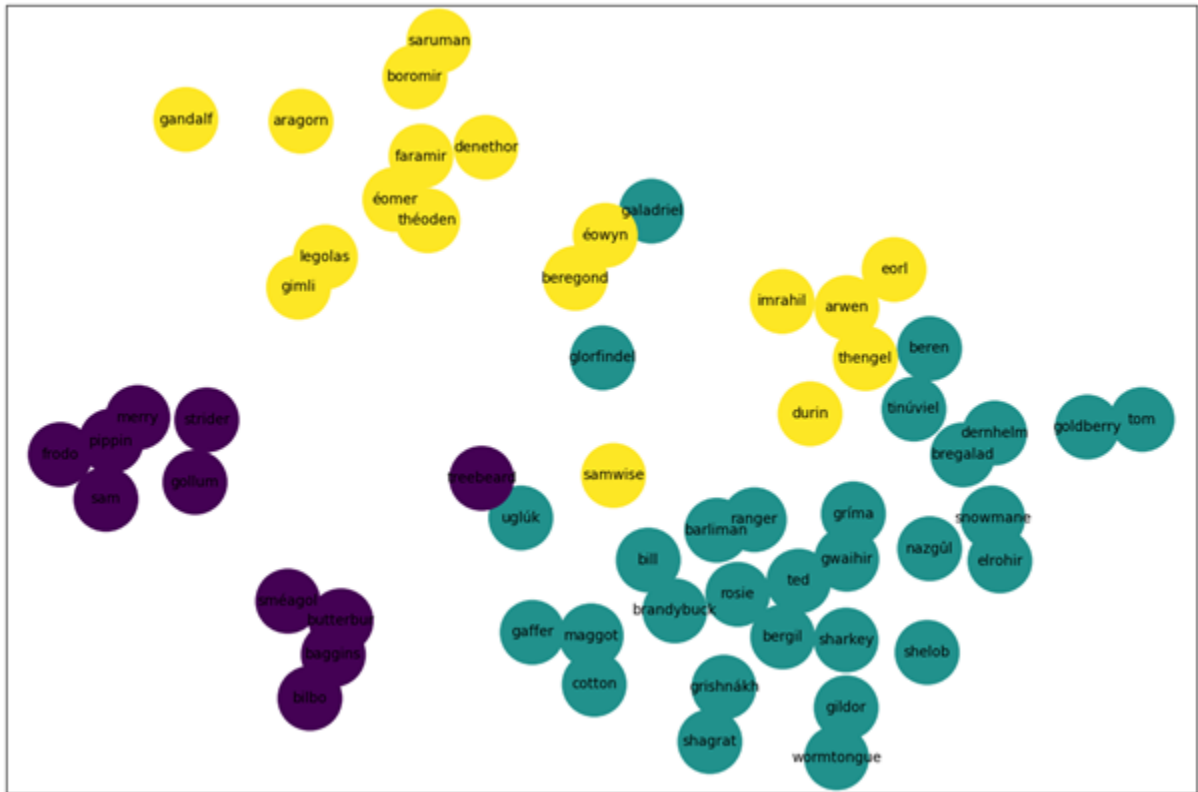
Ces deux notions peuvent être observées dans la tendance des personnages à expérimenter une "libération finale" ou une rédemption, souvent après une période de négativité intense. Ces concepts que l'on peut interpréter dans l'analyse de sentiments, offrent un aperçu de la manière dont Tolkien utilise les fluctuations émotionnelles pour développer ses personnages et faire avancer l'intrigue, reflétant ainsi un mélange complexe de fatalisme et d'espoir dans ses récits.

#### 4.5. ANALYSE DES COMMUNAUTES

Dans cette section, nous avons élaboré divers graphiques pour analyser comment les algorithmes de détection de communauté identifient les groupes en se basant sur une vectorisation des mots via Word2vec et la création de graphes avec NetworkX.

Nous avons d'abord réalisé un graphe pour identifier les communautés au sein de "Le Seigneur des Anneaux" (LOTR), en utilisant deux méthodes d'analyse distinctes : Louvain et Fruchterman-Reingold.

Analyse avec louvain :

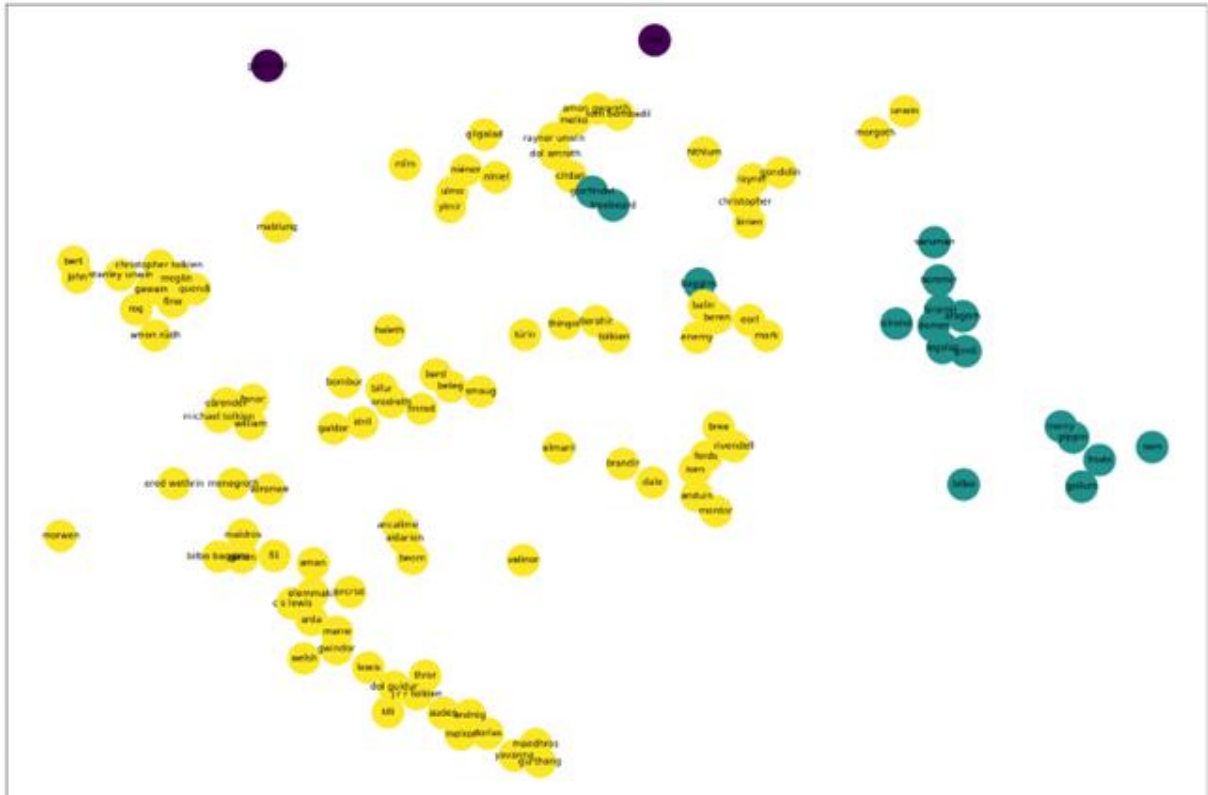


Analyse avec fruchterman-reingold :



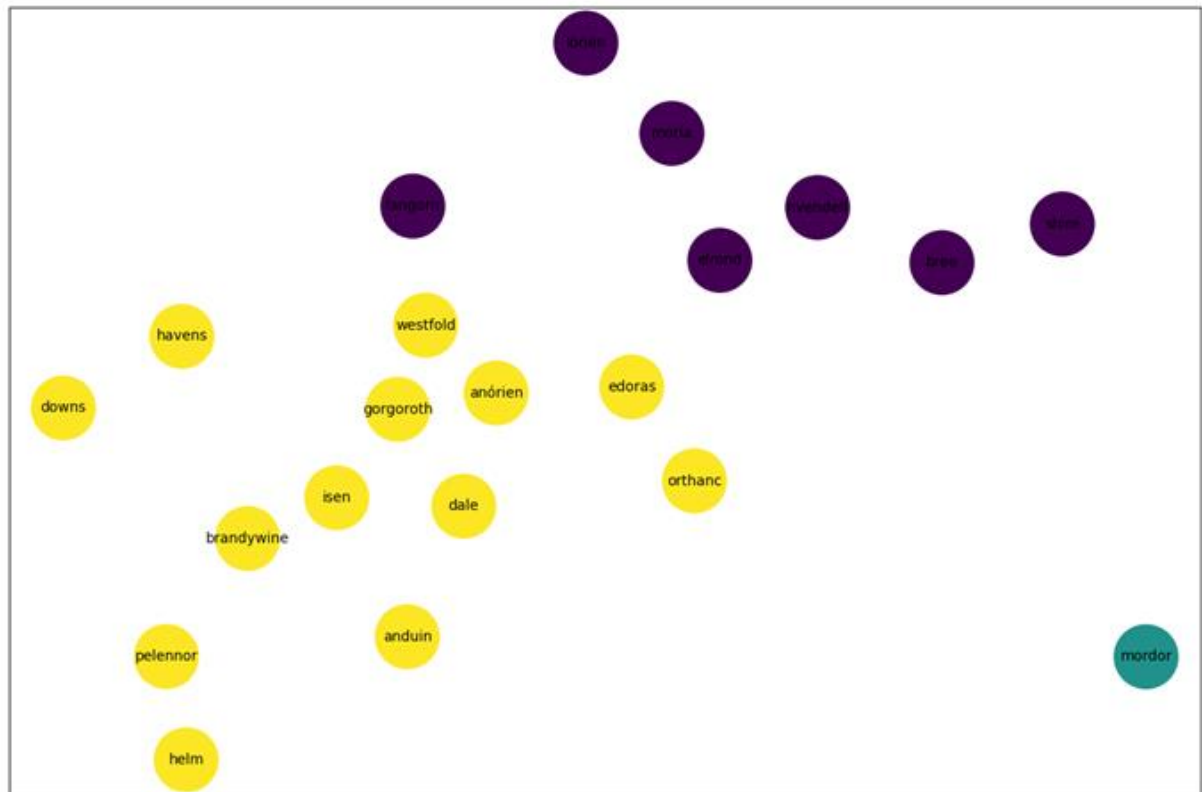
L'analyse avec Louvain a révélé certaines tendances intéressantes. Par exemple, il a été observé que certaines communautés sont formées en fonction de la race (comme les Hobbits) ou des liens récurrents entre les personnages (comme Gimli et Legolas). Trois grandes communautés ont été identifiées, mais cette classification ne reflète pas nécessairement une séparation stricte par race ou par proximité narrative. La détection des liens intercommunautaires demeure un défi.

Ensuite, nous avons étendu notre analyse à tous les personnages des livres en utilisant la méthode Louvain, en nous limitant aux 20 personnages les plus récurrents de chaque livre.



Enfin, nous avons exploré la détection de communautés basée sur les lieux, toujours avec la méthode Louvain.





Cette approche a révélé un schéma de liaison plus marqué entre les lieux en fonction des sentiments associés. Nous avons distingué des lieux neutres (en jaune), des lieux associés au bien (en violet) et des lieux associés au mal, avec le Mordor isolé en bas à droite. Cette classification a mis en évidence des liens plus évidents et significatifs entre les lieux au sein de l'univers de Tolkien.

## 5. CREATION D'UN CHATBOT

Dans le cadre de notre projet, nous avons développé un chatbot, baptisé "The Digital Tolkien", conçu pour répondre avec précision aux questions des lecteurs sur l'univers richement détaillé de J.R.R. Tolkien. Tolkien, connu pour sa passion et son attention méticuleuse aux détails dans ses œuvres, a créé un monde complexe, comprenant des descriptions approfondies des lieux, des personnages, et même des langues uniques. Notre objectif était de capturer cette richesse dans un chatbot interactif.

Pour réaliser cela, nous avons intégré une analyse textuelle approfondie au sein de ChatGPT. Le chatbot tire ses connaissances de plusieurs sources textuelles, notamment "Le Seigneur des Anneaux", "La Chute de Gondolin", "Les Enfants de Húrin", "Le Silmarillion", "Contes et légendes inachevés", et les "Lettres" de Tolkien. Ces textes sont lus, combinés en un seul corpus étendu et ensuite divisés en segments maniables à l'aide d'un séparateur de texte spécifique.

Pour renforcer la base de connaissances du chatbot, nous avons utilisé OpenAIEmbeddings et FAISS, permettant une recherche de similarité sophistiquée dans le corpus. Lorsqu'un utilisateur pose une

question, le système identifie les segments de texte les plus pertinents, qui sont ensuite utilisés pour générer une réponse précise et contextuellement appropriée.

Le chatbot est hébergé sur une interface utilisateur conviviale, créée avec Streamlit. Les utilisateurs sont accueillis par une invite les encourageant à poser leur question à J.R.R. Tolkien. Malgré quelques erreurs mineures inhérentes à tout système basé sur l'IA, le chatbot s'est révélé efficace pour fournir des réponses détaillées et précises, rendant l'exploration de l'univers de Tolkien plus accessible et engageante pour les lecteurs et les fans.

Nous avons également essayé d'intégrer une copie du style de Tolkien à partir de « Letters » mais sans succès.

## 6. CONCLUSION

En réponse à notre problématique initiale, nous avons pu démontrer comment l'application de la technologie NLP peut enrichir notre compréhension de l'univers littéraire de Tolkien. Cependant, il est apparu que, malgré l'efficacité des outils de NLP, la complexité de ces derniers et la complexité de l'écriture de Tolkien présente des défis uniques qui requièrent une approche plus nuancée.

Nos explorations en analyse des sentiments ont révélé une capacité à suivre l'évolution émotionnelle des personnages, capturant leurs phases de rédemption et les oscillations entre des sentiments positifs et négatifs. Néanmoins, cette technologie a parfois eu du mal à classer de manière fiable les personnages selon une polarité positive ou négative, en raison de la prose descriptivement riche et nuancée de Tolkien. L'approche future pourrait inclure l'adaptation de l'outil d'analyse à la lexique spécifique de Tolkien ou l'exclusion des segments textuels excessivement descriptifs pour affiner la précision de l'analyse.

De même, les résultats de la détection de communautés n'ont pas permis de parvenir à une classification claire et exploitable, soulignant la nécessité d'une compréhension plus approfondie du matériel source pour affiner les algorithmes et mieux saisir les intrications de la Terre du Milieu.

À travers ce projet, nous avons appris l'importance vitale de la connaissance du domaine littéraire pour guider l'analyse de données et l'interprétation des résultats obtenus via NLP. Cette expertise est essentielle pour évaluer l'exactitude des outils de NLP et pour ajuster les paramètres de manière à refléter fidèlement les subtilités de l'œuvre étudiée.

Notre travail a souligné l'importance critique de l'expertise humaine dans l'évaluation et l'interprétation des résultats. Les outils de NLP peuvent fournir un aperçu précieux, mais ils doivent être considérés comme des compléments à l'analyse humaine, et non des substituts. La finesse de l'expertise humaine est indispensable pour affiner et interpréter les résultats, en particulier dans le cas d'une œuvre aussi riche et complexe que celle de Tolkien.

En conclusion, notre projet a démontré que, malgré les défis, l'intégration de la NLP dans l'analyse littéraire ouvre des avenues passionnantes pour l'exploration textuelle. Les résultats obtenus sont prometteurs mais mettent également en évidence la nécessité d'une collaboration étroite entre les technologies de pointe et les spécialistes du domaine pour déchiffrer les couches profondes de textes aussi riches que ceux de Tolkien.

