

Analyse lexicale innovante par le TALN pour améliorer les compétences rédactionnelles d'étudiants en FLE

Innovative lexical analysis through NLP to improve the writing skills of students of French as a Foreign Language

ADAIKA Radja^{1,*}

¹ Université Echahid Hamma Lakhdhar d'El-Oued, Algérie

Email: adaika-radja@univ-eloued.dz

Date de soumission: 11/08/2024

Date d'acceptation : 22/09/2024

Date de publication: 30/09/2024

Résumé : Le présent article met en avant l'utilisation d'un outil de Traitement Automatique des Langues Naturelles (TALN) dans l'évaluation linguistique. L'étude vise également à rendre cet outil accessible et applicable pour les étudiants et les chercheurs en sciences humaines, qui s'intéressent à l'analyse des textes. En outre, elle vise l'analyse de la richesse lexicale des étudiants de première année de licence en français à l'Université d'El-Oued. La méthodologie adoptée cherche à répondre à la question principale suivante : quel est l'impact des outils TALN sur le développement des approches de l'analyse des textes ? Cette contribution présente une analyse détaillée des textes narratifs du public étudié révélant que l'analyse automatisée améliore l'efficacité en accélérant le processus et en augmentant la précision des résultats. Par ailleurs, elle démontre que les étudiants possèdent un vocabulaire limité, indiquant clairement qu'une amélioration est nécessaire.

Mots-clés: TALN; intelligence artificielle; voyant Tools; richesse lexicale; analyse automatisée.

Abstract: This article focuses on the utilization of Natural Language Processing (NLP) technologies, a subfield of artificial intelligence, for linguistic assessment. The study also seeks to ensure that this tool is easily available and applicable to students and scholars in the humanities who have an interest in text analysis. Furthermore, it aims to analyze the lexical richness of first-year French undergraduate students at the University of El-Oued. The adopted methodology aims to investigate the primary inquiry: what is the influence of NLP tools on the advancement of text analysis approaches? This work provides a comprehensive analysis of the narrative texts from the examined audience, demonstrating that automated analysis enhances efficiency by accelerating the process and enhancing the precision of outcomes. Moreover, it demonstrates that students have a limited vocabulary, clearly indicating that improvement is necessary.

Key words: NLP; artificial intelligence; Voyant Tools; lexical richness; automated analysis.

Introduction :

L'enseignement du français occupe une position privilégiée au sein des matières scolaires (Richard, 2006). L'apprentissage du français langue étrangère (désormais FLE) représente un défi majeur pour de nombreux étudiants à travers le monde. La maîtrise de cette langue est essentielle pour des raisons académiques et professionnelles. Le français revêt une importance prépondérante dans le paysage linguistique mondial, notamment en Algérie (ADAIKA, 2018).

Pour les étudiants universitaires en FLE, la maîtrise de la langue est essentielle et fondamentale, notamment la maîtrise de l'écrit car aujourd'hui, l'écriture est devenue un instrument de pouvoir (Delay-Goyet, 2012). La production écrite en classe de langue est le reflet des compétences de l'apprenant lors du passage à l'écrit (Audras & Ganascia, 2006).

Enseigner à écrire est un défi pour les enseignants, contraints d'adopter « une méthode éclectique » (Puren, 1994), en l'absence d'une approche efficace (Bensalem, 2016), ce qui rend la tâche de l'écriture plus pénible pour les apprenants parce qu'elle exige des efforts qui vont au-delà des connaissances linguistiques que l'enseignant est généralement chargé de transmettre (Bensalem, 2016).

« *Le texte narratif est un type de texte hautement hiérarchisé. Décrivant une histoire fictive, vraisemblable ou non, il se compose d'une suite d'événements se déroulant dans le temps (Adam & Revaz, 1996, p. 42)* » (Dekhissi & Valetopoulos, 2019). Ce type de texte permet l'évaluation des capacités rédactionnelles des étudiants quant à la structure de leurs idées, l'utilisation d'un vocabulaire riche et varié et le respect des règles grammaticales et syntaxiques. Toutefois, les étudiants en FLE rencontrent souvent des difficultés et des défis multiples lors de la production des écrits narratifs, ce qui peut freiner leur progression académique.

Parmi ces défis, nous pouvons évoquer la richesse lexicale limitée. L'expansion du vocabulaire actif et passif des apprenants constitue l'un des principaux défis dans l'enseignement-apprentissage du FLE (Steffens, 2020). Mais la question qui se pose ne se limite pas à : « Comment didactiser le lexique ? » mais d'abord « Quel(s) lexique(s) développer ? » (Leeman, 2005). Dans cette optique, Leeman souligne la particularité du lexique et la complexité de choisir et d'organiser les éléments à enseigner (Steffens, 2020). Ces problèmes peuvent affecter la qualité globale des textes et diminuer l'efficacité de la communication. Par conséquent, il est essentiel de concevoir des approches pour évaluer et améliorer les compétences linguistiques des étudiants afin de les soutenir dans la résolution de ces défis (Znaidi, 2020).

Le présent travail vise à évaluer les compétences linguistiques rédactionnelles des étudiants en FLE à travers l'analyse de leurs textes narratifs. Les principaux objectifs de cette recherche sont les suivants :

- Analyser les textes narratifs des étudiants en termes de richesse lexicale.
- Utiliser un outil de Traitement Automatique du Langage Naturel (TALN) pour réaliser cette analyse.

- Identifier les faiblesses des étudiants afin de proposer des solutions pour améliorer leurs compétences rédactionnelles.

Pour mener cette analyse, nous utilisons Voyant Tools, un outil de Traitement Automatique du Langage. Le TALN est un domaine de recherche en pleine expansion, offrant des outils puissants pour l'analyse linguistique. Les outils TALN permettent d'automatiser l'analyse des corpus textuels.

Cette recherche apporte une double contribution significative. Premièrement, l'intégration d'un outil TALN dans l'évaluation des compétences linguistiques qui permettent non seulement une analyse plus rapide et plus précise des textes, mais aussi une évaluation plus objective et systématique. Cette approche innovante contribue à l'évolution des pratiques des chercheurs en analyse textuelle. Deuxièmement, en identifiant les forces et les faiblesses des étudiants à travers l'analyse de leurs textes narratifs, cette étude fournit des informations précieuses qui peuvent aider à l'évolution des pratiques pédagogiques pour l'amélioration continue des compétences linguistiques des étudiants.

Pour situer notre étude dans le contexte plus large de la recherche existante, il est essentiel d'examiner les travaux existants sur l'analyse des textes narratifs en FLE. La revue de littérature suivante passe en revue les méthodologies employées. En comparant ces approches traditionnelles avec les outils TALN modernes que nous utilisons, nous mettrons en lumière les avancées et les défis actuels dans ce domaine de recherche.

I- L'analyse des textes narratifs en FLE:

L'analyse des textes narratifs, rédigés par des étudiants universitaires ou d'autres apprenants, en FLE a été largement explorée, mettant en lumière différentes méthodes et approches adoptées. D'autres recherches sur les difficultés rédactionnelles (Znaïdi, 2020), (Khelifi, 2019) et (Guettala, 2018) se sont concentrées sur divers aspects tels que la structure narrative, l'utilisation du vocabulaire et les erreurs grammaticales courantes. Ces études ont montré que les scripteurs ont beaucoup de difficultés dans leurs compétences scripturales mettant en lumière les défis spécifiques liés à l'apprentissage du FLE. Les méthodes traditionnelles employées incluent des approches qualitatives et quantitatives ; souvent combinées pour une analyse plus complète. Par le biais de l'analyse qualitative manuelle, les chercheurs identifient et catégorisent les erreurs linguistiques et les éléments narratifs. Bien que cette approche offre une profondeur d'analyse, elle est souvent limitée par le temps et la subjectivité des chercheurs.

L'analyse de contenu et l'analyse de discours sont couramment utilisées pour examiner les textes narratifs. Plusieurs études ont souligné que la stratégie d'analyse privilégiée est l'analyse de contenu. (Richard, 2006). Un mécanisme d'analyse des textes qui utilise des méthodes systématiques et objectives de description, permettant de traiter de manière méthodique le contenu implicite et explicite des textes afin de classer et d'interpréter, par inférence, les éléments qui les composent. L'analyse minutieuse des textes, le sens des messages présents dans les matériels consultés (Richard, 2006).

Une grande variété de mots est un indicateur essentiel de la maîtrise d'une langue. Elle témoigne de la capacité des étudiants à s'exprimer de façon précise et subtile, ce qui améliore la qualité de leurs

textes narratifs. Elle reflète également leur capacité à utiliser un lexique varié et approprié dans différents contextes, ce qui contribue à leur réussite et rendre leur communication plus efficace.

Selon Tournier, « *se servir des fréquences comme moyens, entre autres moyens, de caractériser toute utilisation du vocabulaire, c'est être tenté à chaque moment de sortir du nombre pour accéder à des significations* ». (Tournier, 1980). Les techniques classiques comprennent le calcul manuel de la densité lexicale et la variété du vocabulaire. L'analyse de la diversité lexicale et la mesure de la fréquence des mots offrent une visualisation claire des mots utilisés dans un texte. D'autres analyses complémentaires peuvent offrir des informations plus détaillées, telles que les mesures de similarités lexicales et l'analyse des réseaux sémantiques qui enrichissent l'évaluation de la richesse lexicale.

Ces méthodes peuvent être complétées ou remplacées complètement par des analyses automatisées par le biais des outils TALN tel que Voyant Tools qui est un environnement en ligne permettant la lecture et l'analyse des textes numériques. Cet outil peut automatiser ces processus, fournissant des statistiques précises sur l'utilisation du vocabulaire.

II- L'intégration des outils TALN dans l'analyse textuelle :

L'analyse textuelle est une méthodologie dont l'utilisation devient de plus en plus légitime dans les sciences sociales. Cependant, cela pourrait être entravé si la méthodologie de cette analyse reste restreinte ou cesse de se développer.

Le TALN (Traitement Automatique du Langage Naturel) ou NLP (*Natural Language Processing*) est une branche captivante de l'intelligence artificielle qui se concentre sur la compréhension et à la génération du langage humain par des machines. Elle comprend différentes techniques et applications. Il utilise des algorithmes et des modèles d'apprentissage automatique pour l'analyse, la compréhension et la génération du texte dans une langue naturelle. (Raus et al., 2022) expliquent qu'il est important d'insérer des critères linguistiques et discursifs en amont de l'entraînement des algorithmes d'IA. De leur côté, ces chercheurs (Guaresi & Mayaffre, 2021) déclarent que « *de manière schématique, l'idée fondamentale de l'algorithmique de l'IA est qu'un mot est un neurone et que des relations synaptiques entre ces mots-neurones finissent par faire réseau, par faire texte, par faire sens* ».

Le TALN joue un rôle crucial dans l'analyse textuelle. L'utilisation de ces outils et leurs méthodologies d'évaluation développées, permet de combiner différentes fonctions qui abordent efficacement des tâches complexes. Les chercheurs peuvent analyser de grandes quantités de données textuelles, car ils permettent l'amélioration de la précision et de la performance des évaluations linguistiques, en fournissant de nouvelles solutions afin d'évaluer des aspects linguistiques variés.

Les premiers systèmes de traitement du langage considèrent que la connaissance de quelques mots et d'un nombre restreint de règles syntaxiques suffisent pour accomplir certaines tâches nécessitant l'utilisation du langage (SABAH, 2004).

Globalement, l'incorporation des outils TALN dans les évaluations linguistiques garantit non seulement une amélioration de l'efficacité, mais également, une évaluation plus exhaustive et plus précise des compétences linguistiques grâce aux différentes méthodologies et techniques.

III- Méthodologie de recherche :

Dans le cadre de cette étude visant à explorer l'analyse des textes narratifs par le biais des outils de TALN ; une approche qualitative a été adoptée. Le public étudié comprend des étudiants de première année de Licence du département de français, l'un des 28 départements de l'Université d'El-Oued qui n'a cessé de s'agrandir (Miloudi, 2020), Ces étudiants sont tous de langue maternelle arabe, le français est leur première langue étrangère.

Le corpus soumis à l'analyse est constitué de textes narratifs rédigés par ces étudiants, selon la consigne suivante : « Ecrivez un récit décrivant les activités et les émotions vécues lors d'une journée spéciale en famille, en mettant en avant les moments significatifs. ». Les participants, regroupés en équipes de 3 à 5, ont produit un total de 13 textes. Les copies manuscrites collectées ont été retranscrites et numérisées en version électronique.

Bien que l'on trouve de nombreuses analyses exclusivement qualitatives en analyse du discours (AD), l'informatique et les statistiques ont aussi constitué une composante essentielle de cette discipline depuis ses débuts (Née, 2017).

Selon les besoins de cette recherche, nous avons sélectionné les outils de Voyant Tools, dont la présentation détaillée sera abordé ultérieurement.

IV- Analyse et interprétation des données :

IV-1- Analyse de la Richesse Lexicale :

Pour mener une analyse approfondie de la richesse lexicale des textes narratifs, nous avons utilisé les outils puissants et polyvalents de Voyant-tools. Selon Aurélien Berrien, voyant Tools est un environnement en ligne de lecture et d'analyse de textes numériques (Berrien). Cet instrument a été choisi pour ses fonctionnalités avancées, sa flexibilité d'utilisation et ses multiples modes de visualisation graphique.

Les treize outils offerts par Voyant-tools sont : Cirrus, Termes, Liens, Lecteur, Grappe de termes, Tendances, Termes du document, Résumé, Documents, Syntagmes, Contextes, Lignes de bulles et Collocations.

Ces outils permettent des modes d'affichage offrant une visualisation claire et détaillée des résultats, facilitant ainsi leur analyse et leur interprétation. Des échantillons aléatoires de notre corpus ont été sélectionnés pour illustrer la méthode, comme le montrent les figures ci-dessous.

ADAIK Radja **Analyse lexicale innovante par le TALN pour améliorer les compétences rédactionnelles d'étudiants en FLE**

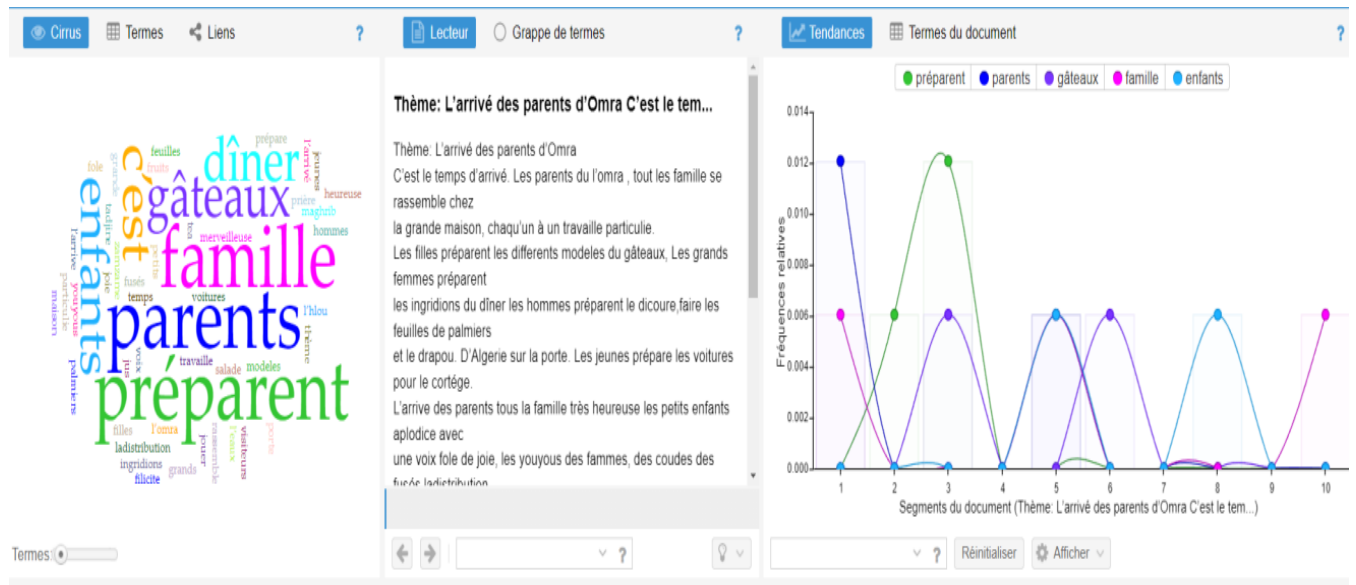


Figure 1 : Outils, Cirrus, Lecteur et Tendances

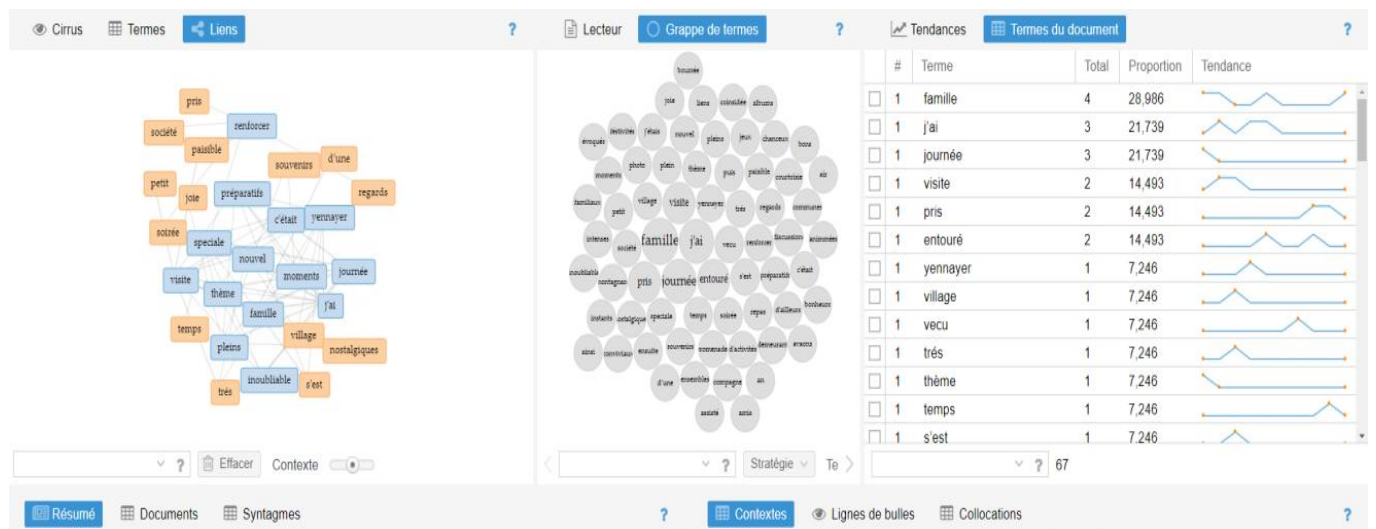


Figure 2 : Outils, Liens, Grappe de termes et Termes du document

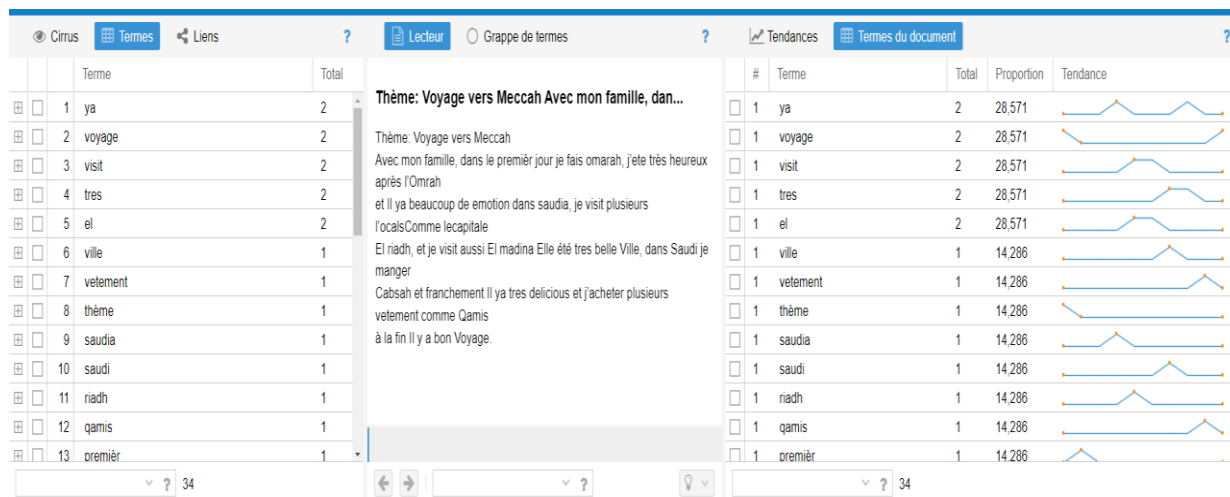


Figure 3 : Outils, Termes, Lecteur et Termes de document

Parmi les treize vues offertes par cette plateforme, pour explorer l'analyse des textes ou des corpus en ligne, nous citons les suivantes :

- **Cirrus** : affiche un nuage de mots des termes les plus fréquents, mettant en évidence les mots significatifs par leur taille.
- **Termes** : liste les mots les plus fréquents et leurs occurrences, fournissant une vue détaillée des termes dominants.
- **Liens** : visualise les relations entre les termes, montrant comment les mots sont connectés.
- **Lecteur** : affiche le texte tel qu'il a été inséré dans la plateforme.
- **Grappe de termes** : identifie des groupes de mots apparaissant fréquemment ensemble, révélant des thèmes et motifs récurrents.
- **Tendances** : représentation sous forme de courbes ou de barres qui suit la fréquence des termes à travers une série de documents ou différentes parties d'un même document, montrant l'évolution de l'utilisation des mots.
- **Termes du document** : représentation sous forme de tableau qui affiche les termes spécifiques à chaque document dans un corpus, permettant des comparaisons intertextuelles.
- **Résumé** : offre une vue d'ensemble du texte traité avec des renseignements généraux sur le texte : Le nombre total de *tokens*, le nombre de mots uniques, le nombre de documents dans le corpus, les mots les plus importants.



Figure 4 : Renseignements offerts par l'outil *Résumé*

Grâce à l'outil *Résumé*, nous avons obtenu des résultats d'une grande précision. Chaque texte a d'abord été analysé individuellement avec l'outil en question, en explorant ses fonctionnalités mentionnées ci-dessus. Les résultats sont résumés dans le tableau ci-dessous pour une analyse comparative. Le tableau inclut le nombre de mots, les formes verbales, la densité du vocabulaire, l'indice de lisibilité, la moyenne des mots par phrase et les mots les plus fréquents de chaque texte.

Tableau : Résultats d'analyse lexicale des Textes

| Textes | Nombre des mots | Formes verbales | Densité du vocabulaire | Indices de lisibilité | Moyenne des mots par phrase | Mots les plus fréquents |
|--------|-----------------|-----------------|------------------------|-----------------------|-----------------------------|---|
| 1 | 125 | 88 | 0.704 | 7.316 | 41.7 | téléphone (2) ; dit (2); " (1); vies (1); vie (1) |
| 2 | 133 | 86 | 0.647 | 5.359 | 11.1 | j'ai (5); spécial (2); père (2); jour (2); famille (2) |
| 3 | 100 | 79 | 0.790 | 11.014 | 50.0 | journée (2); adieu (2); voitures (1); voisins (1); venus (1) |
| 4 | 68 | 56 | 0.824 | 12.986 | 17.0 | Ramadan (2); vraiment (1); village (1); traditionnel (1); thème (1) |
| 5 | 107 | 73 | 0.682 | 6.500 | 21.4 | l'examan (2); jour (2); examan (2); bac (2); vérifié (1) |
| 6 | 116 | 79 | 0.681 | 8.146 | 8.9 | résultat (2); moment (2); jour (2); dieu (2); bac (2) |
| 7 | 70 | 54 | 0.771 | 10.573 | 70.0 | ya (2); voyage (2); visit (2); tres (2); el (2) |
| 8 | 126 | 74 | 0.587 | 8.416 | 25.2 | sœurs (3); sœur (2); salon (2); préparé (2); nouveaux (2) |
| 9 | 123 | 86 | 0.699 | 10.274 | 30.8 | ramadane (2); père (2); nuit (2); frère (2); famille (2) |
| 10 | 166 | 103 | 0.620 | 11.452 | 20.8 | préparent (3); parents (3); famille (3); gâteaux (2); enfants (2) |
| 11 | 115 | 70 | 0.609 | 9.859 | 23.0 | mariage (3); famille (3); frère (2); frère (2); fille (2) |
| 12 | 117 | 83 | 0.709 | 9.476 | 23.4 | n'ai (2); merci (2); maman (2); l'argent (2); l'anniversaire (2) |
| 13 | 138 | 93 | 0.674 | 12.837 | 27.6 | famille (4); j'ai (3); journée (3); visite (2); pris (2) |

Les valeurs moyennes obtenues sont les suivantes :

- Nombre de mots : 78.92
- Formes verbales : 58.61
- Densité du vocabulaire : 0.692
- Indice de lisibilité : 9.554
- Moyenne des mots par phrase : 28.5
- Mots les plus fréquents : père, j'ai, jour, famille

IV-2- Interprétation des résultats :

IV-2-a- Quantité de termes et de formes verbales :

L'étude des textes narratifs met en évidence une différence significative par rapport à la longueur et à l'emploi des formes verbales, ce qui témoigne de la complexité des récits et du style d'écriture des auteurs. Par exemple, le quatrième texte, intitulé "Le Ramadan", est le plus court, comprenant seulement 68 mots. Il se caractérise par une forte densité lexicale, avec des mots particuliers comme « lanternes colorées » et « odeurs délicieuses » qui évoquent l'ambiance festive et sensorielle de cette période. En dépit de la concision du texte, cette diversité d'emplois de l'adjectif et du nom témoigne d'une richesse dans le choix des mots qui contribue à une description vivante.

Par contre, le texte 10, le plus long avec 166 mots, comporte un grand nombre de formes verbales (103), ce qui témoigne d'une dynamique narrative plus développée. L'arrivée des parents de l'Omra est décrite dans ce texte, qui emploie souvent des verbes d'action tels que « préparer », « distribuer » et « accueillir ». Les scripteurs répètent ces verbes afin de mettre en évidence l'activité et l'effervescence qui entourent l'événement, mettant en évidence l'importance des préparatifs et des interactions sociales. La présence de plusieurs verbes d'action dans un texte plus long met en évidence une narration plus dynamique et approfondie, par rapport à des récits plus courts et plus succincts.

Ces exemples démontrent que les disparités en termes et en formes verbales témoignent de choix stylistiques et narratifs variés, ce qui a un impact sur la richesse et la dynamique des récits.

IV-2-b- Densité du vocabulaire :

Le concept de densité lexicale, qui est ici estimé à 0,692 en moyenne, permet d'évaluer la variété des mots en comparant le nombre de mots uniques au nombre total de mots dans un texte. À titre d'exemple, le texte 4 affiche une densité lexicale élevée de 0,824, ce qui témoigne d'une grande variété de mots. Ce texte se caractérise par l'emploi d'un vocabulaire diversifié pour décrire le Ramadan, tels que « lanternes colorées », « délicieuses odeurs », et « repas traditionnel appelé iftar ». Ces sélections de mots témoignent d'une profondeur sémantique et d'une aptitude à susciter des images vivantes et distinctes.

D'un autre côté, le texte 8, qui présente une densité lexicale de 0,587, révèle une utilisation plus limitée du verbe. Les répétitions de termes et d'expressions dans le texte, telles que « préparé nos vêtements » et « préparé le petit déjeuner », peuvent suggérer une restriction de la diversité lexicale et une tendance à recourir aux mêmes mots pour exprimer des idées similaires. Cela pourrait

témoigner d'une maîtrise linguistique plus restreinte ou d'un refus d'explorer une palette de vocabulaire plus étendue.

Ces exemples mettent en évidence la variation significative de la densité lexicale d'un texte à l'autre, ce qui a un impact sur la perception de la richesse linguistique et de la compétence narrative des auteurs.

IV-2-c- Indice de lisibilité :

En fonction de la structure des phrases et du vocabulaire utilisé, l'indice de lisibilité, qui varie entre 5.359 (texte 2) et 12.986 (texte 4), évalue la complexité des textes. Ce critère permet de repérer les niveaux de complexité des textes, qui sont influencés par des éléments tels que la longueur des phrases et la richesse du vocabulaire. Par exemple, le Texte 2 (Indice de Lisibilité: 5.359), intitulé « Jour spécial », offre une structure narrative simple, comprenant des phrases courtes et un vocabulaire accessible. Prenons l'exemple d'une phrase classique du texte : « Le stress est en hausse. Et à partir de 16h, les larmes ont commencé à couler ». La phrase met en évidence l'emploi de structures syntaxiques simples et d'un vocabulaire commun, ce qui rend la lecture plus accessible à un public plus vaste.

D'un autre côté, le Texte 4 (Indice de Lisibilité: 12.986) intitulé « Le Ramadhan » présente un indice de lisibilité supérieur. Le texte emploie des phrases plus longues et un vocabulaire plus descriptif, comme dans l'exemple ci-dessous : « Les familles se réunissaient pour prendre un repas traditionnel appelé iftar afin de rompre le jeûne au coucher du soleil. ». La structure de cette phrase présente une complexité accrue en incluant des détails culturels et des descripteurs, ce qui accroît la complexité du texte et peut rendre la lecture plus difficile pour certains publics. Ces exemples mettent en évidence comment les fluctuations de l'indice de lisibilité traduisent les disparités de style et de complexité dans les textes narratifs étudiés, ce qui a un impact sur l'accessibilité et l'implication des lecteurs.

IV-2-d- Nombre de mots par phrase moyen :

Dans les textes étudiés, le nombre moyen de mots par phrase oscille entre 8,9 mots (texte 6) et 70,0 mots (texte 7), ce qui témoigne d'une grande diversité dans la structure syntaxique des textes. Voici des exemples concrets extraits du corpus afin de mettre en évidence cette diversité : Texte 6 (8,9 phrases par paragraphe): Ce texte utilise des phrases courtes, habituellement d'environ 8 à 10 mots, ce qui favorise un style direct et facile à comprendre. Prenons l'exemple du 16 juillet 2023, une date particulière pour les étudiants de BAC où il s'agit de dévoiler leurs résultats. Un jour différent. Les phrases courtes et concises utilisées contribuent à maintenir une narration claire, ce qui facilite la lecture et la compréhension rapide du texte.

Contrairement à cela, le texte 7 (70,0 mots par phrase) renferme des phrases très longues, parfois complexes, qui peuvent inclure de nombreuses clauses et descriptions. Par exemple, "En compagnie de ma famille, lors de mon premier jour, j'ai participé à l'Omrah. J'étais extrêmement heureux après l'Omrah et il y a une grande émotion en Arabie Saoudite. Je visite différents endroits tels que la capitale Riyadh et Madinah, une ville magnifique en Arabie Saoudite. Je déguste du Kabsa et c'était vraiment délicieux. J'ai également acheté plusieurs vêtements tels que le Qamis. En fin de compte, c'était un voyage agréable." La longueur et la complexité de cette structure de phrase peuvent compliquer la lecture du texte, nécessitant une attention soutenue de la part du lecteur. La différence de longueur des phrases.

La diversité de la longueur des phrases met en évidence non seulement des styles narratifs variés, mais peut également témoigner de différents niveaux de maîtrise linguistique chez les écrivains. Les courtes phrases sont généralement liées à un style plus direct et narratif, tandis que les longues phrases peuvent témoigner d'une tentative de complexité syntaxique, ce qui peut rendre la lecture plus difficile.

IV-2-e- Analyse des mots les plus fréquents :

Il est logique que les mots les plus fréquents soient : « père », « famille », « jour » et « j'ai », puisque la consigne était de raconter une journée particulière en famille. Les mots les plus couramment utilisés dans les textes sont souvent associés à des expériences personnelles et familiales. Les mots « père », « famille », « jour » et « j'ai » sont fréquents, ce qui met en évidence une focalisation sur des événements particuliers et des éléments narratifs récurrents dans les récits. Cela souligne que l'utilisation de ces mots est directement liée à la consigne donnée.

Conclusion :

L'examen des textes analysés reflète une utilisation modérée du vocabulaire, présentant toutefois des variations significatives d'une production à l'autre. Une densité de vocabulaire importante suggère une variété accrue des termes utilisés et peut être considérée comme un signe d'une bonne maîtrise linguistique et d'une capacité d'expression des idées de manière plus nuancée. Par ailleurs, utiliser constamment les mêmes mots ou limiter son lexique pourrait signifier une connaissance limitée de la langue. Les variations de vocabulaire et la complexité des textes suggèrent que certains auteurs possèdent un lexique et une compétence linguistique plus avancés que d'autres, qui pourraient profiter d'un enseignement complémentaire afin d'améliorer leur gamme de mots. Ces conclusions mettent en évidence la nécessité d'utiliser de nouvelles approches spécifiques pour renforcer les compétences linguistiques des étudiants impliquant l'enrichissement de leur vocabulaire et les encourageant à explorer différentes formes syntaxiques et styles narratifs.

En conclusion, les résultats de cette analyse suggèrent plusieurs pistes d'amélioration pour l'enseignement du FLE. En premier lieu, il est recommandé d'introduire les apprenants à une variété de textes, y compris des œuvres littéraires et des textes informatifs. Des ateliers d'écriture focalisés sur la diversification du vocabulaire et l'expérimentation narrative peuvent également être bénéfiques. De plus, encourager la réécriture et la révision des textes peut aider les auteurs à affiner leur expression et à développer une sensibilité à la richesse lexicale.

Pour ce qui est des outils de TALN, ils adoptent des approches modernes et automatisées permettant d'analyser de grands volumes de données textuelles, et fournissant des insights détaillés sur les aspects linguistiques des textes analysés. Dans notre cas, Voyant Tools offre des visualisations interactives et des statistiques sur les fréquences lexicales, les tendances thématiques et la diversité et la densité lexicale, facilitant ainsi l'évaluation lexicale des textes. Nous pouvons dire que cette intelligence artificielle est bel et bien capable d'analyser des données textuelles d'une façon rapide, objective et précise.

En intégrant les outils TALN dans l'analyse des textes, Nous pouvons améliorer non seulement l'efficacité du traitement des données, mais également influencer significativement les méthodologies de l'analyse textuelle.

Références bibliographiques :

- ADAIKA, R. (2018). *Les représentations sociolinguistiques de la langue française chez les étudiants du département de français (Le cas de 3e année de l'Université d'El-Oued)*.
- Audras, I., & Ganascia, J.-G. (2006). *Le TALN au service de la didactique du français langue étrangère écrit*. Paper presented at the 13ème conférence sur le Traitement Automatique des Langues Naturelles.
- Bensalem, D. (2016). Méthodes actuelles de l'enseignement de l'écrit de français langue étrangère: le dispositif du projet. *Multilinguales*(7).
- Berrien, A. Voyant Tools <https://ct3.ortolang.fr/corlidocs/voyanttools/>
- Dekhissi, L., & Valetopoulos, F. (2019). La cohésion dans les textes narratifs: analyse d'un corpus d'apprenants. *Studia Romanica Posnaniensia*, 46(2), 7-24.
- Delay-Goyet, F. (2012). [Quels enjeux pour laproduction d'écrits à l'école? CPC EPS].
- Guaresi, M., & Mayaffre, D. (2021). Intelligence artificielle et discours politique. Quelles plus-values interprétatives? Application aux corpus parlementaire et présidentiel contemporains. *L'intelligence artificielle des textes. Des algorithmes à l'interprétation*, 17, 131-182.
- Guettala, A. (2018). *Étude contrastive des difficultés à l'écrit des étudiants de 3ème année LMD*. Tesis de Doctorado, Universidad de Batna] <http://eprints.univ-batna2.dz/1715>.
- Khelifi, H. (2019). *Didactique du discours: le français langue d'écrit universitaire en Algérie. Étude contrastive entre filières scientifiques et sciences humaines*. Université de Lorraine.
- Miloudi, M. (2020). Les incidences de la politique linguistique algérienne a l'ere de Bouteflika sur les pratiques langagieres des habitants de la commune d'El-Oued. *Synergies Algérie*(28), 221-239.
- Née, É. (2017). *Méthodes et outils informatiques pour l'analyse des discours*: Presses universitaires de Rennes Rennes.
- Raus, R., Tonti, M., Cerquitelli, T., Cagliero, L., Attanasio, G., La Quatra, M., & Greco, S. (2022). *L'analyse du discours et l'intelligence artificielle pour réaliser une écriture inclusive: le projet EMIMIC*. Paper presented at the SHS Web of Conferences.
- Richard, S. (2006). L'analyse de contenu pour la recherche en didactique de la littérature. Le traitement de données quantitatives pour une analyse qualitative: parcours d'une approche mixte. *Recherches qualitatives*, 26(1), 181-207.
- SABAH, G. (2004). Intelligence artificielle, linguistique et cognition. *FUCHS, Catherine (Éd.). La linguistique cognitive. Paris: Éditions de la Maison des sciences de l'homme*, 195-221.
- Steffens, M. (2020). Réemploi lexical et production écrite en FLE: une étude de cas. *Repères. Recherches en didactique du français langue maternelle*(61), 95-115.

ADAI AK Radja Analyse lexicale innovante par le TALN pour améliorer les compétences rédactionnelles d'étudiants en FLE

- Tournier, M. (1980). D'où viennent les fréquences de vocabulaire? La lexicométrie et ses modèles. *Mots. Les langages du politique*, 1(1), 189-209.
- Znaïdi, A. (2020). *Métacognition et apprentissage de l'écriture de textes narratifs: étude de l'autorégulation chez les élèves du 1er cycle de l'enseignement de base*. Université de Lorraine; Université virtuelle de Tunis.

Sitographie : <https://voyant-tools.org/>