**Introduction**

Le présent travail a pour but d’étudier la reconnaissance des entités nommées (REN) sur un corpus de poèmes français de dix auteurs du XIXème et XXème siècles. Étant donné que la REN semble sensible à la variation du genre traité, nous cherchons à réfléchir à la si la nature du poème fait que le modèle utilisé pour la REN fonctionne différemment en comparaison avec les romans, par exemple.

Nous avons utilisé 4 modèles différents afin d’observer quel modèle s’adapte mieux au genre concerné : spaCy (modèle large et small), Flaubert, Camembert et Stanford. Pour l’analyse des résultats, nous avons investigué la source du bruit pour chaque modèle et à la suite nous avons mesuré à quel point la recherche d’un meilleur découpage en phrase et d’une casse plus fidèle à la réalité linguistique peut contribuer à améliorer les performances.

Les ouvrages de poèmes ont été chargés sur Gallica, une bibliothèque numérique de la Bibliothèque nationale de France, utilisant le protocole IIIF (International Image Interoperability Framework) pour annoter des documents à travers des images numérisées. Concernant l'organisation du corpus, un seul répertoire contenant des archives lié à chaque ouvrage a été utilisé pour l'extraction des entités nommées.

XXe siècle :

| **Auteur** | **Ouvrage** | **Date de publication** |
| --- | --- | --- |
| Sophie d'Arbouville | Poésies et nouvelles | XIXe siècle |
| Marceline Desbordes-Valmore | Poésies | 1820 |
| Victor Hugo | Les Contemplations, Tome 2 | 1856 |
| Arthur Rimbaud | Illuminations ; Une saison en enfer | 1873-1875 |
| Paul Verlaine | Sagesse | 1881 |

XXe siècle :

| **Auteur** | **Ouvrage** | **Date** |
| --- | --- | --- |
| Guillaume Apollinaire | Calligrammes | 1918 |
| Jeanne Loiseau | Fleurs d'avril | XXe siècle, date exacte à préciser |
| Anna de Noailles | Derniers vers | XXe siècle, après 1930 |
| Cécile Sauvage | Tandis que la terre tourne | 1910 |
| Renée Vivien | Études et préludes | 1901 |

**État de l’art**

L’article de 2025 « Quelle solution pour améliorer les performances de la reconnaissance d'entités nommées sur des données bruitées, corriger l'entrée ou filtrer la sortie ? » traite de l'amélioration des performances de la tâche de reconnaissance d'entités nommées (REN) sur des textes contenant du bruit provenant des erreurs d'OCR (reconnaissance optique de caractères). Les auteurs cherchent à savoir si la correction préalable des textes traités par OCR peut améliorer de manière significative les résultats de la REN ou s'il est plus efficace de filtrer les sorties après traitement. Leurs résultats suggèrent que, lorsqu'il s'agit de textes contenant du bruit OCR, il est plus efficace d'appliquer des techniques de filtrage aux résultats REN que de corriger les textes avant le traitement, afin d'améliorer la précision de l'identification des entités nommées. Les auteurs ont utilisé deux corpus pour comparer les entités nommées reconnues par *spaCy* sur les versions de référence d’une part et des versions OCR, corrigées automatiquement ou non. Nous allons nous inspirer de cette comparaison pour notre évaluation.

**Résultats**

| **Modèle** | spaCy (sm) | spaCy (lg) | Flaubert | Camembert | Stanford |
| --- | --- | --- | --- | --- | --- |
| **Tokens** | 308204 | 308204 | 337523 |  |  |
| **Entités Nommées** | 3446 | 3277 | 1001 |  |  |
| **Densité EN (%)** | 1,12% | 1,06% | 0,0228 % |  |  |
| **Durée** | 1 minute et 10 secondes | 1 minute et 15 | 37 segundos |  |  |

**TABLE DES TOKENS PAR OUVRAGE**

| **Ouvrage** | **Nb de tokens avec**  **spaCy (sm)** | **Nb de tokens avec**  **spaCy (lg)** | **Nb de tokens avec**  **Flaubert** | **Nb de tokens avec**  **Camembert** | **Nb de tokens avec**  **Stanford** |
| --- | --- | --- | --- | --- | --- |
| **Poésies et nouvelles**  **(Sophie d'Arbouville, XIX)** |  |  | 60937 |  |  |
| **Poésies**  **(Desbordes-Valmore, 1820)** |  |  | 35102 |  |  |
| **Les Contemplations, Tome 2**  **(Victor Hugo, 1856)** |  |  | 75902 |  |  |
| **Illuminations ; Une saison en enfer**  **(Arthur Rimbaud, 1873-1875)** |  |  | 23008 |  |  |
| **Sagesse**  **(Paul Verlaine, 1881)** |  |  | 15719 |  |  |
| **Calligrammes**  **(Guillaume Apollinaire, 1918)** |  |  | 19808 |  |  |
| **Fleurs d'avril**  **(Jeanne Loiseau,XXe)** |  |  | 22432 |  |  |
| **Derniers vers**  **(Anna de Noailles, XXe)** |  |  | 14641 |  |  |
| **Tandis que la terre tourne**  **(Cécile Sauvage, 1910)** |  |  | 24103 |  |  |
| **Études et préludes**  **(Renée Vivien, 1901)** |  |  | 45851 |  |  |

**TABLE DES ENTITÉS NOMMÉS PAR OUVRAGE**

| **Ouvrage** | **Nb d’EN avec**  **spaCy (sm)** | **Nb d’EN avec**  **spaCy (lg)** | **Nbd’EN avec**  **Flaubert** | **Nb d’EN avec**  **Camembert** | **Nb d’EN avec**  **Stanford** |
| --- | --- | --- | --- | --- | --- |
| **Poésies et nouvelles**  **(Sophie D'Arbouville, XIX)** |  |  | 127 |  |  |
| **Poésies**  **(Desbordes-Valmore, 1820)** |  |  | 54 |  |  |
| **Les Contemplations, Tome 2**  **(Victor Hugo, 1856)** |  |  | 94 |  |  |
| **Illuminations ; Une saison en enfer**  **(Arthur Rimbaud, 1873-1875)** |  |  | 91 |  |  |
| **Sagesse**  **(Paul Verlaine, 1881)** |  |  | 125 |  |  |
| **Calligrammes**  **(Guillaume Apollinaire, 1918)** |  |  | 75 |  |  |
| **Fleurs d'avril**  **(Jeanne Loiseau,XXe)** |  |  | 87 |  |  |
| **Derniers vers**  **(Anna de Noailles, XXe)** |  |  | 101 |  |  |
| **Tandis que la terre tourne**  **(Cécile Sauvage, 1910)** |  |  | 90 |  |  |
| **Études et préludes**  **(Renée Vivien, 1901)** |  |  | 157 |  |  |

**TABLE DES DENSITÉS PAR OUVRAGE**

| **Ouvrage** | **Densité (%) avec**  **spaCy (sm)** | **Densité (%) avec**  **spaCy (lg)** | **Densité (%) avec**  **Flaubert** | **Densité (%) avec**  **Camembert** | **Densité (%) avec**  **Stanford** |
| --- | --- | --- | --- | --- | --- |
| **Poésies et nouvelles**  **(Sophie d'Arbouville, XIX)** |  |  |  |  |  |
| **Poésies**  **(Desbordes-Valmore, 1820)** |  |  |  |  |  |
| **Les Contemplations, Tome 2**  **(Victor Hugo, 1856)** |  |  |  |  |  |
| **Illuminations ; Une saison en enfer**  **(Arthur Rimbaud, 1873-1875)** |  |  |  |  |  |
| **Sagesse**  **(Paul Verlaine, 1881)** |  |  |  |  |  |
| **Calligrammes**  **(Guillaume Apollinaire, 1918)** |  |  |  |  |  |
| **Fleurs d'avril**  **(Jeanne Loiseau,XXe)** |  |  |  |  |  |
| **Derniers vers**  **(Anna de Noailles, XXe)** |  |  |  |  |  |
| **Tandis que la terre tourne**  **(Cécile Sauvage, 1910)** |  |  |  |  |  |
| **Études et préludes**  **(Renée Vivien, 1901)** |  |  |  |  |  |