L'impact de genre sur la prédiction de la lisibilité du texte en FLE

Lingyun Gao, Rodrigo Wilkens, Thomas François

L'impact de genre sur la prédiction de la lisibilité du texte en FLE

GAO Lingyun[1] Rodrigo Wilkens[1] Thomas François[1]
(1) CENTAL, IL&C, Université catholique de Louvain, Belgique
lingyun.gao.nom@uclouvain.be

RÉSUMÉ

Cet article étudie l'impact du genre discursif sur la prédiction de la lisibilité des textes en français langue étrangère (FLE) à travers l'intégration de méta-informations du genre discursif dans les modèles de prédiction de la lisibilité. En utilisant des architectures neuronales basées sur CamemBERT, nous avons comparé les performances de modèles intégrant l'information de genre à celles d'un modèle de base ne considérant que le texte. Nos résultats révèlent une amélioration modeste de l'exactitude (*accuracy*) globale lors de l'intégration du genre, avec cependant des variations notables selon les genres spécifiques de textes. Cette observation semble confirmer l'importance de prendre en compte les méta-informations textuelles tel que le genre lors de la conception de modèles de lisibilité et de traiter le genre comme une information riche à laquelle le modèle doit accorder une position préférentielle.

ABSTRACT

The impact of genre on the prediction of text readability in FFL (French as a Foreign Language)

This article examines the influence of discourse genre on readability prediction in texts for French as a foreign language (FLE), focusing on the integration of genre-related meta-information into readability models. Utilizing neural architectures based on CamemBERT, we assessed the performance of models that incorporate genre information against a baseline model that considers only the text. Our findings indicate a modest enhancement in overall accuracy with the inclusion of genre, though little significant variations were observed across specific text genres. These results seem to confirm the importance of taking into account textual meta-information such as genre when designing readability models, and of treating genre as rich information to which the model should give a preferential position.

MOTS-CLÉS : lisibilité, français langue étrangère, genre discursif.

KEYWORDS: readability, French as a foreign language, genre.

1 Introduction

Dans le contexte de la didactique des langues étrangères, la sélection de matériaux pédagogiques adaptés aux étudiants constitue un problème fondamental. Dès lors, certains chercheurs ont entrepris de développer des algorithmes capables de déterminer automatiquement le niveau de compétence nécessaire pour comprendre un texte, en vue de faciliter les tâches de préparation pour les formateurs et les examinateurs. Ces modèles, appelés formules de lisibilité, prennent en compte différentes caractéristiques des textes pour prédire automatiquement le degré de lisibilité d'un texte pour une population ciblée (François, 2011). La notion de lisibilité a été définie par Dale & Chall (1949, 19)

comme : la somme totale de tous les éléments d'un document imprimé qui influencent le succès d'un groupe de lecteurs, mesuré au moyen de la compréhension, de la vitesse de lecture et de l'intérêt.

On connaît généralement les formules de lisibilités, dites classiques, telles que celles proposées par Flesch (Flesch, 1948), Flesch-Kincaid (Kincaid *et al.*, 1975) ou Henry (Ters, 1976). Depuis les années 2000, l'évolution des solutions en traitement automatique du langage (TAL) a transformé la manière d'évaluer la lisibilité des textes écrits. Les approches se sont alors appuyées reposaient sur des modèles statistiques plus complexes et des routines de TAL capables de capturer des caractéristiques textuelles plus complexes, tels que la syntaxe (Schwarm & Ostendorf, 2005), les relations de discours (Pitler & Nenkova, 2008), ou des caractéristiques acquisitionnelles (Vajjala & Meurers, 2012). Enfin, suite à la révolution neuronales, les formules de lisibilité se sont majoritairement appuyées sur des réseaux de neurones profond (Azpiazu & Pera, 2019), l'architecture transformer (Yancey *et al.*, 2021) et les plongements de mots (Filighera *et al.*, 2019).

La plupart des travaux récents se concentrent soit sur la nature des informations textuelles à capturer (variables ou plongements de mots) ou sur les algorithmes, mais très peu sur les caractéristiques de la situation de lecture. Les chercheurs comme Hiebert & Pearson (2010) ont constaté que les méta-informations textuelles liés à la situation de communication (ou de lecture ici) telle que genre discursif peuvent jouer un rôle essentiel dans la compréhension des textes. Pourtant, Beier *et al.* (2022) indiquent que cette méta-information est souvent négligée dans les études sur la lisibilité des textes. Ignorer le genre dans l'évaluation de la lisibilité risque de conduire à une compréhension inexactes des structures, une incapacité à prendre en compte les spécificités lexicales (termes de spécialité, éléments brachygraphiques, etc.) et à une normalisation des conventions stylistiques.

L'objectif de cet article est donc explorer l'impact du genre discursif sur la qualité des prédictions de la lisibilité des textes en français langue étrangère (FLE). Nous proposons d'explorer différentes approches pour intégrer la notion de genre dans des modèles de lisibilité neuronaux et d'évaluer son effet sur la performance de ces modèles. Avec cette étude, notre contribution constitue un premier pas en vue de démonter l'importance de prendre en compte le genre dans des modèles neuronaux pour la lisibilité du FLE. Plus généralement, cela ouvre des perspectives quant à la prise en plus des caractéristiques de la situation de lecture dans la lisibilité computationnelle.

Dans la suite de l'article, nous brossons un état de l'art centré sur les études réalisées selon le paradigme de la lisibilité computationnelle en mettant l'accent sur l'importance du "genre" (section 2.1). Nous rappelons également quelques notions fondamentales concernant les genres discursifs (section 2.2). Dans la partie suivante (section 3), nous décrivons la méthodologie de cette étude, qui comprend la construction des corpus et la modélisation. La section des résultats (section 4) met en lumière l'impact du genre sur la performance des modèles. Enfin, nous discutons nos principaux résultats avant de conclure (section 5).

2 État de l'art

2.1 Lisibilité computationnelle et genre

Durant la période classique, les formules de Flesch (1948) et de Dale & Chall (1948) avaient déjà été critiquées pour leur manque de robustesse sur des textes de nature différente que ceux du corpus d'entraînement. Par exemple, Brown (1965) montre que la formule de Dale and Chall surestime

la complexité des textes scientifiques. En réaction, Jacobson (1965) entraîne sa propre formule spécialisée pour les textes scientifiques.

Avec l'avènement de l'apprentissage automatique basé sur l'ingénierie de caractéristiques, les approches en lisibilité sont devenues plus aptes à capturer automatiquement diverses caractéristiques textuelles comme la distribution des mots, (Collins-Thompson & Callan, 2005), les relations syntaxiques (Schwarm & Ostendorf, 2005) ou des relations de discours (Pitler & Nenkova, 2008). Ceci a permis de développer des modèles plus robustes, supposément moins sensibles aux particularités des textes analysés.

Par la suite, le développement de l'apprentissage profond a ouvert de nouvelles perspectives en permettant d'encoder directement les propriétés linguistiques pertinentes sans l'intermédiaire d'une ingénierie de caractéristiques complexe. Les travaux récents réalisés dans cette veine privilégient des architectures neurales avancées. Azpiazu & Pera (2019) ont proposé une architecture de Réseau Neuronal Récurrent multi-attentive pour évaluer la lisibilité selon une approche multilingue. Filighera *et al.* (2019) ont employé les réseaux neuronaux et les plongements lexicaux tels que word2vec, GloVe et ELMo, qui offrent des performances comparables aux approches de pointe basées sur l'ingénierie de caractéristiques, tout en étant plus faciles et moins coûteux à adapter à de nouveaux types de textes. Jian *et al.* (2022) ont introduit un modèle hybride basé sur les réseaux de neurones convolutionnels pour évaluer automatiquement la lisibilité des textes anglais, améliorant ainsi leur efficacité et leur précision. Dans sa thèse, Ma (2022) a employé des modèles Transformer pré-entraînés et a obtenu des améliorations par rapport aux approches basées sur les n-grammes. Enfin, l'exploration de l'apprentissage par renforcement profond, notamment par Mohammadi & Khasteh (2019), a également montré des résultats prometteurs.

Depuis quelque temps, certains chercheurs comme Beier *et al.* (2022) ont cependant mentionné du point de vue de la psychologie cognitive et de l'éducation, la nécessité de prendre en considération, non seulement des éléments linguistiques, mais aussi les informations métalinguistiques liées à la situation de lecture telles que la langue première des lecteurs, le sujet/domaine traité du texte, et le genre textuel. Il n'est toutefois pas le premier à se pencher sur cette question. Certains chercheurs ont déjà rapporté une corrélation entre les performance de modèles de prédiction de la lisibilité et le genre. Ainsi, Nelson *et al.* (2012) ont noté une différence de performance de plusieurs formules sur les genres informatif versus narratif. Dell'Orletta *et al.* (2012, 2014) ont montré, sur un corpus en italien, que la prédiction de la lisibilité était fortement influencée par le genre de textes et que, pour cette raison, une notion de lisibilité orientée par genre est nécessaire. Sheehan *et al.* (2013) se distinguent par leur utilisation explicite du genre, via une méthodologie en deux phases qui intègre une classification initiale des textes par genre (informatif, littéraire, ou mixte) avant l'étape d'évaluation de la lisibilité par un modèle de régression.

Plutôt que d'entraîner des modèles différents par genre, Kate *et al.* (2010) exploitent des modèles n-grams entraînés sur des genres différents au sein d'un modèle unique et observent une amélioration de performance. À l'inverse, il est intéressant de noter que Falkenjack *et al.* (2016) ont cherché à prédire l'appartenance à un genre sur la base de variables textuelles de lisibilité, prouvant ainsi l'existence d'un lien fort entre le genre et la lisibilité. Enfin, Li (2022) a évalué la lisibilité des textes de manuels selon leur genre en appliquant des métriques de la linguistique mathématique et a montré l'influence des genres sur la lisibilité et également sur la compréhensibilité des étudiants non natifs de la langue cible.

Dans le domaine du FLE, qui nous intéresse plus particulièrement, Yancey *et al.* (2021) ont évalué la performance d'un modèle de lisibilité affiné avec CamemBERT en fonction de huit genres différents

couramment utilisés en FLE. Ils ont observé un écart de performance de 20 % pour leur meilleur modèle entre les dialogues et des genres atypiques (tels que les chansons, prospectus, recettes, etc.). Toutefois, le genre n'est pas inclus en tant qu'une variable dans leur modèle, ce qui nous a poussé à prolonger cette expérience, avec une perspective centrée sur le genre.

De ces différents travaux, on peut retenir que l'effet du genre sur les performances des modèles de lisibilité a déjà été démontré. Une approche consiste à entraîner des modèles différents par genre (personnalisation), mais elle s'avère coûteuse. Par conséquent, nous privilégierons la stratégie de Sheehan *et al.* (2013), qui consiste à informer un modèle unique du genre de textes qu'il doit analyser. Nous nous démarquons toutefois en adoptant une architecture neuronale, en l'appliquant, pour la première fois, à la lisibilité du FLE et en explorant la meilleure stratégie pour informer le modèle.

2.2 Genre

La question du genre s'ancre dans la pluralité des disciplines et des points de vue : diverses perspectives coexistent pour aborder cette notion (Ablali, 2010). En linguistique, la définition du genre varie selon les écoles. Ainsi, dans le monde anglophone, trois courants principaux dominent les études de genre : les études de genre rhétoriques (Bazerman *et al.*, 1988), la linguistique systémique fonctionnelle et l'anglais sur objectifs spécifiques (Swales, 1990; Bhatia, 1993). Pour le premier, le genre est décrit comme une action sociale, dans de différents contextes de communication (Miller, 1984). Quant à la linguistique systémique fonctionnelle, le genre y est défini comme un processus social comportant plusieurs étapes (*stage*, en anglais) et axé sur des objectifs spécifiques dans différents contextes sociaux (Martin *et al.*, 2010). Enfin, en anglais sur objectifs spécifiques, le genre est considéré comme « des événements communicatifs reconnaissables caractérisés par des objectifs communicationnels et par différents motifs au niveau de la structure, du style, du contenu, et du public visé » (Swales, 1990, 58).

Dans le contexte francophone, la définition de cette notion reflète également une multitude de points de vue et de disciplines. Plusieurs théoriciens tels que Jean-Michel Adam, Dominique Maingueneau et Patrick Charaudeau y ont apporté leur contribution. Adam (1997) met l'accent sur les structures textuelles, analysant comment les textes se constituent en genres selon leurs caractéristiques formelles et fonctionnelles. Maingueneau (2007) considère le genre comme une notion discursive, se concentrant sur les contextes d'énonciation et les conventions influençant la production et la réception des textes. Enfin, Charaudeau (2011) s'intéresse à la dimension communicative et pragmatique des genres, étudiant comment ils structurent la communication dans divers contextes sociaux. Dans le domaine de la didactique du FLE, Beacco (2013) a approfondi la notion de genre de discours. Selon ces auteurs, les genres de discours représentent une forme métalinguistique de communication spécifique à une communauté de discours donnée, et guident les locuteurs dans leur communication verbale.

Il est à noter également que dans le monde francophone, il existe des hésitations terminologiques et un certain flottement des cadres théoriques de référence sur le genre (Adam, 2011). Les notions de "type de texte" et de "genre de texte" sont parfois considérés comme synonyme. La définition du genre/type varie, selon le point de vue du théoricien ou de la traduction, comme démontré par Adam (2011) dans son livre sur la base des titres des numéros de revues de linguistique et de didactique consacrés à cette question.

Nous considérons que le genre discursif est défini par la situation de production et de réception du texte (Adam, 2011). Pour cette étude, qui se concentre sur la lisibilité des textes en FLE, il est

essentiel de prendre en compte la réalité sur le terrain. Dans l'enseignement du FLE, les "types de texte" sont souvent inclus dans la notion de genre. Selon Adam (2011), les textes sont catégorisés en types et prototypes, selon leurs caractéristiques structurelles et fonctions communicatives. Cette notion se distingue de la notion de genre, mais est souvent confondue avec le genre à cause des confusions terminologiques historique détaillées dans (Adam, 2011). Ainsi sur le terrain, ces deux notions sont souvent mélangées. Ce flottement terminologique se retrouve dans le corpus utilisé dans cette étude (cf. section 3.2), mais, dans cet article, la notion de "genre" est utilisée comme un terme parapluie pour les méta-informations textuelles incluant le genre textuel et le type des textes.

2.3 Modélisation de genre en TAL

Cette étude cherchant également à prédire le genre de textes automatiquement, il convient de souligner que des recherches ont observé l'impact de genre textuel sur la performance des modèles de TAL dans diverses circonstances. Dans sa thèse, Frérot (2005) a montré que, selon le genre de corpus, les performances de différentes stratégies de rattachement prépositionnel pour une tâche d'analyse syntaxique varient. Dans le domaine de la recherche d'informations, Mothe & Tanguy (2005) ont découvert que la difficulté des requêtes peut être calculée et prédite à partir de l'analyse d'un certain nombre de traits linguistiques.

L'objectif principal de la modélisation de genre consiste à regrouper les documents par genre. Ainsi, la classification des genres est appliquée dans des tâches différentes. De l'extraction de terminologie pour des textes spécialisés (Todirascu & Guillaume, 2011; Todriascu *et al.*, 2012), à la recherche d'information sur les genres journalistiques (Petrenz & Webber, 2011), genres du Web (Mehler *et al.*, 2010) et genres littéraires (Ollagnier *et al.*, 2015).

Dans les travaux de l'enseignement du FLE assisté par le TAL, le genre est également pris en considération, comme dans l'étude d'analyse linguistique des productions écrites dans apprenants (Audras & Ganascia, 2005).

3 Méthodologie

L' objectif de cette étude est déterminer dans quelle mesure l'intégration de l'information de genre influe sur la performance des modèles de prédiction de la lisibilité. Nous cherchons à savoir si l'ajout de cette information peut enrichir les représentations des textes et quelle méthode d'intégration s'avère être la plus efficace. Pour ce faire, nous comparons les performances d'une architecture informée du genre de texte à celles d'une architecture de référence qui ne connaît pas le genre. Nous avons limité notre étude à l'architecture transformers, car elle correspond à l'état de l'art en lisibilité et peut apprendre le genre indirectement, offrant ainsi un modèle de référence difficile à atteindre.

3.1 Corpus

3.1.1 Présentation des corpus sources

Cette étude se concentre sur l'effet de l'intégration du genre discursif des textes en tant que méta-information sur la performance des modèles de lisibilité en FLE, visant à prédire les niveaux de

compétence linguistique selon le CECR (Conseil de l'Europe, 2001). Le corpus utilisé est issu du corpus élaboré par François (2011) dans sa thèse de doctorat et comprend des extraits des manuels de FLE couvrant les six niveaux du CECR ainsi qu'une diversité de genres textuels tels que textes, dialogues, lettres/emails, publicités, poèmes/chansons, et recettes. En 2021, Yancey *et al.* (2021) ont constitué FLE-CORP, en suivant la même méthodologie, mais en intégrant des genres supplémentaires et en regroupant les textes des niveaux C1 et C2 pour mieux équilibrer les classes. Une portion de ce corpus a déjà servi à une recherche antérieure par Yancey *et al.* (2021), axée sur l'apprentissage profond et l'emploi de variables cognitives et pédagogiques. La section restante du corpus servira à l'entraînement de notre modèle de prédiction de genre, soulignant la continuité et l'évolution dans l'utilisation des ressources pour affiner les outils de mesure de la lisibilité.

3.1.2 Préparation du corpus

À l'origine, nous avions prévu de réutiliser le corpus FLE-CORP (Yancey *et al.*, 2021), qui comprend 8 genres différents. Cependant, les observations des auteurs, qui ont noté une performance élevée de leurs modèles sur les dialogues (niveaux A1, A2) et faible sur des genres diversifiés comme "varias", nous ont conduit à réévaluer la pertinence des données pour notre étude. Une exploration fine du corpus a révélé que certains textes classés comme "Texte à trous" ou "Varias" devaient être éliminés. Le genre "Texte à trous" correspond à des exercices à trous dans les manuels, tandis que "Varias" regroupe plusieurs genres peu organisés. La catégorie "Texte" mélange des textes narratifs et informatifs. Idéalement, il aurait été préférable de réannoter ces textes, mais cette entreprise dépasse le périmètre de cette étude. Nous avons finalement décidé de garder ces genres par souci de comparaison avec les travaux de Yancey *et al.* (2021).

Pour favoriser l'équilibre des classes, les niveaux C1 et C2 ont été regroupés pour former une classe "C" plus peuplée. Ce corpus comprend ainsi 5 classes au total (A1, A2, B1, B2, C). Une fois les textes du genre "Varias" éliminés, nous avons observé un certain déséquilibre des classes, surtout pour le niveau B2. Pour limiter ce déséquilibre, nous avons récupéré des textes du niveau B2 dans le corpus de François (2011). Le corpus final, nommé Corpus-FLE-GENRES, est ainsi construit à partir des données provenant des deux corpus.

Comme l'information du genre des textes n'est pas toujours disponible sur le terrain réel, nous avons décidé de créer un corpus pour entraîner notre modèle de prédiction des genres, à partir des données qui n'ont pas été intégrées dans le Corpus-FLE-GENRES. Ce corpus, nommé Corpus-Genre, est constitué des textes et de leur genre. L'information de genre semble être partiellement associée à la lisibilité des textes dans nos données (cf. table 1), par exemple, les dialogues apparaissent davantage aux niveaux A1 et A2. Nous avons donc préparé un jeu de données dans lequel nous avons éliminé le texte de chaque donnée, ne conservant que la variable de genre en vue d'entraîner un modèle de lisibilité. Ce corpus sera nommé Genre-Text.

3.1.3 Analyse du corpus

Corpus-FLE-GENRE et Genre-Text contiennent 2 101 données au total, ce qui fait 650 données de moins par rapport au corpus utilisé dans l'étude de Yancey *et al.* (2021). Quant à corpus-GENRE contient, il inclut 1 220 textes. La table 1 présente la distribution des données en ce qui concerne les niveaux CECR et les genres dans Corpus-FLE-GENRE. Ce corpus est par la suite réparti en corpus d'entraînement, de validation et de test ayant une proportion de 70/10/20 pour les expériences.

Niveau/Genre	Dialogue	Informative	Mail	Narrative	Phrase	Texte	Total
A1	105	68	43	27	109	98	450
A2	53	131	31	42	65	128	450
B1	28	86	28	28	26	254	450
B2	17	67	24	28	65	100	301
C	0	104	8	55	39	244	450
Total	203	456	134	180	304	824	2 101

TABLE 1 – Distribution des données par genre et par niveau dans Corpus-FLE-GENRE

Nous pouvons constater que Corpus-FLE-GENRE offre une belle variété au niveau du nombre de textes associés à chaque genre et à chaque niveau. Le niveau B2 a toutefois un nombre de textes inférieur de 149 par rapport aux autres niveaux (qui en comportent 450 chacun). Le niveau A1 comporte un nombre relativement élevé de dialogues et de phrases, ce qui est cohérent avec les compétences de communication basique et la construction de phrases simples à ce stade de l'apprentissage. À mesure que le niveau augmente, on note une augmentation du nombre de textes informatifs et narratifs, ce qui va de pair avec la progression des compétences linguistiques vers un niveau plus élaboré. Enfin, le niveau C met particulièrement l'accent sur les textes informatifs et narratifs, reflétant la maîtrise avancée de la langue nécessaire pour comprendre et produire des textes détaillés et nuancés. Cependant, quand on prête attention aux genres majoritaires dans chaque niveau, nous pouvons observer des similitudes entre chaque niveau à savoir que le genre "Text" est majoritaire dans les niveaux A2, B1, B2 et C. La situation est la même pour le genre informatif. Cette situation est très prononcée dans la classe B1 et surtout B2. Quant à Genre-Text, il ne comprend pas de données issues du niveau B2 et très peu du niveau C, ce qui ne semble néanmoins pas constituer une limite. Par contre, le corpus a un grand nombre de textes du genre texte et Phrase, mais très peu de textes du genre narratif et mail.

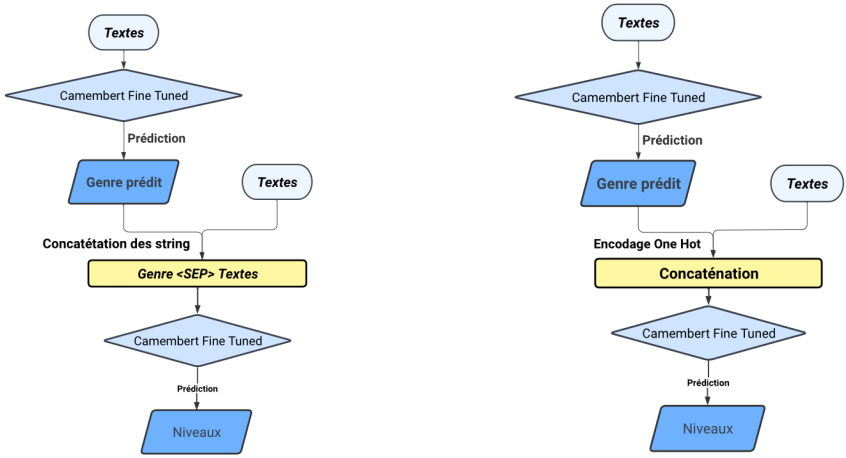
3.1.4 Architectures

Pour notre modèle de référence, nous avons opté pour un CamemBERT affiné (voir figure en annexe 3) disposant uniquement des informations textuelles (sans inclure le genre de texte), car cela correspond à l'approche couramment appliquée en lisibilité.

Ensuite, pour évaluer l'apport de l'information de genre, nous avons entraîné un modèle CamemBERT affiné similaire au modèle de référence, mais avec l'information sur le genre du texte en tant que paramètre d'entrée donné en même temps que le texte. Nous avons exploré deux stratégies pour informer les modèles de lisibilité avec le genre. Premièrement, une chaîne de caractères indiquant le genre (modèle $\text{Concat}_{\text{théorique}}$, voir figure en annexe 5b) est concaténé avec le texte à analyser, et cet input est utilisé pour affiner CamemBERT. Cette stratégie vise à tirer parti de l'information déjà encodée dans CamemBERT concernant les tokens associés au genre. Deuxièmement, nous concaténons le vecteur du texte d'entrée après encodage par CamemBERT avec un vecteur "one-hot" représentant le genre (modèle $\text{One-Hot}_{\text{théorique}}$, voir figure en annexe 4b). Le résultat de cette concaténation est ensuite transmis à la tête de classification de CamemBERT. Ces deux stratégies sont illustrées respectivement dans la Figure 4b en annexe.

La comparaison de ces modèles permet d'évaluer l'apport de l'information de genre à l'identification automatique de la lisibilité d'un texte. Néanmoins, ces modèles constituent un idéal théorique, partant

du principe que la méta-information de genre est connue avec certitude, ce qui n'est pas réalisable avec les limites de la technologie actuelle. Nous avons donc envisagé deux autres modèles afin d'évaluer notre question de recherche dans un environnement plus proche de la réalité des modèles de lisibilité. Ces modèles partagent les mêmes structures que les modèles de concaténation présentés ci-dessus, mais les informations sur le genre proviennent d'un modèle de prédiction automatique du genre. Ces modèles alternatifs sont appelés $\text{Concat}_{\text{réaliste}}$ (voir figure 6b) et $\text{One-Hot}_{\text{réaliste}}$ (voir figure 6c).



(a) Modèle basé sur le nom du genre ($\text{Concat}_{\text{réaliste}}$) (b) Modèle basé sur le vecteur de genre ($\text{One-hot}_{\text{réaliste}}$)

FIGURE 1 – Modèle de concaténation de genre réaliste

Dans cette optique, un modèle automatique est nécessaire pour prédire le genre. Toutefois, un modèle pour prédire les genres en FLE, et plus précisément les genres fournis dans notre contexte d'étude, n'est pas disponible. Nous avons donc développé notre propre modèle d'identification du genre. Comme ce travail ne vise pas à apporter une contribution directe au domaine de la classification automatique du genre, nous avons opté pour un modèle relativement standard, reposant également sur un CamemBERT affiné pour prédire les genres à partir du texte. Il est entraîné sur un échantillon de textes de FLE de notre corpus non utilisés pour l'entraînement des modèles de lisibilité, afin d'éviter d'introduire un biais (pour plus de détails, voir la section 3.1.2).

Tous nos modèles ont été évalués en utilisant une approche de validation croisée à 5 plis pour garantir la robustesse et la généralisabilité des résultats. De plus, l'arrêt anticipé (*early stopping*) basé sur la métrique F-mesure a été employé pour optimiser chaque modèle sans risque de sur-apprentissage. Les performances ont été mesurées à l'aide de métriques standards de classification : l'exactitude , la précision, le rappel, et le F-mesure. Ces mesures sont rapportées sur l'ensemble du corpus, mais également pour chaque niveau de lisibilité et pour chaque genre. Des analyses statistiques ont été menées pour déterminer si les améliorations observées étaient significatives, permettant ainsi d'évaluer de manière rigoureuse l'impact de l'intégration de l'information de genre sur la performance des modèles.

Un modèle de prédiction de la lisibilité basé exclusivement sur les genres sera également développé, mais nous ne discutons pas sa performance, car nous sommes intéressés par son évaluation extrinsèque,

au sein de la tâche de lisibilité.

4 Résultats

Notre étude sur l'intégration de l'information de genre dans la prédiction de la lisibilité des textes en français langue étrangère (FLE) a évalué cinq modèles : CamemBERT, $\text{Concat}_{\text{théorique}}$, $\text{One-Hot}_{\text{théorique}}$, $\text{One-Hot}_{\text{réaliste}}$, et $\text{Concat}_{\text{réaliste}}$. La table 2 présente les scores d'exactitude obtenus par ces modèles sur l'ensemble des données (Global) et sur chacun des genres considérés, ainsi leurs écart-types et leur intervalle de confiance.

Genre	**CamemBERT**	**$\text{Concat}_{\text{théorique}}$**	**$\text{One-Hot}_{\text{théorique}}$**	**$\text{One-Hot}_{\text{réaliste}}$**	**$\text{Concat}_{\text{réaliste}}$**	**Écart-type**
Texte	0,53	0,51	0,53	0,51	0,48	0,04
Informative	0,64	0,65	0,65	0,59	0,56	0,03
Mail	0,43	0,54	0,67	0,52	0,43	0,09
Phrase	0,42	0,46	0,53	0,44	0,53	0,04
Dialogue	0,69	0,74	0,69	0,66	0,68	0,02
Narrative	0,56	0,61	0,57	0,52	0,53	0,06
Global	0,55	0,57	0,59	0,54	0,55	0,02

TABLE 2 – Scores d'exactitude moyens par genre pour chaque modèle avec écart-type et intervalle de confiance à 95 %.

L'intégration de l'information de genre montre des gains d'exactitude de 0,02 ($\text{Concat}_{\text{théorique}}$) et 0,05 ($\text{One-Hot}_{\text{théorique}}$) par rapport au modèle de référence (CamemBERT), bien que ces gains ne soient pas statistiquement significatifs. Certains genres bénéficient davantage de l'intégration de l'information de genre. Par exemple, pour le genre "Dialogue", $\text{Concat}_{\text{théorique}}$ améliore l'exactitude de 0,05 par rapport à CamemBERT. Le genre "Mail" voit l'exactitude du modèle $\text{One-Hot}_{\text{théorique}}$ atteindre 0,67 contre 0,43 pour CamemBERT. L'utilisation de modèles réalistes, qui prédisent le genre des textes, diminue logiquement les performances. Ainsi, l'exactitude du modèle $\text{One-Hot}_{\text{réaliste}}$ diminue de 0,05 par rapport à $\text{One-Hot}_{\text{théorique}}$, et $\text{Concat}_{\text{réaliste}}$ perd 0,02.

Les intervalles de confiance et les écarts-types fournissent des informations cruciales sur la stabilité et la précision des performances des modèles. L'analyse des intervalles de confiance (IC) (voir Figure 5)pour l'accuracy des modèles sur différents genres de textes révèle que les modèles '$\text{Concat}_{\text{réaliste}}$' et 'Baseline' présentent généralement des IC plus étroits, indiquant une plus grande stabilité dans leurs prédictions. Par exemple, '$\text{Concat}_{\text{réaliste}}$' affiche l'IC le plus étroit pour les genres 'texte' (0.039668) et 'mail' (0.104243), ce qui suggère une performance cohérente. En revanche, les modèles '$\text{One-hot}_{\text{réaliste}}$' et '$\text{Concat}_{\text{théorique}}$' montrent des IC plus larges dans certains genres, tels que 'sentence' et 'dialogue', indiquant une variabilité plus importante dans leurs performances de prédiction. Les genres 'narrative' et 'mail' montrent une plus grande variabilité dans les performances de certains modèles, suggérant que ces types de textes posent plus de défis pour des prédictions stables. Par exemple, le modèle 'Baseline' a un IC très large pour le genre 'narrative' (0.277162), tandis que '$\text{Concat}_{\text{théorique}}$' affiche un IC large pour 'mail' (0.324186). Le tableau 5 montre que la stabilité des performances des modèles varie non seulement entre les modèles, mais aussi en fonction des genres de texte.

T-test sur les intervalles de confiance par genres de chaque modèles et les écarts-types sur l'exactitude fournissent des informations cruciales sur la stabilité et la précision des performances des modèles(voir Figure 6 & 7). Les modèles $\text{Concat}_{\text{théorique}}$ et $\text{One-hot}_{\text{théorique}}$ semblent offrir une stabilité comparable

avec une variabilité de performance relativement faible. Les modèles hybrides montrent plus de variabilité, suggérant une performance moins stable. La plupart des comparaisons montrent des P-Values non significatives, indiquant que les améliorations de performance par rapport au modèle Baseline ne sont pas statistiquement significatives pour la plupart des types de texte, à quelques exceptions près comme $\text{Concat}_{\text{réaliste}}$ pour certains genres.

Quant à l'analyse des erreurs, nous allons la traiter selon deux perspectives : l'évaluation des matrices de confusion au niveau des genres et la corrélation entre les caractéristiques linguistiques et les erreurs de prédiction. Les matrices de confusion montrent que les modèles proposés améliorent leurs performances par rapport au modèle de référence (voir Figures 5 et 6). Pour le genre "Texte", les modèles $\text{Concat}_{\text{théorique}}$ et $\text{Concat}_{\text{réaliste}}$ réduisent légèrement les confusions, notamment aux niveaux A1 et C. Cependant, des confusions persistent entre les niveaux intermédiaires tels que A2, B1 et B2, souvent dues à la similarité des caractéristiques entre ces niveaux et genres, indiquant un besoin d'amélioration dans la différenciation des caractéristiques.

Avec l'outil FABRA (Wilkens *et al.*, 2022), nous avons analysé la corrélation entre certaines caractéristiques linguistiques (longueur des textes et longueur moyenne des mots) et les erreurs de prédiction (voir Figure 7). Les corrélations observées sont faibles, indiquant que ces caractéristiques n'ont pas une influence significative sur les erreurs. Cela suggère que d'autres facteurs, comme la complexité syntaxique, la fréquence des mots, et les relations sémantiques, pourraient être plus pertinents pour expliquer les erreurs.

Pour améliorer les modèles, il serait bénéfique d'explorer ces autres variables linguistiques plus en profondeur. Une analyse détaillée de la complexité syntaxique, de la fréquence des mots et d'autres aspects linguistiques pourrait fournir des diagnostics précieux pour affiner les modèles et améliorer la précision des prédictions de lisibilité.

5 Conclusion

Notre étude approfondit la compréhension de l'impact du genre sur la prédiction de la lisibilité en FLE, confirmant l'importance cruciale de ces méta-informations. Bien que l'intégration du genre n'ait pas conduit à des améliorations significatives globalement, nos analyses par genre révèlent des nuances spécifiques, confirmant l'impact soulevé dans certaines études en lisibilité mentionnées dans la section 2.1. Même avec ce signal faible, nous concluons que les informations sur le genre aident à identifier le niveau de lisibilité d'un texte. Toutefois, cette tâche est loin d'être achevée. En effet, notre travail montre que plusieurs éléments doivent être mis en place afin d'obtenir un véritable modèle de lisibilité intégrant le genre. Pour les futures recherches, nous estimons que la collecte de données variées est fondamentale, y compris la nécessité d'un corpus doté d'informations bien calibrées sur le genre. D'autre part, il est important de développer des architectures qui prennent efficacement en compte le genre. Notre étude préliminaire montre que le genre doit être traité comme une information riche et le modèle doit lui accorder une position préférentielle, afin d'affiner et d'améliorer la performance des modèles de lisibilité.

Références

ABLALI D. (2010). Linguistique des genres. exploration sur corpus. *Linguistique & Littérature : Cluny*, **40**, 251.

ADAM J.-M. (1997). Genres, textes, discours : pour une reconception linguistique du concept de genre. *Revue belge de philologie et d'histoire*, **75**(3), 665–681.

ADAM J.-M. (2011). La linguistique textuelle. introduction à l'analyse textuelle des discours. *Semen*, **32**, 182–185. DOI : https ://doi.org/10.4000/semen.9411.

AUDRAS I. & GANASCIA J.-G. (2005). Des outils informatiques au service du passage à l'écrit d'apprenants.

AZPIAZU I. & PERA M. (2019). Multiattentive recurrent neural network architecture for multilingual readability assessment. *Transactions of the Association for Computational Linguistics*, **7**, 421–436. DOI : 10.1162/tacl_a_00278.

BAZERMAN C. *et al.* (1988). *Shaping written knowledge : The genre and activity of the experimental article in science*, volume 356. University of Wisconsin Press Madison.

BEACCO J.-C. (2013). L'approche par genres discursifs dans l'enseignement du français langue étrangère et langue de scolarisation. *Pratiques. Linguistique, littérature, didactique*, (157-158), 189–200. Number : 157-158 Publisher : Association CRESEF, DOI : 10.4000/pratiques.3838.

BEIER S., BERLOW S., BOUCAUD E., BYLINSKII Z., CAI T., COHN J., CROWLEY K., DAY S. L., DINGLER T., DOBRES J., HEALEY J., JAIN R., JORDAN M., KERR B., LI Q., MILLER D. B., NOBLES S., PAPOUTSAKI A., QIAN J., REZVANIAN T., RODRIGO S., SAWYER B. D., SHEPPARD S. M., STEIN B., TREITMAN R., VANEK J., WALLACE S. & WOLFE B. (2022). Readability Research : An Interdisciplinary Approach. *Foundations and Trends® in Human–Computer Interaction*, **16**(4), 214–324. DOI : 10.1561/1100000089.

BHATIA V. (1993). *Analysing Genre : Language Use in Professional Settings*. Applied linguistics and language study. Longman.

BROWN W. (1965). Science textbook selection and the Dale-Chall formula. *School Science and Mathematics*, **65**(2), 164–167.

CHARAUDEAU P. (2011). Chapitre 1. l'information comme acte de communication. *Medias-Recherches*, **2**, 21–28.

COLLINS-THOMPSON K. & CALLAN J. (2005). Predicting Reading difficulty with Statistical Language Models. *JASIST*, **56**, 1448–1462. DOI : 10.1002/asi.20243.

CONSEIL DE L'EUROPE (2001). *Cadre Européen Commun de Référence pour les Langues : Apprendre, enseigner, évaluer*.

DALE E. & CHALL J. (1948). A formula for predicting readability. *Educational research bulletin*, **27**(1), 11–28.

DALE E. & CHALL J. (1949). The concept of readability. *Elementary English*, **26**(1), 19–26.

DELL'ORLETTA F., VENTURI G. & MONTEMAGNI S. (2012). Genre-oriented Readability Assessment : a Case Study. p. 91–98.

DELL'ORLETTA F., WIELING M., VENTURI G., CIMINO A. & MONTEMAGNI S. (2014). Assessing the Readability of Sentences : Which Corpora and Features ? In *Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications*, p. 163–173, Baltimore, Maryland : Association for Computational Linguistics. DOI : 10.3115/v1/W14-1820.

FALKENJACK J., SANTINI M. & JONSSON A. (2016). An Exploratory Study on Genre Classification using Readability Features.

FILIGHERA A., STEUER T. & RENSING C. (2019). *Automatic Text Difficulty Estimation Using Embeddings and Neural Networks*, p. 335–348. DOI : 10.1007/978-3-030-29736-7_25.

FLESCH R. (1948). A new readability yardstick. *Journal of Applied Psychology*, **32**(3), 221–233. Place : US Publisher : American Psychological Association, DOI : 10.1037/h0057532.

FRANÇOIS T. (2011). *Les apports du traitement automatique du langage à la lisibilité du français langue étrangère*. Thèse de doctorat, Université Catholique de Louvain.

FRÉROT C. (2005). *Construction et évaluation en corpus variés de lexiques syntaxiques pour la résolution des ambiguïtés de rattachement prépositionnel*. Thèse de doctorat, Université de soutenance.

HIEBERT E. H. & PEARSON P. D. (2010). An examination of current text difficulty indices with early reading texts. reading research report #10-01.

JACOBSON M. (1965). Reading difficulty of physics and chemistry textbooks. *Educational and Psychological Measurement*, **25**(2), 449–457.

JIAN L., XIANG H. & LE G. (2022). English text readability measurement based on convolutional neural network : A hybrid network model. *Computational Intelligence and Neuroscience*, **2022**, 1–9. DOI : 10.1155/2022/6984586.

KATE R., LUO X., PATWARDHAN S., FRANZ M., FLORIAN R., MOONEY R., ROUKOS S. & WELTY C. (2010). Learning to Predict Readability using Diverse Linguistic Features. p. 546–554.

KINCAID J. P., FISHBURNE J., ROBERT P. R., RICHARD L. C. & BRAD S. (1975). *Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel :*. Rapport interne, Defense Technical Information Center, Fort Belvoir, VA. DOI : 10.21236/ADA006655.

LI W. (2022). Text genres, readability and readers' comprehensibility. *European Journal of Computer Science and Information Technology*, **10**, 52–62. DOI : 10.37745/ejcsit.2013/vol10n45262.

MA C. (2022). *Readability Assessment with Pre-trained Transformer Models*. Thèse de doctorat.

MAINGUENEAU D. (2007). Genres de discours et modes de généricité. *Le français aujourd'hui*, (4), 29–35.

MARTIN J., CHRISTIE F. & ROTHERY J. (2010). Social processes in education : a reply to sawyer and watson. *Metaphor*, (4), 51–52.

MEHLER A., SHAROFF S. & SANTINI M. (2010). *Genres on the Web : Computational Models and Empirical Studies*, volume 42. DOI : 10.1007/978-90-481-9178-9.

MILLER C. R. (1984). Genre as social action. *Quarterly journal of speech*, **70**(2), 151–167.

MOHAMMADI H. & KHASTEH S. H. (2019). Text as environment : A deep reinforcement learning text readability assessment model.

MOTHE J. & TANGUY L. (2005). Linguistic features to predict query difficulty.

NELSON J., PERFETTI C., LIBEN D. & LIBEN M. (2012). Measures of Text Difficulty :. *Council of Chief State School Officers, Washington, DC*.

OLLAGNIER A., FOURNIER S. & BELLOT P. (2015). Analyse en dépendance et classification de requêtes en langue naturelle, application à la recommandation de livres [dependency parsing and classification of natural language queries : application to book recommendation]. *Traitement Automatique des Langues*, **56**(3), 23–47.

PETRENZ P. & WEBBER B. (2011). Stable classification of text genres. *Computational Linguistics*, **37**, 385–393. DOI : 10.1162/COLI_a_00052.

PITLER E. & NENKOVA A. (2008). Revisiting readability : a unified framework for predicting text quality. In *Proceedings of the Conference on Empirical Methods in Natural Language Proces-*

sing - EMNLP '08, p. 186, Honolulu, Hawaii : Association for Computational Linguistics. DOI : 10.3115/1613715.1613742.

SCHWARM S. & OSTENDORF M. (2005). Reading level assessment using support vector machines and statistical language models. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, p. 523–530.

SHEEHAN K. M., FLOR M. & NAPOLITANO D. (2013). A two-stage approach for generating unbiased estimates of text complexity. In *Proceedings of the Workshop on Natural Language Processing for Improving Textual Accessibility*, p. 49–58.

SWALES J. (1990). *Genre Analysis : English in Academic and Research Settings*. Cambridge Applied Linguistics. Cambridge University Press.

TERS F. (1976). Henry (Georges). — Comment mesurer la lisibilité. *Revue française de pédagogie*, **36**(1), 71–74.

TODIRASCU A. & GUILLAUME B. (2011). Classyn : classer les documents selon le genre textuel.

TODRIASCU A., PADÓ S., KISSELEW M., KRISCH J. & HEID U. (2012). French and german corpora for audience-based text type classification.

VAJJALA S. & MEURERS D. (2012). On improving the accuracy of readability classification using insights from second language acquisition. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, p. 163–173.

WILKENS R., ALFTER D., WANG X., PINTARD A., TACK A., YANCEY K. P. & FRANÇOIS T. (2022). FABRA : French aggregator-based readability assessment toolkit. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, p. 1217–1233, Marseille, France : European Language Resources Association.

YANCEY K., PINTARD A. & FRANCOIS T. (2021). Investigating readability of french as a foreign language with deep learning and cognitive and pedagogical features. *Lingue e linguaggio*, **20**(2), 229–258.

Annexes

Annexe 1 : Distribution de longueur de texte par genre et par niveau

Ce tableau présente le nombre de mots maximum, minimum et moyen par genre dans le corpus FLE-GENRE.

Genre	Dialogue	Informative	Mail	Narrative	Phrase	Texte
MaxNbmot	522	1682	646	1427	340	2485
MinNbmot	12	15	23	3	18	15
MoyenNbmot	146,67	228,22	138,52	225,21	73,15	251,48

TABLE 3 – Nombre de mots par genre dans Corpus-FLE-GENRE

Ce tableau montre la distribution de la longueur des textes par niveau de compétence. Elle permet de visualiser comment la complexité textuelle varie en fonction des niveaux de compétence des apprenants.

Niveau	A1	A2	B1	B2	C
MaxNbmot	346	644	719	2485	1682
MinNbmot	12	15	24	3	15
MoyenNbmot	94,01	134,58	191,60	254,48	348,26

TABLE 4 – Nombre de mots par niveau dans Corpus-FLE-GENRE

Annexe 2 : Intervalles de Confiance

Ce tableau présente les intervalles de confiance de précision pour chaque genre et chaque modèle. Ces données sont essentielles pour évaluer la variabilité et la stabilité des performances des modèles.

Modèle	**Texte**	**Phrase**	**Narratif**	**Informatif**	**Dialogue**	**Mail**
Baseline	0.0693	0.0909	0.2772	0.0436	0.2749	0.1286
Concat_théorique	0.0827	0.1515	0.1922	0.1184	0.2749	0.3242
One_hot_théorique	0.0834	0.1773	0.1676	0.1303	0.1632	0.1749
One_hot_réaliste	0.1071	0.0611	0.0850	0.1639	0.2290	0.2145
Concat_réaliste	0.0397	0.0645	0.0862	0.0702	0.1524	0.1042

TABLE 5 – Intervalles de Confiance de Précision par Genre et par Modèle

Annexe 3 : Comparaisons Statistiques

Ces tableaux présentent les résultats des tests T et les valeurs P pour les comparaisons des performances des modèles par genre. Ils aident à déterminer si les différences observées entre les modèles sont statistiquement significatives.

Type de Texte	Modèle de Base	Modèle de Comparaison	Statistique T	Valeur P
text	Baseline	Concat	-0.2733	0.7930
sentence	Baseline	Concat	-1.0075	0.3434
narrative	Baseline	Concat	-0.4031	0.6993
informative	Baseline	Concat	0.3593	0.7287
dialogue	Baseline	Concat	0.0798	0.9384
mail	Baseline	Concat	0.8824	0.4033
text	Baseline	One_hot	-0.5426	0.6081
sentence	Baseline	One_hot	-1.1712	0.2758
narrative	Baseline	One_hot	-0.5081	0.6278
informative	Baseline	One_hot	0.2621	0.8002
dialogue	Baseline	One_hot	0.1984	0.8487
mail	Baseline	One_hot	-0.3696	0.7214
text	Baseline	One_hot_réaliste	1.5300	0.1848
sentence	Baseline	One_hot_réaliste	-0.6657	0.5271
narrative	Baseline	One_hot_réaliste	-0.3196	0.7615
informative	Baseline	One_hot_réaliste	0.2390	0.8174
dialogue	Baseline	One_hot_réaliste	0.2809	0.7876
mail	Baseline	One_hot_réaliste	-1.8058	0.1088
text	Baseline	Concat_réaliste	-6.1665	0.0016
sentence	Baseline	Concat_réaliste	-2.7598	0.0250
narrative	Baseline	Concat_réaliste	-0.8314	0.4445
informative	Baseline	Concat_réaliste	-2.8235	0.0239
dialogue	Baseline	Concat_réaliste	-0.8194	0.4421
mail	Baseline	Concat_réaliste	-2.7423	0.0295

TABLE 6 – Statistiques T et Valeurs P pour les Comparaisons de Modèles par Genre-1

Type de Texte	Modèle de Base	Modèle de Comparaison	Statistique T	Valeur P
text	One_hot	One_hot_réaliste	1.6450	0.1399
sentence	One_hot	One_hot_réaliste	0.7413	0.4848
narrative	One_hot	One_hot_réaliste	0.3397	0.7441
informative	One_hot	One_hot_réaliste	-0.0228	0.9823
dialogue	One_hot	One_hot_réaliste	0.1095	0.9155
mail	One_hot	One_hot_réaliste	-1.5003	0.1719
text	One_hot	Concat_réaliste	-2.4254	0.0688
sentence	One_hot	Concat_réaliste	-1.3219	0.2249
narrative	One_hot	Concat_réaliste	-0.4000	0.7026
informative	One_hot	Concat_réaliste	-2.6811	0.0335
dialogue	One_hot	Concat_réaliste	-1.4040	0.1980
mail	One_hot	Concat_réaliste	-2.4572	0.0429
text	One_hot_réaliste	Concat_réaliste	-4.1242	0.0134
sentence	One_hot_réaliste	Concat_réaliste	-2.6081	0.0334
narrative	One_hot_réaliste	Concat_réaliste	-1.0042	0.3458
informative	One_hot_réaliste	Concat_réaliste	-2.6820	0.0332
dialogue	One_hot_réaliste	Concat_réaliste	-1.5581	0.1578
mail	One_hot_réaliste	Concat_réaliste	-0.7142	0.4977

TABLE 7 – Statistiques T et Valeurs P pour les Comparaisons de Modèles par Genre-2

Annexe 4 : Distribution de longueur de texte par genre et par niveau

Cette figure montre la distribution de la longueur des textes par genre et par niveau de compétence.

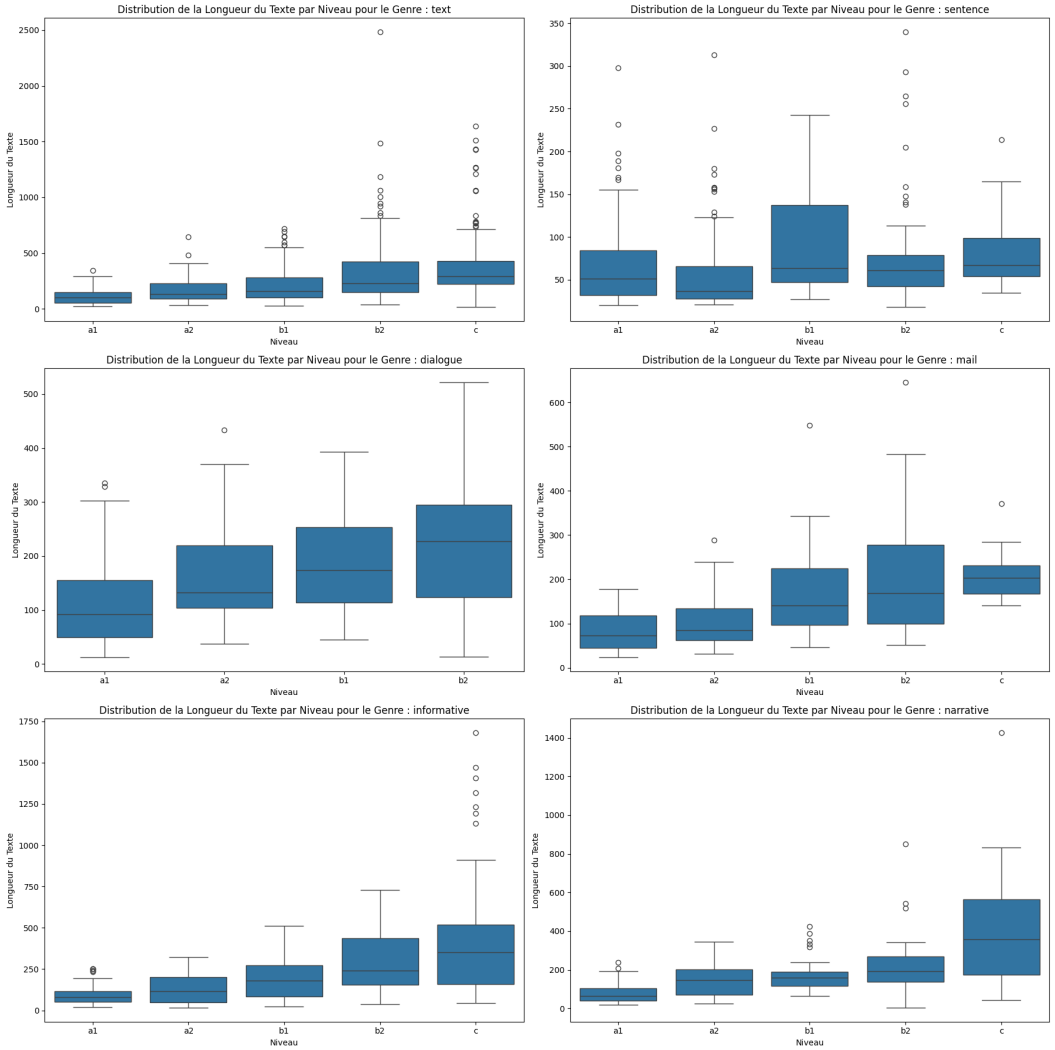


FIGURE 2 – Distribution de longueur de texte par genre et niveau

Annexe 5 : Modèles et Architectures

Ces figure illustre le modèle de base (Baseline) utilisé dans l'étude pour prédire la lisibilité des textes qui ne prend en compte que les informations textuelles sans intégrer les méta-informations de genre.

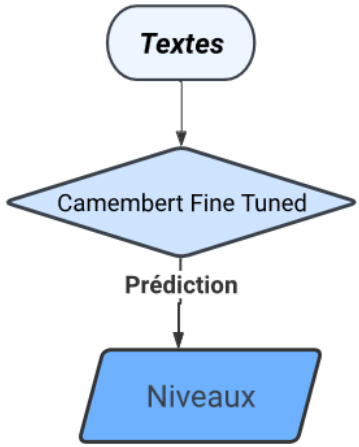
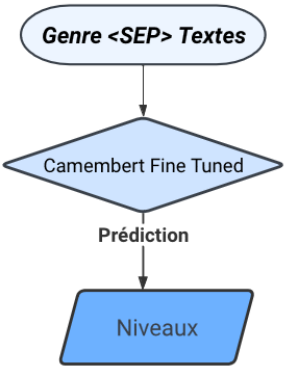
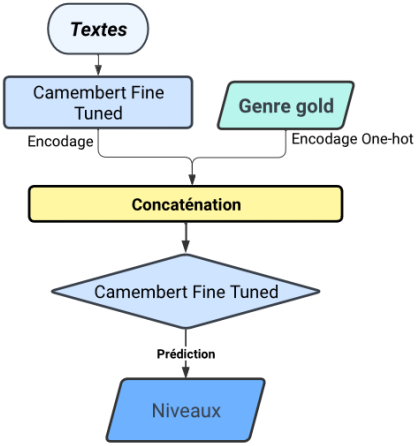


FIGURE 3 – Modèle Baseline

Cette figure présente deux variantes de modèles théoriques utilisant la concaténation de l'information de genre avec les textes. Le modèle (a) utilise le nom du genre, tandis que le modèle (b) utilise un vecteur "one-hot" pour représenter le genre.



(a) Modèle basé sur le nom du genre ($\text{Concat}_{\text{théorique}}$)

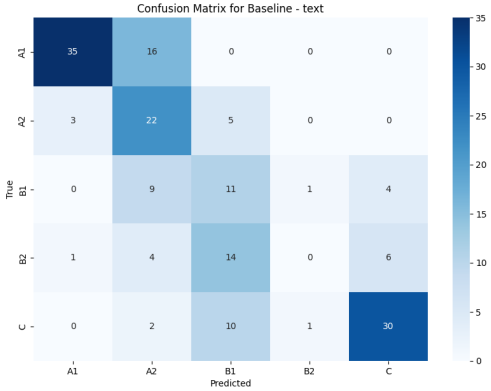


(b) Modèle basé sur le vecteur de genre ($\text{One-hot}_{\text{théorique}}$)

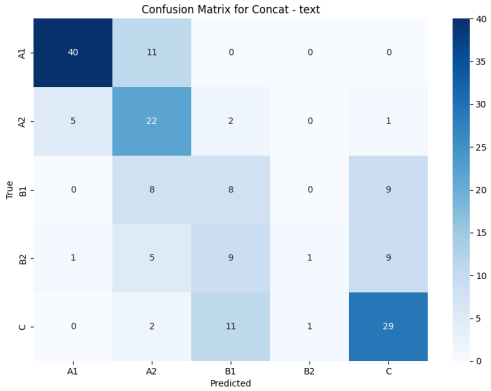
FIGURE 4 – Modèle de concaténation de genre Théorique

Annexe 6 : Matrices de Confusion

Ces matrices de confusion montrent les performances des différents modèles dans la prédiction de la lisibilité pour le genre 'Texte'. Elles illustrent les confusions entre les niveaux de compétence, ce qui aide à identifier les domaines où les modèles peuvent être améliorés.

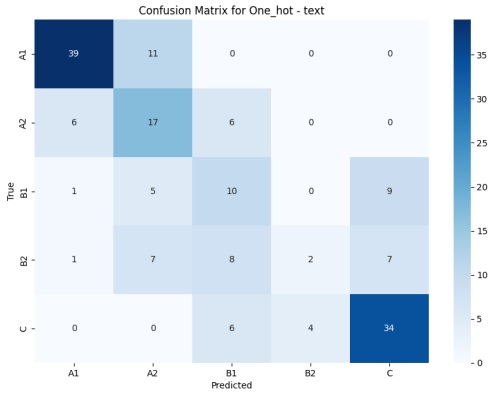


(a) Matrice de confusion - Texte (Baseline)

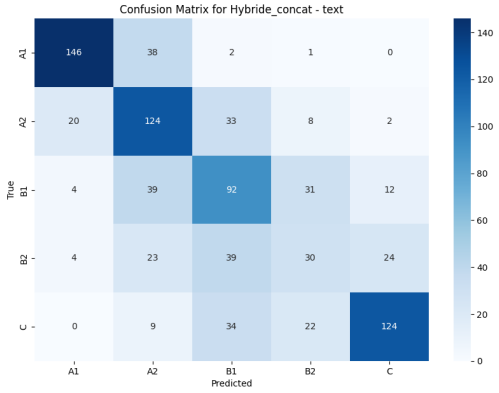


(b) Matrice de confusion - Texte ($\text{Concat}_{\text{théorique}}$)

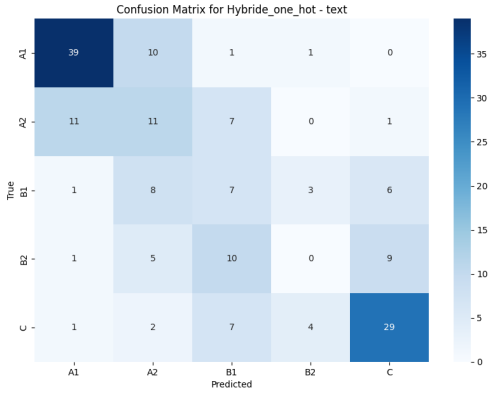
FIGURE 5 – Matrices de confusion de la prédiction de la lisibilité pour le genre 'Texte'



(a) Matrice de confusion - Texte (One-hot$_{\text{théorique}}$)



(b) Matrice de confusion - Texte (Concat$_{\text{réaliste}}$)



(c) Matrice de confusion - Texte (One-hot$_{\text{réaliste}}$)

FIGURE 6 – Matrices de confusion de la prédiction de la lisibilité pour le genre ’Texte’ bis

Annexe 7 : Corrélations et Analyses d'Erreur

Cette figure montre les corrélations entre certaines caractéristiques linguistiques (comme la longueur des textes et la longueur moyenne des mots) et les erreurs de prédiction. Les corrélations observées sont faibles, suggérant que d'autres facteurs linguistiques pourraient être plus pertinents pour expliquer les erreurs.

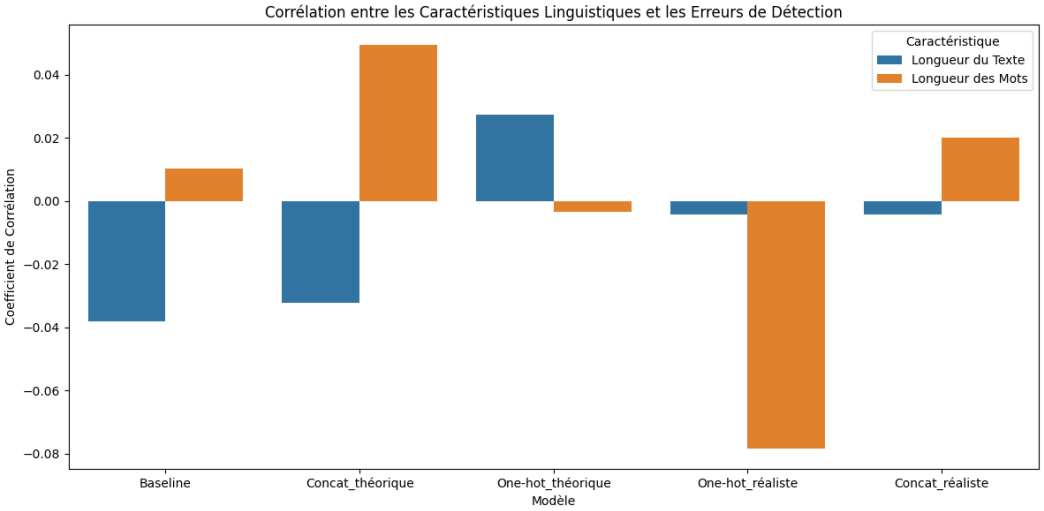


FIGURE 7 – Visualisation Corrélation Caractéristiques linguistiques et Erreurs