



Rapport de stage De Master 2 Informatique ATAL

Alexandre TABOT

Stage Recherche

CATEGORISATION AUTOMATIQUE DE PERSONNAGES POUR LA
SYNTHESE DE LIVRES AUDIO

22 août 2019

Rapport de stage Présenté en vue d'obtenir Le Diplôme de Master Informatique ATAL

Alexandre TABOT

Stage Recherche

CATEGORISATION AUTOMATIQUE DE PERSONNAGES POUR LA
SYNTHESE DE LIVRES AUDIO

ENTREPRISE : Laboratoire d'Informatique de l'Université du Mans (LIUM)

Responsable du stage : Marie Tahon

Référent enseignant : Marie Tahon

RESUME :

Nous avons exploré la catégorisation des personnages au sein de livres audio sur un corpus annoté. Nous nous appuyons sur les descripteurs prosodiques et linguistiques pour effectuer un clustering de type KMeans. Nous appliquons ensuite sur ce même corpus, un algorithme de classification SVM pour corroborer les tendances observées lors du clustering. Nous nous appuyons également sur le ressenti perspectif. Cette étude a pour but de mettre en lumière les éléments nécessaires au système de synthèse vocale afin qu'il puisse produire un jeu de voix capable de narrer correctement les œuvres diverses de la littérature en français.

Mots clés : Catégorisation, livres, audio, clustering, classification

Table des matières

1	Introduction	1
1.1	Preamble	1
1.2	Contexte	2
2	Etat de l'art	3
2.1	La construction des corpus dans Le champ de recherche sur les livres audios	3
2.2	Le type de clustering et de classification pour les livres audios	5
2.3	L'étude et l'extraction d'information sur les romans	7
3	Objectif de recherche	8
4	Expérimentation.....	9
4.1	Protocole	9
4.1.1	<i>Les descripteurs utilisés.....</i>	<i>9</i>
4.1.2	<i>Les deux groupes Personnage et Narrateur.....</i>	<i>11</i>
4.1.3	<i>La méthode appliquée.....</i>	<i>12</i>
4.2	Le Deroulement de l'experience.....	13
5	Résultats	15
5.1	Presentation	15
5.2	Analyse des resultats obtenus	16
5.2.1	<i>Répartition des clusters issue du k-means clustering</i>	<i>16</i>
5.2.2	<i>Barycentre des clusters.....</i>	<i>21</i>
5.2.3	<i>Eléments sémantiques et perceptifs mis en avant par les descripteurs.....</i>	<i>22</i>
5.3	Validation des resultats	23
5.3.1	<i>Validation par la classification SVM.....</i>	<i>23</i>
5.3.2	<i>Validation par le perceptif.....</i>	<i>27</i>
6	Conclusion.....	28
6.1	Retour du projet.....	28
6.2	Apport du stage au sein du projet	29
6.3	Perspectives pour le projet	30
6.4	Bilan personnel	31
7	Références.....	32
8	Glossaire / Abréviation	33
9	Annexes	1

Remerciements

Je tiens à remercier tout particulièrement ma responsable de stage, Marie Tahon pour m'avoir proposé ce stage de recherche. Son accompagnement et son aide tout au long de ce stage furent très précieux.

Je tiens à remercier tous mes collègues du laboratoire pour les conversations enrichissantes, leur soutien et les différents événements auxquels j'étais invité à participer. Ceci m'a permis de connaître les différentes facettes du monde de la recherche.

Je voudrais également dire un grand merci à tous mes professeurs présents durant mon cursus universitaire en informatique. Ils m'ont permis d'avoir une grande ouverture sur le monde de l'informatique et ainsi acquérir des connaissances diversifiées.

Enfin, je souhaiterais remercier tous mes camarades étudiants, spécialement ceux de cette 2e année de Master 2. L'entraide et la dynamique de notre groupe ont été essentielle pour que cette année soit très enrichissante et mémorable.

1 Introduction

1.1 PREAMBULE

Ce stage est un stage de recherche réalisée au sein du Laboratoire d'Informatique de l'Université du Mans, le LIUM. Le LIUM est composé de 2 équipes. L'équipe LST, Language Speech Translate, est axée sur la parole. Elle traite les champs de la traduction de la parole, de la synthèse de la parole et de la reconnaissance de la parole. Chaque champ à des sous domaines précis comme la reconnaissance du locuteur, ou différents types de paires de langues pour la traduction.

La seconde équipe, IEIAH, Ingénierie des Environnements Informatiques pour l'Apprentissage Humain, est axée sur les jeux sérieux. L'équipe IEIAH s'occupe plus spécifiquement des environnements informatiques pour l'apprentissage humain par les outils de suivi et d'aide à cet apprentissage. Cette équipe est aussi présente dans l'IUT de Laval, spécialisée également dans la réalité virtuelle.

Ce stage s'est fait au sein de l'équipe de la parole, dans le champ de la synthèse vocale, sous la direction de Marie Tahon. Il est relié à un autre stage fait par Thomas Granjon sur la comparaison et l'évaluation de vocodeurs pour la synthèse neuronale, ainsi qu'avec d'autres stages axés sur l'annotation de livres audio. Il fait partie d'un projet d'élaboration automatique de livres audio en voix de synthèses.

1.2 CONTEXTE

Ce stage a été réalisé dans le cadre de ma fin d'études de Master 2 d'informatique ATAL, Apprentissage et Traitement Automatique de la langue.

Le but de ce stage recherche, est de mener une démarche scientifique complète sur le thème de la catégorisation de personnages dans les livres audio. Cette démarche permet également de voir tous les éléments annexes propres au travail de recherche, comme les conférences, les séminaires, les journées d'étude.

La raison de ce choix de sujet de stage, est que l'étude définissant l'expressivité d'une voix pour la narration d'histoire est à l'heure actuelle très peu explorée. Le développement de la demande de livres audio est très important. Or l'enregistrement de livres audio de qualité nécessite du temps et des lecteurs ou lectrices expérimentées. Actuellement, la lecture de livres numériques par voie de synthèse, notamment celle des assistants vocaux commercialisés, sont loin de donner des résultats comparables à des lecteurs humains. Ce stage s'inscrit dans un projet plus global de l'équipe de parole du laboratoire, visant à créer une voix de synthèse capable de narrer une histoire qui soit, au niveau perceptif, confortable. Ce projet est notamment constitué de plusieurs stages de recherche en parallèle.

Nous verrons dans l'état de l'art que la voix est surtout étudiée sur le plan de l'émotivité ou de l'intelligibilité. Des corpus annotés d'extraits d'enregistrements vocaux sont en construction. Les corpus sont des ensembles de données sélectionnés et regroupés afin de réaliser plusieurs objectifs, dont la création de réseaux de neurones capables d'apprendre comment réaliser une tâche avec des exemples annotés, et ensuite de réaliser cette tâche lors de nouvelles situations.

La catégorisation de personnages de livres audios, explore les différents descripteurs prosodiques et linguistiques pouvant être extraits des corpus, afin d'explorer les phénomènes qui sous-tendent une bonne narration. Lorsque ces phénomènes seront identifiés de manière complète, la connaissance de ces phénomènes pourrait être transmis à un système de synthèse vocale « lecteur ».

2 Etat de l'art

Nous allons présenter l'état de l'art des différents domaines sur lesquels s'appuie la catégorisation de personnages pour les livres audio. Nous commençons tout d'abord par le domaine des corpus de livres audio, des techniques et outils d'annotation de ces corpus spécifiques. Nous verrons ensuite les méthodes de clustering et de classification pour la catégorisation de groupe au sein de livres, ou plus généralement d'œuvres littéraires. Nous terminons par un état de l'art sur le genre littéraire présent dans notre champ d'étude et les méthodes d'extraction d'informations sur ce champ.

2.1 LA CONSTRUCTION DES CORPUS DANS LE CHAMP DE RECHERCHE SUR LES LIVRES AUDIOS

Nous allons aborder la présentation du corpus exploité. Les méthodes d'annotation du corpus SyPaFlex¹ ont pour but de mettre en lumière l'émotivité présente dans le discours de la narration des livres audio, afin d'étudier plus précisément ces mécanismes. Lors de la création du corpus, des descripteurs prosodiques ont été définis, comme il est mentionné dans la partie 4.1 Prosodic Patterns de l'article associé à la création du corpus SyPaFlex¹. Il est à noter que la méthode implémentée pour l'obtention de la fréquence fondamentale, un des descripteurs prosodiques défini, est la méthode EPS (Talkin, 1995)

Nous avons utilisé des livres qui proviennent de la base Librivox, Ces livres sont une partie du corpus SynPaFlex¹ de 87 heures d'enregistrement audio de bonne qualité. La base Librivox, couvre différents genres littéraires : roman, nouvelle, conte de fées, fables et poèmes. Ce large éventail littéraire permet d'étudier différents phénomènes phonétiques, linguistiques, acoustiques et phonologiques sur un vaste corpus d'enregistrement. Il permet également d'étudier la méthode d'élocution de longs discours par des compteurs d'histoire. La lectrice de ces livres est une femme, qui n'est pas une professionnelle du doublage, mais une amatrice expérimentée. Cette lectrice utilise un style de voix expressives et donne à chaque personnage une voix propre, pour que ceux-ci soient distincts, le plus possible pour les auditeurs. Ceci donne des récits racontés expressifs. Le corpus ainsi construit répond au 5 contraintes¹ fixées, en prélude de son élaboration (Tahon et al 2015) :

- une quantité large de données exprimées par un seul lecteur
- la disponibilité des textes lus à voix haute par ce lecteur unique.
- La grande variété de styles de discours et des genres littéraires lors de ses lectures à voix haute.
- Expressivité des émotions dans ces récits comptés.

Nous allons passer à la présentation de l'outil ROOTS². ROOTS² est un outil d'annotation de livres audio et d'extraction de caractéristiques prosodiques et

¹ Sini et al., [Référence 01](#)

² Chevelu, Lecorve, et Lolive, [Référence 04](#)

linguistiques. Ces éléments sont synthétisés dans un fichier JSon. Il permet la construction simplifiée de gros corpus d'écoute. Il reprend un ensemble très complet de caractéristiques. Cet outil a permis de construire un corpus annoté sur 1300 e-books issus du site Web e-books gratuits (Chevelu et al 2014) ². Ce corpus est une collection de livres audio en langue française, où seul un lecteur est présent par livre. L'organisation et la synchronisation des différents descripteurs pris en charge par l'outil, est présenté par le schéma de l'organigramme de ROOTS².

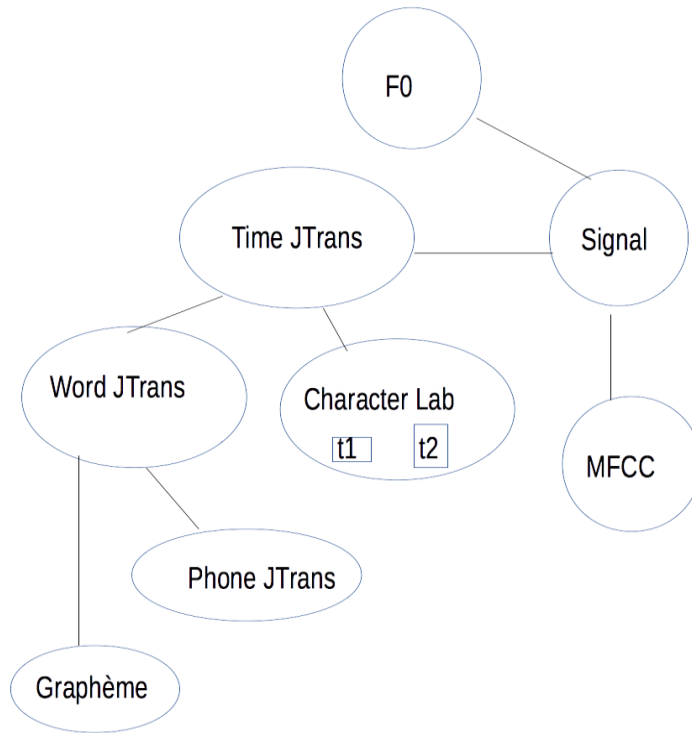


Schéma 01 : Organigramme partiel des relations de synchronisation des données descripteurs de ROOTS²

2.2 LE TYPE DE CLUSTERING ET DE CLASSIFICATION POUR LES LIVRES AUDIOS

Pour l'état de l'art de l'utilisation des méthodes de clustering et de classification, nous nous basons particulièrement sur l'ouvrage : le machine Learning avec python³. Les différents types de méthodes abordées dans cet ouvrage, sont présentées de manières moins détaillées dans les autres publications⁴ lues pour le projet durant ce stage.

Le clustering est une méthode qui appartient à l'apprentissage non supervisé au sein du machine Learning. Ce type de méthode est très bien approprié pour des données de nature qualitative et dans la phase exploratoire de nouveaux ensembles de données sans connaissance préalable. Il existe 3 méthodes couramment utilisées : le partitionnement en k-moyenne (K-Mean), le clustering agglomératif et le DBSCAN (Density-Based Spatial Clustering of Applications with Noise).

Le partitionnement en k-moyenne est un algorithme très rapide et optimisé, mais il nécessite d'avoir une connaissance a priori des tendances que l'on cherche à mettre en évidence au sein d'un ensemble d'éléments. Cette connaissance a priori nous permet de fixer le nombre de clusters que nous souhaitons obtenir.

Le clustering agglomératif ne nécessite pas de connaissances a priori sur les éléments. Les clusters obtenus sont cependant sous-tendus par une relation hiérarchique. Ceci est une hypothèse forte sur le corpus des éléments étudiés.

Le DBSCAN, est un algorithme à plus long terme d'exécution que les deux algorithmes précédemment décrits. Il ne nécessite pas de connaissances a priori et n'impose pas d'hypothèse forte sur les éléments de l'apprentissage. Il calcule la densité de chaque groupe d'éléments possibles, pour ensuite former des clusters ayant des densités homogènes et distinctes entre les clusters.

Ces trois méthodes permettent de moduler la granularité des clusters et ainsi, de sélectionner le degré de finesse de l'analyse de données soumises pour l'apprentissage. La méthode des K-moyennes est généralement utilisée comme première méthode d'exploration de nouvelles données, d'après l'expérience acquise dans les cours de 2e année de Master ATAL de l'université du Mans, et d'après la lecture de différents ouvrages et publications sur le Machine Learning.

La classification appartient au domaine de l'apprentissage supervisé au sein d'une machine Learning. Les principaux modèles pour la classification, sont : les K plus proches voisins (KNN), les classifieurs bayésien naïfs, les arbres de décision et leurs ensembles, et les séparateurs à vastes marges (SVM). L'apprentissage supervisé est constitué d'une phase d'entraînement et d'une phase de test. La phase d'entraînement permet d'apprendre au système la méthode de classification voulue grâce à des données annotées comportant la classe d'appartenance correcte. Les données ou éléments sont présentés au système, celui-

³ C.Müller et Guido, [Référence 05](#)

⁴ Pedregosa et al., [Référence 07](#)

ci les classer, puis une erreur sur la classification est calculée par rapport à l'annotation correcte. Le système adapte alors son modèle en fonction de cette erreur pour terminer ce cycle d'entraînement, avant de recommencer le cycle avec de nouvelles données. La phase suivante, celle du test, permet d'évaluer la capacité du modèle appris, à prédire la classe de nouveaux éléments qui lui sont présentés.

Dans l'apprentissage supervisé, les K plus proches voisins est une méthode très usitée. Elle vise à regrouper les éléments qui sont proches en distance quand ils sont projetés dans un espace de représentation spécifique. Cette méthode nécessite de construire cet espace de représentation pour pouvoir évaluer les distances entre les éléments, et ainsi les classer.

Les classifieurs bayésiens naïfs appliquent le même principe que les modèles linéaires. Ils construisent des frontières de classe par assemblage de lignes droites. Ces lignes sont définies par des fonctions mathématiques linéaires. Ces classifieurs permettent un apprentissage rapide mais avec une capacité de généralisation réduite. Ils sont une des premières méthodes de base de l'apprentissage supervisé.

Les arbres de décision se basent sur l'homogénéité des classes obtenues, après une succession de choix de séparation d'éléments par rapport à des paramètres ou des caractéristiques intrinsèques aux éléments. Il existe des méthodes ensemblistes de ces arbres de décision, comme le Random Forest, qui permet d'être plus souple aux situations complexes. En effet, l'arbre de décision tend à établir des règles rigides, ce qui limite sa capacité de généralisation, d'adaptation à des cas non rencontrés lors de l'entraînement.

Les Séparateurs à Vastes Marges (SVM) ou Support Vector Machine, ont pour objectif de maximiser les marges aux abords des frontières de classe. Ceci permet de prendre en compte plus facilement et correctement les nouvelles données rencontrées après la phase de l'entraînement, au cours de l'apprentissage du modèle. Le SVM a donc un pouvoir de généralisation plus important. Le SVM est utilisé comme système de validation d'un modèle issu de l'apprentissage non supervisé car il a un pouvoir de généralisation fort, tout en définissant des frontières de classe dont les paramètres sont facilement récupérables.

2.3 L'ETUDE ET L'EXTRACTION D'INFORMATION SUR LES ROMANS

Dans les études littéraires sur les romans du XIX^e siècle auxquels appartiennent la majorité des œuvres du corpus de notre recherche, les champs lexicaux employés ainsi que le vocabulaire sont des éléments importants. La richesse et la diversité de ces derniers pour chaque groupe de personnages identifiés au cours du déroulé de l'histoire sont souvent utilisés comme base pour la description d'autres phénomènes littéraires. Une étude sur l'œuvre de Victor Hugo⁵, composé de romans a mis en évidence la relation entre fréquence des mots, champs lexicaux et richesse du vocabulaire. Nous avons également exploré l'état de l'art au niveau de la catégorisation des personnages en études littéraires sur des œuvres appartenant au même genre littéraire que celle étudiées dans le cadre de ce stage. Il est apparu que le nombre moyen de personnages clés des œuvres est de 6.

Dans le domaine de la notation automatique, une publication sur l'intégration de données lexicales dans un CRF⁶, permet d'apprendre plus aisément un système de segmentation et d'annotation automatique robuste. Les systèmes d'identification d'entités nommées⁷ actuelles apportent des informations supplémentaires sur la structure narrative d'un roman de manière automatique. Ces contributions facilitent l'utilisation d'outils d'annotation, comme ROOTS, décrit précédemment, et la constitution de grands corpus variés. La plus grande diversité des informations extraites des romans, favorise la recherche de relations entre phénomènes textuels et phénomènes perçus dans la voix de lecture.

Ces phénomènes de voix de lecture s'intègrent dans la classification des types de discours. Cette classification a plusieurs branches, dont celle des livres audios⁸, et celle de la classification entre discours narratif et discours direct⁹. Ces phénomènes vocaux interviennent également dans l'identification des personnages dans des livres audio pour enfants.

Ces champs de recherche sont très utiles pour le projet car ils abordent tous une des facettes de la catégorisation de personnages. Cette catégorisation se fait sur plusieurs plans, notamment prosodiques et linguistiques. La prosodie concerne les caractéristiques de la voix qui lit un livre audio, pour notre domaine de recherche. La linguistique concerne le texte qui est lu dans un livre audio.

⁵ Brunet, [référence 03](#).

⁶ Constant et al., [référence 06](#).

⁷ Bikel, Schwartz, et Weischedel, [référence 02](#).

⁸ Aghilas Sini, Delais-Roussarie, et Lolive, [référence 01](#).

⁹ Tahon et Lolive, [référence 09](#).

3 Objectif de recherche

L'objectif est de catégoriser, et d'évaluer les personnages de livres de manière automatique. Nous cherchons les caractéristiques distinctives d'un personnage, afin d'attribuer une voix de synthèse cohérente à ce personnage. Le récit ainsi synthétisé sera plus intelligible et fluide, que les récits par voix de synthèse actuelle.

Nous l'avons vu dans l'état de l'art, l'identification et la classification d'œuvre narrée est une préoccupation actuelle. L'objectif de la recherche entrepris durant ce stage, est également d'apporter à ces deux domaines, des éléments supplémentaires. Cet apport est attendu via l'exploration des phénomènes présents dans le corpus d'œuvres étudiées. Nous souhaitons également apporter une contribution à la normalisation des méthodes de construction de corpus de livres audio.

Nous pensons que la catégorisation de personnages dans le cadre de livres audio pourra aider à la catégorisation de contexte d'emploi de la synthèse vocale. Ces contextes d'emploi sont en expansion actuellement, du fait de l'intégration de systèmes de synthèse vocale dans un nombre de plus en plus important d'objets connectés. Les éléments de contexte des œuvres lues seront donc également observés.

Dans le cadre plus global du projet comprenant plusieurs stages dont celui-ci, nous voudrions participer à l'amélioration l'évaluation de la perception de l'écoute de voix humaine et synthétique. Une meilleure évaluation nous permettrait de réaliser des livres audios par voix de synthèse de meilleure qualité. Une plateforme d'évaluation perceptive d'écoute est réalisée par Thomas Granjon lors de son stage dans le cadre de ce projet global.

4 Expérimentation

4.1 PROTOCOLE

Nous allons maintenant décrire le protocole que nous avons mis en place lors de l'expérience. Nous verrons d'abord les descripteurs utilisés, puis les groupes attendus au sein des livres audio, et enfin, la méthode appliquée durant la réalisation de l'expérience. Ce protocole a pour objectif de faciliter la réitération de l'expérience, dans les mêmes conditions que celles rencontrées lors de ce stage.

Les données sources de l'expérimentation du stage, sont accessibles sur demande auprès de Damien LOLIVE du laboratoire IRISA. Il est l'un des participants à la création de l'outil ROOTS. Les demandes d'accès peuvent se faire par mail, à l'adresse suivante :

Damien.lolive@irisa.fr.

Il souhaite contrôler leur utilisation en sachant qui les utilise et à quelles fins. Cela permet de créer une communauté autour de l'outil ROOTS et de mutualiser l'expérience d'utilisation de cet outil. La remontée d'informations sur l'emploi et les possibles nécessités de mises à jour de l'outil ROOTS, est également, une des finalités de ce suivi.

Les productions de cette expérimentation durant ce stage sont de trois types. Un outil informatique sous la forme d'un notebook Jupyter en Python a été créé dans un premier temps. Les fichiers de sauvegarde des résultats obtenus par l'utilisation de ce notebook ont été créés tout au long de l'expérimentation. Une documentation du notebook et des résultats, a été faite lors de l'analyse de ces résultats. Ces productions, ainsi que le présent rapport de stage, sont disponibles sur mon espace Github :

<https://github.com/AlexTAB/CategorisationPersoLivreAudio.git>

4.1.1 Les descripteurs utilisés

Nous avons utilisé des descripteurs prosodiques, c'est-à-dire des descripteurs de la voix dans les différentes phases de lecture des œuvres. Ces phases de lecture sont les passages, ou segments, associés à chacun des personnages distincts. Nous utilisons quatre descripteurs prosodiques.

Le descripteur Fo est la fréquence fondamentale. Elle permet de caractériser un extrait sonore du grave vers l'aigu. La fréquence fondamentale d'un signal audio est un indicateur de tendance des fréquences composant ce même signal audio. La fréquence perçue par l'oreille humaine est établie sur une échelle logarithmique. C'est une caractéristique couramment utilisée dans l'étude des signaux sonores, dont fait partie la voix. La méthode implémentée pour l'obtention de la fréquence fondamentale, est la méthode EPS (Talkin, 1995), d'après la publication de la construction du corpus SynPaFlex¹⁰. La fréquence fondamentale, Fo est constituée dans cette expérimentation de

¹⁰ Sini et al., [référence o8](#).

quatre composantes sur un segment audio : la moyenne du descripteur sur le segment, l'écart type, le minimum et le maximum.

Le descripteur du speech Rate est la vitesse du discours prononcé. Cette vitesse est facilement perceptible par l'oreille humaine, elle peut donc être annotée de manière simple. Elle est influencée par l'état émotionnel, la personnalité du locuteur. Le speech Rate est facilement quantifiable. Il est constitué pour l'expérimentation de quatre composantes pour chaque segment : la moyenne, l'écart type, le maximum et le minimum.

Le descripteur de la durée utilisée est le temps de chaque segment de parole attribuée à un personnage distinct dans un paragraphe de texte. En effet, l'unité de base de l'annotation de l'outil ROOTS, est l'utterance. Cette unité, utterance, a pour taille maximale, un paragraphe de texte. Le descripteur de la durée permet de quantifier la participation un personnage à l'histoire.

Le descripteur MFCC est constitué de l'énergie et de 12 coefficients cepstraux de MEL (Gravier, 2003) selon les spécifications du corpus SyPaFlex¹⁰. Ils sont couramment utilisés dans l'identification du locuteur. Ils ont une grande richesse de variabilité, par leur nature multidimensionnelle. Cette richesse est un atout pour les tâches de clustering et de classification. La catégorisation de personnage partage des objectifs communs avec l'identification du locuteur et la tâche de classification. Les constituants de ce descripteur MFCC, lors de cette expérimentation, sont formés chacun, de quatre composantes pour chaque segment : la moyenne, l'écart type, le maximum et le minimum.

4.1.2 Les deux groupes Personnage et Narrateur

Nous débutons cette partie par présenter quelques points de nomenclature. Le personnage narrateur de toutes les œuvres utilisées dans cette expérimentation est représenté par : «_», le underscore. Le narrateur peut être externe à l'action de l'histoire et omniscient. Son rôle est alors essentiellement descriptif. Il peut être, pour des histoires dites « à la première personne », incarné dans le personnage principal. Le lecteur ne perçoit, alors, l'histoire qu'à travers le personnage principal. Les différents types de narrateur forme le groupe Narrateur. Les autres personnages, aussi appelés personnages non-narrateur constituent le groupe Personnage. La distinction de ces deux groupes est le premier objectif établi, pour la phase de clustering au cours de l'expérience. Cette distinction est, en effet, la première étape pour la compréhension de tout type de récit. C'est une distinction aisément validée par le texte.

Nous l'avons énoncée précédemment, la notation spécifiant l'identité du narrateur est commune à toutes les œuvres. Au sein de l'ensemble des œuvres, le narrateur est omniscient et n'est pas associé à un personnage particulier. De plus, il n'y a pas de changement de lectrice dans le corpus d'études de notre recherche. Ces deux facteurs, nous amènent à faire l'hypothèse que les descripteurs prosodiques associés au narrateur, sont constants. Les valeurs de ces descripteurs pour le narrateur nous délivrent donc l'état de voix neutre de la lectrice.

Le groupe personnage, n'a pas de similitude d'annotation entre les œuvres. Pour toutes les œuvres, sauf les contes, les personnages non-narrateur sont nommés Pers avec un chiffre à la suite. La liste détaillée de ces œuvres est présentée dans le tableau des annexes, Tableau Ao2. La numérotation des personnages non-narrateur Pers n'a pas de lien avec une hiérarchisation des personnages dans le déroulement de l'histoire, ou une caractéristique propre au personnage. La validation par le texte de la séparation en sous-groupes du groupe personnage, nécessite donc une phase d'annotation supplémentaire de l'identifiant du personnage. La catégorisation des personnages est exploratoire pour cette expérimentation. Les catégories ne sont pas prédéfinies. Le nombre de sous-groupes a été borné entre 2 et 6, d'après les informations recueillies lors l'état de l'art.

La connaissance du groupe narrateur et du groupe personnage permet à l'auditeur de suivre correctement et aisément le récit. Il est donc essentiel lors de cette expérimentation, de bien décrire les contours de ces deux groupes.

4.1.3 La méthode appliquée

Lors des différentes expériences, constituant de cette expérimentation, durant ce stage, nous avons appliqué une méthode de type exploratoire. Nous avons appliqué le principe du tamis sur les groupes, c'est-à-dire que nous sommes partis d'éléments globaux, nous avons appliqué une granularité plus fine pour chaque classe formant un sous-groupe qui ressortait particulièrement. Nous avons appliqué un clustering de 2 à 6 clusters par pas de 1.

Pour déterminer la pertinence des descripteurs prosodiques, nous avons appliqué un processus agglomératif. Les descripteurs sont pris individuellement, puis nous avons testé les combinaisons de ces descripteurs entre eux, en prenant en compte l'efficacité de chaque descripteur dans la tâche de clustering. L'efficacité est définie comme la capacité à créer des clusters distincts équilibrés.

Nous avons d'abord appliqué ce processus de clustering à une seule œuvre La Fille du Pirate, d'Henri Chevalier, du fait de difficultés de traitement pour fusionner la représentation des différentes œuvres. Nous sommes passés ensuite à l'étude globale du corpus. Nous avons analysé les différents clusters pour pouvoir faire des clustering successifs sur les clusters de personnage présentant des particularités notables.

4.2 LE DEROULEMENT DE L'EXPERIENCE

Le déroulement de l'expérimentation se fait en plusieurs phases. Chaque phase met en place et applique un traitement spécifique sur les données transmises par la phase précédente. Ces données traitées sont ensuite sauvegardées dans des fichiers, afin d'avoir des points de contrôle et de transmissions à la phase suivante, et toute au long de l'expérimentation.

Nous utilisons, au sein du corpus d'œuvres SyPaFlex, les livres suivants : La Fille du Pirate, d'Henri Chevalier ; La Vampire, de Paul Cézanne ; Madame Bovary, de Gustave Flaubert ; Les contes du Sénégal et du Niger, de de Frantz De Zeltner. Nous réalisons l'extraction de données, issues de l'application de l'outil ROOTS sur ces livres, puis nous effectuons des traitements de mise en forme et de tri des données lacunaires. Nous obtenons ainsi les descripteurs prosodiques et linguistiques, dont nous avons déterminé le besoin, grâce aux informations recueillies pendant l'état de l'art et à la planification de cette expérimentation. Nous réunissons les descripteurs prosodiques dans un dictionnaire adapté pour le clustering par la bibliothèque python sklearn. Nous mettons également, des descripteurs linguistiques dans un dictionnaire. Ces descripteurs linguistiques sont composés de fréquences de mots, de fréquences de parole étiquetée et d'un ensemble de vocabulaires propres à toutes les œuvres du corpus.

Nous appliquons ensuite un clustering par K-means de chacun des descripteurs prosodiques, sur chacune des œuvres utilisées dans notre expérimentation. Puis nous traitons l'ensemble du corpus de ces œuvres, avec un clustering des descripteurs prosodiques. Lors de ce clustering, chaque donnée est traduite par la moyenne des valeurs des composantes des descripteurs qui les constituent. Le cluster regroupe les moyennes proches. Cette technique nécessite de déterminer à l'avance le nombre de clusters voulus. Dans un premier temps, nous nous sommes concentrés sur un clustering à deux clusters, appliqué au livre La Fille du Pirate, de Henri Chevalier. Nous trouvons dans ce livre une grande variété de personnages, ce qui explique cette attention.

Nous analysons les clusters grâce à leurs représentations en tableau et en graphique. Afin d'améliorer l'analyse de ces clusters, nous réalisons un complément d'annotations, notamment pour l'identification précise du personnage dans l'œuvre La Fille du Pirate, d'Henri Chevalier. Ce complément est présent dans la partie Annexes, dans le Tableau A01. Cette analyse nous permet d'orienter notre exploration des données vers des phénomènes pertinents.

Un étiquetage des clusters de présence majoritaire est réalisé pour chaque personnage. La majorité est définie comme le pourcentage le plus élevé personnage dans un clusters par rapport aux autres pourcentages de ce même personnage dans les autres clusters. Le cluster majoritaire est donc considéré comme le cluster qui caractérise le mieux le personnage. Nous avons ensuite déterminé la composition des clusters. Cette composition prend uniquement en compte les personnages possédant plus de vingt segments audios et ayant une appartenance majoritaire à un des clusters.

La préparation pour l'application de la méthode de classification SVM est ensuite effectuée. Cette classification SVM est une des deux validations, du clustering KMeans effectué. La classe recherchée est attribuée à un cluster et associée à un segment audio.

Chaque segment est défini par un vecteur des composantes du descripteur prosodique, obtenu lors du clustering. La classification SVM a pour objectif d'apprendre et de prédire l'association de vecteur de descripteur avec une classe, et ceci pour chaque segment audio. Deux types de préparation sont réalisés pour prendre en compte les deux types possibles de classe recherchée. Le premier type de classe est le cluster dans lequel est étiqueté le segment lors du clustering. Le second type de classe est le cluster majoritaire auquel appartient le personnage du segment.

La seconde validation du clustering est la validation par le perceptif. La préparation de cette validation est effectuée par l'analyse de la répartition des segments de chaque personnage dans les différents clusters. Les personnages ayant une répartition équitable de leurs segments, entre des clusters, sont identifiés. L'écoute de ces personnages permet de confirmer ou d'infirmer la différence de classe de leurs segments. Cette écoute permet également d'expliquer cette différenciation effectuée par le clustering. L'analyse par écoute, est réalisée sur l'œuvre, *La Fille du Pirate*, d'Henri Chevalier.

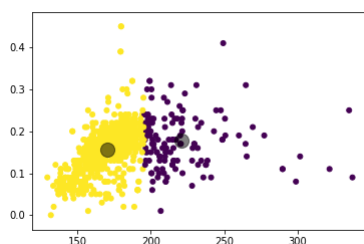
Une analyse des résultats obtenus à la suite des deux validations est menée et permet de formuler des explications sur les phénomènes observés. Une description de ces phénomènes est ainsi produite. Des éléments clés pour la synthèse vocale de livres audio sont établis. L'ensemble de cette production d'analyse et de validation de résultats, donne lieu à la constitution d'une documentation et à l'ouverture de nouvelles pistes pour le projet.

5 Résultats

5.1 PRESENTATION

Nous avons tout d'abord obtenu des tableaux de répartition en clusters, ainsi que la représentation graphique des clusters, pour une meilleure visualisation. Ces éléments ont été sauvegardé dans des fichiers CSV, Texte et Image. Voici un exemple de tableau de répartition et sa représentation graphique associée, pour un clustering avec $k = 2$ sur l'œuvre La Fille du Pirate, d'Henri Chevalier :

Pers	Cluster 0	Cluster 1	Total
—	20	522	542
pers1	9	13	18
pers3	1	0	1
Pers7	0	1	1
pers5	2	0	2
pers2	5	5	10
Total Tout Pers	132	732	864



Représentation 01 : Clustering KMean avec $K=2$ par le descripteur Fo du Narrateur et des personnages Non-Narrateur de l'œuvre La Fille du Pirate, d'Henri Chevalier

Le tableau représente le nombre de segments audio de chacun des personnages attribuer à chaque cluster. Le tableau de la Représentation 01, est un extrait du tableau d'Annexes A03. Chaque ligne représente un personnage, Caractères du livre. Les lignes sont composées de l'identifiant, attribué par l'outil ROOTS, du personnage, du nombre de segments du personnage pour chaque cluster et du total des segments qui sont attribuées à l'identifiant du personnage pour le livre. La dernière ligne est une ligne de totaux de nombre de segments de chaque cluster et de l'ensemble. Elle permet de constater l'importance de chaque cluster. Le nommage ultérieur de chaque personnage sera fait manuellement en s'appuyant sur le livre dans sa version électronique et sur le dictionnaire des paroles, segments, de chaque personnage. Ce dictionnaire est généré automatiquement d'après les données extraites de l'outil ROOTS. Dans ce dictionnaire, les Caractères et leurs segments audio sont également triées par ordre d'apparition dans le livre. Les premiers segments sont les premiers apparus.

5.2 ANALYSE DES RESULTATS OBTENUS

5.2.1 Répartition des clusters issue du k-means clustering

Les proportion d'importance des clusters pour les personnages (% Pers), sont le rapport du nombre de segments, d'un personnage, appartenant à un cluster, sur le nombre total des segments de ce même personnage. Ces proportions sont présentes dans les tableaux de résultats présentés ci-après

Nous constatons sur l'ensemble du corpus des œuvres, que les MFCC et le Speech Rate donne, pour un KMeans avec $K = 2$, un cluster qui est constitué de l'ensemble des paroles, segments, du groupe Narrateur et du groupe Personnage non-narrateur. L'autre cluster ne contient qu'une parole sur les 1967 de l'ensemble. Le pourcentage plus précis sont indiqués dans mon Github du projet, avec les autres tableaux de résultats.

5.2.1.1 *Clustering de l'ensemble des personnages Narrateur et Non-Narrateur*

5.2.1.1.1 Clustering avec les descripteurs prosodiques pris individuellement

En ne considérant que le descripteur FO, nous avons le narrateur qui se situe quasiment dans le seul cluster o. Les personnages non-narrateur du groupe Personnage, se situent quant à eux pour 30 % dans le cluster 1. Ils forment 90% du cluster 1. Les 70 % restants des segments audios des personnages non-narrateur, se retrouve placé dans le cluster o. Nous constatons également que le cluster o est un peu plus de 5 fois plus grands que cluster 1. Le tableau suivant illustre cette analyse :

Pers	Cluster o % Pers	Cluster 1 % Pers	Total segments personnage 100% Pers
Narrateur _	97,68	2,32	1122
Non-Narrateur	70,65	29,35	845
Total	86,07	13,93	1967

Tableau 01 : Répartition des segments audio personnages et leurs proportions dans les clusters selon Fo (Fréquence fondamentale o) pour un KMeans avec $K = 2$

En considérant uniquement la caractéristique durée, nous constatons que les 2 clusters sont équivalents en taille. Nous constatons que pour le narrateur les personnages non-narrateur, le cluster de présence majoritaire est différent. Cependant, cette majorité est très légère (moins de 10 %). Une répartition aléatoirement des segments entre les deux clusters, amènerait ce même constat. La durée ne semble donc pas être un caractère discriminant pour les classes narrateurs/personnages Le tableau suivant illustre cette analyse :

Pers	Cluster o	Cluster 1	Total Personnage
Narrateur _	57,31	42,69	1122
Non-Narrateur	40,12	59,88,	845
Total	49,92	50,08	1967

Tableau 02 : Répartition des segments audio dans les clusters selon la Durée des segments pour un KMeans avec $K = 2$

Nous considérons maintenant les descripteurs Fo, et la durée ensemble. La répartition au sein des cluster est identique à la répartition lorsque seule la Fo est pris en compte. Le poids du descripteur de la durée est donc beaucoup moins important que le poids du descripteur Fo. Il est à noter que le narrateur, en nombre de segments, représente 57 % de l'ensemble des segments du corpus étudié.

5.2.1.1.2 Clustering avec des combinaisons de descripteurs prosodiques

Nous considérons maintenant ensemble la Fo, le Speech Rate, et la durée. La constatation est que la répartition ne change pas significativement par rapport à la situation de l'utilisation uniquement de la Fo.

Nous allons maintenant tester la situation de l'union des deux descripteurs les moins discriminants, constaté de manière individuelle. Nous observons que l'union de ces deux descripteurs, conduit aux mêmes résultats que ceux obtenus lors de l'utilisation individuelle de ces descripteurs. Ils ne discriminent donc pas les personnages.

5.2.1.1.3 Mise en forme et analyse rapide des résultats du clustering

Nous ressentons de cette première phase d'expérimentation, que la Fo possède le pouvoir de discrimination le plus important de l'ensemble des descripteurs prosodiques testés. Le cluster 0 selon Fo permet de discriminer correctement 98 % des segments de narrateur, d'après le Tableau 01. Le cluster 1 permet de discriminer 30 % des segments de personnage non-narrateur. Les personnages non-narrateur représentent 43 % des paroles totales prononcées ou segments. Nous allons utiliser la formule apparentée au TF-IDF, suivante pour calculer le pouvoir de discrimination de Fo :

$$Pd(Fo) = \frac{(Nb_segment_Narrateur_cluster_0 + Nb_segment_Non-Narrateur_cluster_1)}{Nb_total_des_segments_de_toutes_les_oeuvres_utilisées}$$

Application Numérique

$$Pd(Fo) = (1096 + 251) / 1967 = 0,68$$

La Fo discrimine donc 68 % du corpus de segments, correctement. À partir de ces analyses, nous allons explorer plus en détail la répartition des personnages non-narrateur, qui constitue le groupe Personnages.

5.2.1.2 Clustering du groupe Personnages dit Non-Narrateur

Le groupe Personnages des personnages non-narrateur, est maintenant étudié dans le cadre des descripteurs : Fo, la durée, la durée et la Fo ensemble.

5.2.1.2.1 Mesures de synthèse du clustering avec les descripteurs prosodiques sur l'ensemble des œuvres étudiées

Nous allons établir une mesure de pouvoir de discrimination basée sur la formule présentée plus haut. Cette mesure est comprise entre 0 et 1, 1 étant, une catégorisation maximale des personnages par le descripteur. Une telle catégorisation associe un seul Cluster à un seul personnage Pers. Cette mesure permet d'éviter de favoriser les clusters géants qui ne permettent pas de faire de séparation significative.

Descripteurs	Nb de caractéristiques du descripteur	Proportion moyenne de Pers dans Cluster	Proportion moyenne de Cluster dans Pers	F-mesure moyenne
F0	4	0.03	0.71	0.06
Durée	1	0.04	0.95	0.08
F0Durée	5	0.03	0.71	0.06
Mfcc	52	0.01	0.99	0.02
SpRa	4	0.01	0.99	0.02
ToutProsodie	61	0.03	0.71	0.06

Tableau 03 : Synthèse des descripteurs de clustering de k = 6 du Groupe Personnage

Nous voyons dans ce tableau que la durée catégorise mieux les sous-groupes du groupe Personnage. Le descripteur Fo associé au descripteur de la durée, impose sa tendance de catégorisation. Nous allons donc nous concentrer sur la Fo dans la suite de ces analyses. Nous allons, maintenant, mettre en lumière les clusters d'appartenance majoritaire des personnages non-narrateur, sur l'ensemble du corpus des œuvres

5.2.1.2.2 Mise en évidence des clusters majoritaires

Nous allons poursuivre ces analyses du groupe Personnage. Les personnages possédant un nombre de segments inférieur ou égale à 20, sont susceptibles d'avoir la voix neutre du narrateur et d'engendrer du bruit dans la recherche de classes pour les personnages. En effet, ces personnages représentent entre 0 et 2% de l'ensemble des 1967 segments du corpus. Ce sont des personnages de faible importance pour le récit. La probabilité que la lectrice leur donne une voix proche de celle du narrateur est élevée. Ces personnages « bruit », sont donc désormais ôtés du corpus des segments pour les études.

Nous déterminons les clusters majoritaires pour les personnages sur l'ensemble des œuvres. Le cluster majoritaire d'un personnage est défini, par le cluster ayant le nombre le plus élevée de segments de ce personnage, par rapport à l'ensemble des segments de ce même personnage. Le nombre de segments, d'un personnage, appartenant à un cluster sur le nombre total des segments de ce même personnage, forment la proportion d'importance du cluster pour ce personnage. Les clusters majoritaires permettent de former des classes pour la validation des résultats par la classification SVM. Le tableau de répartition des clusters majoritaire, avec les pourcentages des segments, ou proportion d'importance de chaque cluster pour chaque personnages, Pers, est le suivant :

Pers.Livre	C0 % Pers	C1 % Pers	C2 % Pers	C3 % Pers	C4 % Pers	C5 % Pers	Total Segments Personnage 100% Pers	cluster appartenance majoritaire
chevalier_filledupirate. pers1	54,55	9,09	4,55	27,27	4,55	0	22	C0
chevalier_filledupirate. pers11	68,18	9,09	4,55	0	18,18	0	22	C0
merimee_venusdille. pers7	15,38	23,08	15,38	0	46,15	0	26	C4
chevalier_filledupirate. pers12	42,86	14,29	10,71	0	32,14	0	28	C0
chevalier_filledupirate. pers10	25	25	10,71	0	39,29	0	28	C4
chevalier_filledupirate. pers26	64,52	12,9	0	0	22,58	0	31	C0
zeltner_contes. Koli	9,09	18,18	18,18	0	54,55	0	33	C4
merimee_venusdille. pers2	13,51	32,43	0	0	54,05	0	37	C4
merimee_carmen. pers3	0,2444	13,33	17,78	0	44,44	0	45	C4
chevalier_filledupirate. pers25	12,77	44,68	14,89	0,0426	23,4	0	47	C1
merimee_venusdille. pers3	21,43	37,5	17,86	0	23,21	0	56	C1
merimee_venusdille. pers1	22,95	27,87	16,39	6,56	26,23	0	61	C1
merimee_carmen. pers6	2,6	27,27	36,36	14,29	19,48	0	77	C2
Total							513	

Tableau 04 : Répartition du groupe Personnage dans les clusters selon Fo avec K = 6

Nous constatons grâce au tableau de répartition en clusters majoritaires, que quatre clusters sont utiles après l'application du Kmeans, avec K = 6. Ce tableau permet également d'établir des classes pour la méthode de classification SVM. La recherche de la composition de ces clusters majoritaires pour les personnages, donne le tableau suivant :

Cluster / Classe	Personnages Représentés par le Cluster / Classe				
C0	La Fille Du Pirate pers1	La Fille Du Pirate pers11	La Fille Du Pirate pers12	La Fille Du Pirate pers26	
C1	La Fille Du Pirate pers25	Venus d'Ille pers3	Venus d'Ille pers1		
C2	Carmen pers6				
C4	Venus d'Ille pers7	La Fille Du Pirate pers10	Zeltner Contes Koli	Venus d'Ille pers2	Carmen pers3

Tableau 05 : Composition des Clusters-Classes issus du clustering exploratoire

Nous constatons que, trois clusters majoritaires sur les quatre, sont bien équilibrés les uns par rapport aux autres. Le cluster C2 implique un nommage et une étude par le perceptif, du personnage Pers6 de l'œuvre Carmen, de Prosper Mérimée. L'annotation complémentaire de l'œuvre la fille du pirate d'Henri Chevalier, permettra plus loin dans cette analyse, de décrire les trois clusters majoritaires équilibrés. L'œuvre La Vampire de Pierre Féval, n'est pas présente dans ces clusters majoritaires, du fait du nombre réduit de segments attribués à chacun des personnages de cette œuvre. Ce nombre réduit de segments de personnages non-narrateur, a donc ôté l'œuvre du corpus, selon le principe exposé plus haut sur les personnages « bruit ».

5.2.2 Barycentre des clusters

Nous amenons maintenant aborder la partie de l'extraction des barycentres des clusters durant les différents clustering de l'expérimentation. Nous rappelons que la Fo et le speech rate, ont, lors de cette expérimentation 4 composantes : la moyenne, l'écart type, le maximum et le minimum. La durée ne possède qu'une seule composante qui est une valeur de temps. Les MFCC, sont constituées de 13 coefficients qui ont chacun 4 composantes : la moyenne, l'écart type, le minimum et le maximum de leurs valeurs. Le barycentre d'un cluster est le point de gravité de celui-ci, qui possède toutes les composantes du descripteur sur lequel s'applique le clustering.

Le barycentre des clusters permet de caractériser le type de personnage appartenant au cluster selon le point de vue apporté par le descripteur prosodique sur lequel a été basé le clustering. Il permet de créer un modèle de personnages de classe, qu'un algorithme d'apprentissage supervisé de classification, comme le SVM peut apprendre et retrouver dans des données non rencontrées lors de la phase d'entraînement de l'apprentissage supervisé. Le tableau des barycentres de clusters pour le KMeans avec $K = 6$, appliqué sur le Groupe Personnage.

Centre Cluster	Fo moyenne	Fo écart-type	Fo maximum	Fo Minimum
C0	1,59 ^e +02	1,25 ^e -01	1,35 ^e +00	7,93 ^e -01
C1	2,03 ^e +02	1,97 ^e -01	1,52 ^e +00	6,64 ^e -01
C2	2,33 ^e +02	1,88 ^e -01	1,40 ^e +00	6,64 ^e -01
C3	2,85 ^e +02	1,83 ^e -01	1,31 ^e +00	6,46 ^e -01
C4	1,83 ^e +02	1,83 ^e -01	1,56 ^e +00	6,86 ^e -01
C5	-1,15 ^e +01	3,08 ^e +00	-5,03 ^e -01	1,94 ^e 00

Tableau 06 : Tableau des barycentres du K-means($K=6$)

Nous observons que les barycentres sont de taille réduites et proches les uns des autres. Les clusters sont donc cohérents et relativement denses. Le centre du cluster C5 est défini par une moyenne Fo négative. Ceci peut s'expliquer par l'absence de données réelles associées à ce cluster, comme le souligne le Tableau 04.

5.2.3 Éléments sémantiques et perceptifs mis en avant par les descripteurs

Nous avons constaté, d'après le clustering, que la Fo est un descripteur discriminant entre le narrateur et les autres personnages pour un livre utilisant le discours direct. Nous faisons maintenant une étude sur tous les segments, de tous les personnages du livre *La Fille du Pirate* d'Henri Chevalier, afin de faire ressortir des éléments sémantiques et de perception d'écoute d'un livre audio.

5.2.3.1 Éléments sémantiques et perceptifs du livre *La Fille du Pirate* d'Henri Chevalier

Le degré de distinction du genre masculin féminin par la Fo, est un des éléments, qui ressort de cette étude. Il est illustré par le tableau de répartitions des segments audios pour les Caractères principaux du récit de l'œuvre *La Fille du Pirate*, d'Henri Chevalier :

nom Caractère	Id	cluster 0	cluster 1	Total Segments Caractère
Narrateur	-	21	552	573
Charles	pers2	5	5	10
Alphonse Maigret	pers15	3	9	12
Angèle	pers25	29	18	47
Pierre Morlaix	Pers10	9	19	28
Monsieur Jaques	pers26	2	29	31
Tous	Pers	132	732	864

Tableau 07 : Extrait répartition des segments de personnages avec annotation de leurs noms lors d'un k-means avec 2 clusters ($k = 2$) pour la Fo sur l'œuvre *La Fille du Pirate*

Le Tableau 07 montre une catégorisation masculin-féminin, présente, mais moins marquée que celle de Narrateur Non-Narrateur, au vu des écarts de répartition de segments. Le personnage de Charles, dont les segments sont équitablement répartis sur les deux clusters, est un personnage spécifique. La mise en relation du Tableau 05 des clusters majoritaires avec le Tableau 07 des noms des personnages de *La Fille du Pirate*, permet d'apporter une description aux clusters majoritaires. Le cluster majoritaire C0 contient des personnages masculins. Le cluster majoritaire C1 contient des personnages féminins. Les clusters majoritaires discriminent le genre des personnages. Le cluster majoritaire C3 contient des personnages âgés. La discrimination des clusters majoritaires s'applique sur le genre et l'âge des personnages.

L'écoute des personnages spécifiques déterminés par le Tableau 07 et l'observation des représentations de cluster, comme la Représentation 01, permettent de faire des constatations sur la Fo. L'amplitude de Fo pris par la voix du narrateur est beaucoup plus réduite que celle des autres voix des personnages. Ceci permet d'indiquer aux constructeurs de voix de synthèse quelles sont les caractéristiques dont l'apprentissage de la modulation est important.

5.3 VALIDATION DES RESULTATS

La validation des résultats présentés ici, permet d'établir si les résultats de notre recherche exploratoire par clustering sont reproductibles et peuvent être appris par un système automatique. Cette validation se déroule en deux étapes. Une première partie de cette validation est l'apprentissage par la méthode de classification SVM, des classes de personnages, mises en avant précédemment. Cette méthode valide que nos résultats peuvent être appris en apprentissage supervisé. La deuxième phase de la validation est la validation par le perceptif. Cette validation se base sur l'écoute de segments audios, établis comme caractéristiques d'un groupe de personnages.

5.3.1 Validation par la classification SVM

La validation par la classification SVM nécessite de construire un corpus d'apprentissages, composé d'un jeu de données d'entraînement et d'un jeu de tests. Notre corpus de résultats étant de petite taille, nous réalisons, après un entraînement classique, un entraînement avec une cross validation. Le corpus de jeu de données pour la SVM, comprend le vecteur de valeurs des composantes du descripteur du segment audio, ainsi qu'une annotation supplémentaire du cluster auquel ce segment est rattaché. Le cluster de rattachement du personnage lié au segment audio, devient la classe du personnage attendue. La cross validation est une technique qui permet de réaliser l'apprentissage et le test de généralisation du modèle sur un jeu de données de petite taille.

5.3.1.1 *Classification SVM sur l'ensemble des personnages Narrateur et Non-Narrateur*

Nous allons d'abord présenter les résultats de l'apprentissage supervisé de classification SVM, entraîné par les données du clustering sur les deux groupes narrateur et personnage non-narrateur. Ces deux groupes ayant le même cluster majoritaire, le cluster de chaque segment est utilisé comme classe à apprendre pour le modèle. Deux méthodes d'entraînement sont utilisées.

5.3.1.1.1 Méthode d'entraînement classique sur Narrateur / Non-Narrateur pour la SVM

La méthode d'entraînement utilisé est classique. Le jeu du test pour vérifier la capacité de généralisation et de prédiction du modèle ainsi appris, représente 25 % du corpus de données. Le tableau de répartition des classes prédites (en-tête horizontal) par rapport aux classes attendues (en-tête vertical), est associé à la mesure de justesse, accuracy, du modèle. Cet apprentissage se fait en utilisant le module SVM de la bibliothèque python sklearn.

Classes	Co, prédite par SVM	C1, prédite par SVM
Co, définie par KMeans	87	0
C1, définie par KMeans	0	13

Tableau o8 : répartition des occurrences de classe selon les classes prédites et attendues par le modèle de classification SVM sur les classes Narrateur et Non Narrateur par Fo définies par le clustering KMeans.

Le score de justesse du modèle est le suivant : Accuracy 1.0

Le score de justesse obtenue par le modèle est maximum. Ceci peut signifier soit que le modèle est parfait, soit que le modèle sur apprend, car les données utilisées ne sont pas assez nombreuses et variées. En effet un score de justesse de 1 est généralement le signe d'un biais d'apprentissage supervisé. Nous effectuons maintenant un entraînement avec la méthode de cross validation

5.3.1.1.2 Méthode de la cross validation sur Narrateur / Non-Narrateur pour la SVM

Le modèle est maintenant entraîné par la méthode de cross validation, qui consiste à entraîner les modèles par de petits sous-ensembles de données issues de l'ensemble du corpus de jeu de données. Cette méthode est très utilisée pour les corpus de petite taille comme celui du corpus de cette expérimentation. Il permet d'éviter le biais de la concentration de tendances dans certaines parties du corpus. Le tableau de répartition des classe et la précision du modèle issu de ce nouvel apprentissage, sont présents ci-dessous.

	Co	C1
Co	86	0
C1	0	14

Tableau 09 : répartition des occurrences de classe appris selon modèle SVM par cross validation sur les classes Narrateur et Non Narrateur par Fo.

Le score de justesse du modèle est le suivant : Cross Val : Accuracy 1. 1. 1.

Le constat de ce nouvel apprentissage, est que le modèle entraîné par cross validation obtient la même précision maximum que le précédent apprentissage. Il sera nécessaire de confirmer cette performance de détermination des classes narrateur et non narrateur, par un corpus de données beaucoup plus importants. Le biais possible du surapprentissage sera ainsi levé. Nous passons maintenant à l'étude du groupe Personnages, des personnages non-narrateur

5.3.1.2 Classification SVM sur le groupe Personnages : les personnages non-narrateur

Le clustering du groupe des personnages non-narrateur est maintenant soumis à la même procédure de validation. Les personnages non-narrateur ayant plus 20 segments, représentant plus de 10% du corpus des segments de départ, sont intégrés à ce nouveau corpus pour la classification SVM. Les quatre clusters majoritaires mis préalablement en évidence et décrit par le tableau 05 de composition des clusters classes, sont utilisés comme classe pour l'apprentissage. Nous effectuons en premier, comme dans l'étude précédente, un apprentissage avec la méthode d'entraînement classique.

5.3.1.2.1 Méthode d'entraînement classique sur le groupe Personnages pour la SVM

L'entraînement du premier modèle est classique, 25 % des données constituent le corpus de tests pour la prédiction du modèle. Le descripteur étudié est, comme auparavant la fréquence fondamentale Fo. Le tableau de répartition et la justesse du modèle, sont présentés ci-après

	C ₀	C ₁	C ₂	C ₄
C ₀	0	0	0	0
C ₁	3	8	6	13
C ₂	1	3	4	0
C ₄	12	24	6	20

Tableau 10 : répartition des occurrences de classe étiquetée par le clustering et la classification SVM pour les quatre classes du groupe personnages non-narrateur par la Fo.

Le score de justesse du modèle est le suivant : Accuracy : 0.323943661971831

Pour la catégorisation des personnages non-narrateur, la classification et le clustering sont d'accord dans 30 % des cas présentés par le corpus de tests. Ce corpus représente 25 % du corpus de jeu de données utilisé pour l'apprentissage de la classification SVM. La classe C₄ obtient le plus d'accord entre les deux méthodes, avec 20 occurrences d'accord. La frontière des classes est donc plus complexe. Nous passons maintenant à un entraînement de modèle par cross validation.

5.3.1.2.2 Méthode de cross validation sur le groupe Personnages pour la SVM

La cross validation est maintenant appliquée sur l'ensemble du corpus du descripteur Fo. Les personnages non-narrateur ayant plus 20 segments, représentant plus de 10% du corpus des segments de départ, sont les seuls intégrés à ce nouveau corpus de jeu de données, pour la classification SVM. Le modèle est maintenant entraîné par la méthode de cross validation sur trois petits sous-ensembles de données. Le tableau des répartitions des occurrences et la précision du nouveau modèle, sont présentés ci-après.

	C ₀	C ₁	C ₂	C ₄
C ₀	5	3	0	3
C ₁	3	10	6	11
C ₂	1	1	3	0
C ₄	0	20	4	21

Tableau 11 : répartition des occurrences de classe étiquetée par le clustering et la classification SVM par cross validation pour le groupe non-narrateur par la Fo.

Le score de justesse du modèle est le suivant :

Cross-validation scores: [0.51832461 0.27894737 0.36898396]

Une petite majorité d'accord se dégage sur un set de validation. Cette justesse est cependant bien inférieure aux justesses obtenues en général lors de l'utilisation de la cross validation. Ce résultat annonce donc de bonnes pistes, mais nécessite d'être confirmé. La classe C₄, est, comme pour la validation simple, la classe qui obtient le plus grand nombre d'accord entre la classification SVM et le clustering KMeans. Nous analysons maintenant, l'ensemble des résultats obtenues par la validation de la classification SVM

5.3.1.3 *Analyse de la validation par la classification SVM*

Les résultats obtenus permettent de conclure que le descripteur prosodique de la fréquence fondamentale F_0 , a la capacité de différencier le narrateur des personnages non-narrateur. Au sein du groupe des personnages narrateur, quatre pistes de catégories se dégagent. Il est nécessaire de confirmer ces catégories par d'autres descripteurs prosodiques et linguistiques, ou par l'étude plus approfondie des relations entre les différents descripteurs prosodiques utilisés lors de cette expérimentation. Il serait intéressant d'étudier les mécanismes qui donnent à la fréquence fondamentale une prédominance par rapport aux autres descripteurs prosodiques utilisés.

Les valeurs obtenues par la classification SVM durant ce stage sont en dehors de l'intervalle habituel de valeurs obtenues pour ce type de classification. Des erreurs algorithmiques sont donc envisagées. Une vérification est en cours. Les erreurs estimées se situent dans l'utilisation et la séparation de données, soumises au module d'apprentissage SVM de la bibliothèque python Sk-learn. Elles sont donc actuellement, toujours recherchées. Les corrections futures possibles de ces dernières, seront publiées sur mon Github. Les apports futurs de l'utilisation de descripteurs linguistiques supplémentaires, pour la confirmation ou l'infirmerie des analyses des résultats obtenus à ce jour, seront également publiés sur mon GitHub. La date de rendu de ce rapport implique d'exposer ce point d'avancement dans notre recherche avec les résultats mentionnés plus haut.

5.3.2 Validation par le perceptif

La validation par le perceptif s'est déroulée en deux temps : validation d'éléments communs entre personnages d'un même cluster et étude des personnages se répartissant de manière équitable sur plusieurs clusters. Pour réaliser cette validation, nous avons d'abord identifié les fichiers audio contenant des segments caractéristiques. Ces éléments devaient soit appartenir à des personnages du même cluster, soit appartenir au même personnage, mais à des clusters dans lequel les segments de ce personnage sont répartis uniformément.

Nous avons consigné notre ressenti perceptif dans le tableau d'évaluation d'écoute de segments audio de deux personnages. La description du ressenti d'appartenance du personnage à sa classe se fait d'après le genre, le niveau d'importance dans le récit et l'âge supposé du personnage. L'évaluation de la différence de deux segments appartenant au même personnage, est faite en s'appuyant sur le contexte de l'histoire et la connaissance des descripteurs prosodiques utilisés pour l'étude. Une ébauche d'échelle d'évaluation perceptive de comparaison de segment est réalisée. Elle s'appuie sur les retours de l'utilisation de la plateforme d'évaluation de systèmes de synthétiseurs vocaux d'un autre stage du projet.

Personnages	Genre	Jeunesse	Colère
Charles / La Fille du Pirate	Plus Grave	Plus aigu	Plus Grave
Angèle / La Fille du Pirate	Plus aigu	Plus Grave	Plus Grave

Tableau 12 : Evaluation perceptive d'écoute de segments audio de deux personnages.

Les personnages changent de Fo en fonction du degré d'écart de genre, de jeunesse et d'état émotionnel, des personnages qui les accompagne dans les différentes scènes. Les écarts perçus donnent une explication des difficultés d'accord sur les classes des personnages, entre et au sein des méthodes de clustering et de classification, exposés plus haut. L'influence de la présence des autres personnages dans le récit sur l'inflection de la voix donnée à un personnage appuie l'importance de la distinction des voix et donc l'importance de la capacité de modulation vocal du conteur.

6 Conclusion

Pour la conclusion de ce rapport de stage recherche, nous allons faire un retour d'impression sur le déroulé de ce stage et l'avancement du projet global, puis nous verrons les perspectives et enfin nous terminerons par un bilan personnel. Les buts préalables du stage ont été atteints par la réalisation d'une production de recherche scientifique dans un laboratoire. Une connaissance du monde professionnel de la recherche a été acquise, ainsi qu'une certaine expertise dans le domaine de la synthèse vocale et de l'étude des livres audio.

6.1 RETOUR DU PROJET

Ce stage de recherche et cette participation au projet ont été très enrichissants. Les débuts, avec la réalisation d'un état de l'art de niveau recherche, et la familiarisation avec de nouveaux outils, ont nécessité un temps d'adaptation et l'acquisition de nouveaux comportements de travail. Les difficultés de programmation et la recherche de solutions à ces dernières, ont été longues mais ont permis de consolider des savoirs.

Les échanges réguliers avec ma responsable de stage et mes collègues, ont permis de surmonter les difficultés rencontrées lors des différentes phases du projet et du stage. Nous avons pu mettre en évidence des phénomènes de catégorisation, lors de cette exploration de recherche. Les échanges avec les autres stagiaires participant au projet ont été fructueux. L'intégration du projet au sein du laboratoire a bien été réalisée, notamment grâce au petit séminaire du laboratoire et à la plate-forme d'évaluation réalisée dans le cadre du stage de recherche de Thomas Granjon, participant également à ce projet. La diversité des profils des stagiaires du projet était très précieuse. Les événements de recherche annexes à l'expérimentation proprement dite, comme les journées des jeunes chercheurs en acoustique et audition du signal (JJCAAS), ont apporté une ouverture au stage et au projet.

Le stage recherche s'est bien déroulé. Il s'est révélé riche d'expériences à titre personnel. Il a apporté une contribution au projet et au domaine.

6.2 APPORT DU STAGE AU SEIN DU PROJET

Les apports du stage au projet sont de deux types sur le plan qualitatif : les indications des descripteurs prosodiques à fort impact et l'identification des groupes de personnages facilitant l'écoute d'un récit. Les apports, par l'échange d'idées et par l'analyse de conception d'outils, sont également à prendre en compte.

La recherche exploratoire sur l'impact des descripteurs prosodiques a permis de confirmer ou d'infirmer certaines hypothèses. Elle est utile pour la conception de futurs outils d'analyse de livres audio, en indiquant sur quels éléments cette analyse doit être plus fine. Cette recherche exploratoire a aussi permis la construction d'une bibliothèque de fonctions python, pour le traitement et l'analyse des données d'annotation de livres audio. Cette bibliothèque a été conçue en vue d'être facilement réutilisable, modifiable et maintenable.

L'identification des groupes de personnages permettant une facilité d'écoute de livres audios, participe à l'intégration de données textuelles, en complément des descripteurs prosodiques. Ces données contextuelles mises en avant pendant le stage, ont été reliées à des études littéraires et linguistiques. La création de corpus, via l'établissement des différents dictionnaires, durant l'expérimentation du stage, participe à l'avancée de l'apprentissage automatique en analyse de livres audio.

Les apports du stage sur le plan quantitatif, via les différentes productions réalisées pendant le stage, sont présentés dans la liste suivante :

- 1 notebook Jupyter Python, contenant l'outil de l'expérimentation du stage.
- 6 fichiers de données mise en forme pour le clustering des six œuvres utilisées lors de l'expérimentation.
- 6 dictionnaires des segments textuels, ou tirades, de chaque personnage pour chaque œuvre, leur vocabulaire associé, ainsi que les descripteurs linguistiques TF et IDF.
- 1 dictionnaire de vocabulaire de l'ensemble du corpus des œuvres avec le descripteur linguistique TF pour chaque mot.
- 100 fichiers de suivis et de résultats de clustering, comprenant des tableaux et des représentations graphiques.
- 8 fichiers de résultats de classification SVM.
- 1 fichier d'annotation complémentaire de La Fille du Pirate d'Henri Chevalier.
- 1 rapport de stage avec les méthodes utilisées et l'étude du perceptif.
- 1 diaporama de présentation de soutenance de stage.

6.3 PERSPECTIVES POUR LE PROJET

Pour la suite du projet, nous souhaitons aussi étudier les caractéristiques linguistiques, notamment la fréquence d'utilisation d'un vocabulaire, pour la discrimination de personnages, en utilisant la technique du TF-IDF, adapté aux tirades de personnages. L'établissement des liaisons entre les descripteurs linguistiques et prosodiques, par la suite, permettrait de tester l'enrichissement apporté par chaque type de descripteur, comme dans le cas de la publication sur les CRF¹¹.

La définition des groupes de personnages basée sur des études littéraires et sur le ressenti perceptif de l'écoute d'un récit dans un apprentissage supervisé, est une des pistes à explorer pour ce projet de synthétiseur pour livres audio. D'après mon expérience d'écoute régulière de livres audio, les groupes définissant le niveau d'importance des personnages, leur genre et leur âge, sont cruciaux.

¹¹ Constant et al., [référence 06](#).

6.4 BILAN PERSONNEL

Ce stage m'a apporté au niveau personnel, une connaissance du monde de la recherche, à travers toutes les étapes de l'expérimentation, de la rédaction du rapport de stage et de la présentation de la soutenance. J'ai pu également l'appréhender lors d'événements de recherche, comme les mini séminaires du laboratoire et les journées des jeunes chercheurs en acoustique et signal. Les échanges avec mes collègues et ma responsable de stage m'ont également beaucoup appris.

Ce stage m'a également permis d'acquérir plus d'expérience pratique dans l'application des méthodes et concepts appris lors du premier semestre de ce Master 2 ATAL. J'ai également découvert d'autres outils, comme ROOTS. J'ai pu continuer l'assimilation des connaissances acquises tout au long de mon parcours universitaire dans la filière informatique. J'ai acquis des méthodes pour réaliser une veille technologique, afin de continuer à assimiler de nouvelles connaissances futures dans le domaine de l'informatique.

J'ai pour objectif après ce stage de fin de Master, de rechercher du travail dans le domaine du machine Learning appliquée à l'audio, pour acquérir de l'expérience pratique afin de pouvoir poursuivre par une thèse dans ce domaine. Je compte poursuivre les pistes de recherche abordée lors de ce stage, dans le cadre d'expérimentation, à plus petite échelle et sur mon temps personnel, afin de créer mes propres adaptations en livre audio. Ceci me permettrait également de poursuivre mon projet d'aide à l'apprentissage par synthèse vocale de cours ou d'ouvrage de formation.

Pour terminer ce bilan, j'ai acquis une expertise dans l'utilisation de systèmes de reconnaissance et de synthèse vocale, que ce soit par l'étude théorique durant les cours, par l'observation pratique au sein du laboratoire du LIUM, ou par l'utilisation de systèmes pour la rédaction de ce rapport de stage et de tous les éléments de communication que j'ai réalisée durant ce projet. Cette expertise est essentielle dans ma recherche d'emploi et dans les futures situations professionnelles, de formation ou personnelles, que je suis susceptible de rencontrer.

7 Références

- Référence 01 : Aghilas Sini, Elisabeth Delais-Roussarie, et Damien Lolive. « Annotation automatique des types de discours dans des livres audio en vue d’une oralisation par un système de synthèse », s. d., 9.
- Référence 02 : Bikel, Daniel M, Richard Schwartz, et Ralph M Weischedel. « An Algorithm That Learns What’s in a Name », s. d., 21.
- Référence 03 : Brunet, Étienne. « LA STRUCTURE LEXICALE dans l’oeuvre de Hugo », s. d., 20.
- Référence 04 : Chevelu, Jonathan, Gwenole Lecorve, et Damien Lolive. « ROOTS: A Toolkit for Easy, Fast and Consistent Processing of Large Sequential Annotated Data Collections », s. d., 8.
- Référence 05 : C.Müller, Andreas, et Sarah Guido. *Le Machine learning avec Python*. First Interactive. O’REILLY, 2018.
- Référence 06 : Constant, Mathieu, Isabelle Tellier, Denys Duchier, Yoann Dupont, Anthony Sigogne, et Sylvie Billot. « Intégrer des connaissances linguistiques dans un CRF: application à l’apprentissage d’un segmenteur-étiqueteur du français », s. d., 13.
- Référence 07 : Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. « Scikit-Learn: Machine Learning in Python », s. d., 7.
- Référence 08 ; Sini, Aghilas, Damien Lolive, Gaëlle Vidal, Marie Tahon, et Elisabeth Delais-Roussarie. « SynPaFlex-Corpus: An Expressive French Audiobooks Corpus Dedicated to Expressive Speech Synthesis », s. d., 9.
- Référence 09 : Tahon, Marie, et Damien Lolive. « Discourse Phrases Classification: Direct vs. Narrative Audio Speech ». In *9th International Conference on Speech Prosody 2018*, 433-37. ISCA, 2018. <https://doi.org/10.21437/SpeechProsody.2018-88>.

8 Glossaire / Abréviation

LIUM : Laboratoire d'Informatique de l'Université du Mans.

LST : Language Speech Translate qui se traduit par transcription du langage parlé. Cette équipe de recherche du Laboratoire d'Informatique de l'Université du Mans, est axée sur la parole. Cet axe se décline par la reconnaissance, la traduction et la synthèse de cette parole.

IEIAH : Ingénierie des Environnements Informatiques pour l'Apprentissage Humain. Cette équipe de recherche du Laboratoire d'Informatique de l'Université du Mans, est axée sur les jeux sérieux.

IRISA : Institut de Recherche en Informatique et Systèmes Aléatoires.

KMean : Méthode de clustering utilisant un nombre K de moyennes pour former un nombre K de clusters contenant des éléments proches de la même moyenne.

DBSCAN : Density-Based Spatial Clustering of Applications with Noise qui se traduit par Application avec bruit de clustering base sur la densité de zones de l'espace. Cet algorithme de clustering est basé sur le regroupement de zones de densités d'éléments.

KNN : K Near Neighbor qui se traduit par K plus proches voisins. Cette méthode de classification d'éléments est basée sur la distance des éléments entre eux quand ils sont projetés dans un espace de représentation.

SVM : Support Vector Machine qui peut se traduit par Séparateur à Vaste Marge. Méthode de classification visant à maximiser les marges de séparation entres les classes.

CRF : Conditional Random Fields, qui peut se traduit par champs markoviens conditionnels.

JJCAAS : Journées Jeunes Chercheurs en Acoustique musicale, Audition et Signal audio.

TALN : Traitement Automatique du Langage Naturel est une discipline informatique qui vise à extraite des informations de données linguistiques. Elle étudie notamment les données textuelles.

MFCC : Mel Frequency Clepstral Coefficient.

SpRa : Speech Rate. Vitesse du discours prononcé.

Fo : Fréquence fondamentale d'un signal audio.

9 Annexes

Les annexes ci-après, ainsi que le présent rapport de stage, sont disponibles sur mon espace Github :

<https://github.com/AlexTAB/CategorisationPersoLivreAudio.git>

idPersROOTS	PersAnnote	premiereParole	GenreAnnote	CatSoc	Qualife
–	narrateur	chapitre un du prologue	Neutre	non	non
Pers1	Commandant/ capitaine François	Range à carguer la grande voile	Masculin	Moyenne	Marin
Pers2	Charles fils armateur Alycon	Qu'y a-t-il donc ?	Masculin	Moyenne	Armateur
Pers3	chef timonier	Elle gouverne plus	Masculin	Modeste	Marin
Pers10	Pierre Morlaix	Et avant deux	Masculin	Moyenne	Conducteur
Pers4	Capitaine pirate du Corbeau	Comment s'appelle ta coquille	Masculin	Moyenne	Marin
Pers5	Un calier	Nous faisons eau	Masculin	Modeste	Marin
Pers6	Juliette Guilloux / la Camarade	Ah! c'est vous	Féminin	Moyenne	Domestique
Pers7	Second du capitaine François	Le Corbeau	Masculin	Moyenne	Marin
Pers8	Des matelots	Le Corbeau	Masculin	Modeste	Marin Foule
Pers9	Le troisième rameur	Cela signifie	Masculin	Modeste	Marin
Pers10	Pierre Morlaix	Et avant deux !	Masculin	Moyenne	Conducteur
Pers11	Mr Larençon	Eclairez-moi	Masculin	Aisée	Bourgeois
Pers12	Mike	Qu'est-ce qu'il	Masculin	Modeste	Marin
Pers13	Factionnaire 1 sentinelle de prison	Numéro 1. Onze Heures! Rien ne	Masculin	Moyenne	Soldat
Pers14	Factionnaire 2 sentinelle de prison	Numéro 2. Onze Heures! Rien ne	Masculin	Moyenne	Soldat

Tableau A01.P01 : Partie 01 de l'Annotation complémentaire de l'œuvre La Fille du Pirate, d'Henri Chevalier

idPersROOTS	PersAnnote	premiereParole	GenreAnnote	CatSoc	Qualife
Pers15	Alphonse Maigret	Bon; le factionnaire	Masculin	Modeste	Ouvrier / Prisonnier
Pers16	Foule Sentinelles de prison			Moyenne	Foule Soldat
Pers17					
Pers18	Madame Morlaix	Tout de même que c'est z-	Féminin	Moyenne	Commère
Pers19	Madame Raviot	Oui, en effette	Féminin	Modeste	Commère
Pers20	Madame Roger	Moi, d'abord, j'vas	Féminin	Modeste	Commère
Pers21	Commère non nommée	Et que vous aurez raison	Féminin	Modeste	Commère
Pers22	Mère Cadet	P'têt'ben qu'elle était malade	Féminin	Modeste	Commère
Pers23					
Pers24					
Pers25	Angèle	Non, monsieur Jacques	Féminin	Aisée	Principal féminin
Pers26	Mr Jacques	Mais,Angèle?	Masculin	Moyenne	Prétendant
Pers27					
Pers28					
Pers29					
Pers30					
Pers31					
Pers32					
Pers33					
Pers34					

Tableau A01.P02 : Partie 02 de l'Annotation complémentaire de l'œuvre La Fille du Pirate, d'Henri Chevalier

Titre de l'œuvre	Auteur de l'œuvre	Année de parution de l'œuvre
La fille du Pirate	Henri Chevalier	1878
Madame Bovary	Gustave Flaubert	1857
La Vénus d'Ille	Prosper Mérimée	1837
Carmen	Prosper Mérimée	1845
La Vampire	Paul Féval	1865
Contes du Sénégal et du Niger	Frantz De Zeltner	1913

Tableau A02 : Liste des œuvres utilisées lors de l'expérimentation du stage de recherche du présent rapport

Pers	Cluster o	Cluster 1	Total
—	20	552	572
pers1	8	10	18
pers3	1	0	1
pers1/t0	1	3	4
pers5	2	0	2
pers2	5	5	10
pers7	0	1	1
pers8	1	0	1
pers9	1	1	2
pers9/t0	0	1	1
pers4	1	5	6
pers4/t0	0	1	1
pers6	2	11	13
pers11/t0	0	1	1
pers11	3	18	21
pers12	7	15	22
pers12/t0	0	6	6
pers13	1	1	2
pers14	1	1	2
pers15	3	9	12
pers16	0	1	1
pers16/t0	1	0	1
pers10	9	19	28
pers18	3	3	6
pers19	2	2	4
pers20	4	1	5
pers21	2	1	3
pers22	1	1	2
pers23	3	1	4
pers24	1	0	1
pers18/t0	9	4	13
pers25	29	18	47
pers26	2	28	30
pers27	0	1	1
pers27/t0	0	1	1
pers28	2	0	2
pers0/t25	1	0	1
pers29	0	3	3
pers30	2	2	4
pers29/t0	0	1	1
pers31	0	1	1
pers32	1	0	1
pers26/t0	0	1	1
pers33/t0	0	2	2
pers33	1	0	1
pers34	2	0	2
Total	132	732	864

Tableau A03 : Répartition des segments des personnages du livre La Fille du Pirate d'Henri Chevalier par un clustering KMean avec $K=2$.