Modéliser la facilité d'écoute en FLE : vaut-il mieux lire la transcription ou écouter le signal vocal ?

Minami Ozawa, Rodrigo Wilkens, Kaori Sugiyama, Thomas François

Modéliser la facilité d'écoute en FLE : vaut-il mieux lire la transcription ou écouter le signal vocal ?

Minami Ozawa[1, 2] Rodrigo Wilkens[1] Kaori Sugiyama[2] Thomas François[1]
(1) Université catholique de Louvain, IL&C, CENTAL, Louvain-la-Neuve, Belgique
(2) Université Seinan Gakuin, Fukuoka, Japon
minami.ozawa@uclouvain.be, rodrigo.wilkens@uclouvain.be,
sugiyama@seinan-gakuin.jp, thomas.francois@uclouvain.be

RÉSUMÉ

Le principal objectif de cette étude est de proposer un modèle capable de prédire automatiquement le niveau de facilité d'écoute de documents audios en français. Les données d'entrainement sont constituées d'enregistrements audios accompagnés de leurs transcriptions et sont issues de manuels de FLE dont le niveau est évalué sur l'échelle du Cadre européen commun de référence (CECR). Nous comparons trois approches différentes : machines à vecteurs de support (SVM) combinant des variables de lisibilité et de fluidité, wav2vec et CamemBERT. Pour identifier le meilleur modèle, nous évaluons l'impact des caractéristiques linguistiques et prosodiques ainsi que du style de parole (dialogue ou monologue) sur les performances. Nos expériences montrent que les variables de fluidité améliorent la précision du modèle et que cette précision est différente par style de parole. Enfin, les performances de tous les modèles varient selon les niveaux du CECR.

ABSTRACT

Modelling listenability for FFL : is it better to read the transcript or listen to the speech signal ?

Our main goal is to design a model able to automatically predict the level of listenability of audio documents in French as a foreign language (FFL). The training data consists of audio recordings accompanied by their transcriptions and are extracted from FFL textbooks whose level is assigned accordingly to the Common European Framework of Reference (CEFR) scale. We compare three different approaches : support vector machines combining readability and fluency variables, wav2vec, and CamemBERT. To identify the best model, we evaluate the impact of linguistic and fluency features as well as of speech style (dialogue or monologue) on performance. Our experiments show that fluency variables improve model accuracy, and that this accuracy differs by speech style. Finally, the performance of all models varies according to CEFR scales.

MOTS-CLÉS : facilité d'écoute, lisibilité, FLE, wav2vec.

KEYWORDS: listenability, readability, FFL, wav2vec.

1 Introduction

La compréhension orale est une activité langagière dynamique activement impliquée dans l'interprétation. Le développement de cette compétence est directement lié à l'acquisition de la compétence langagière en général : si la compréhension orale ne progresse pas, cela peut nuire à la bonne communication dans des situations authentiques (Kumai, 1992). Pour soutenir son développement

chez les apprenants de langue étrangère, il est courant de faire écouter divers documents audios dans le cadre d'exercices de compréhension à l'audition. L'efficacité de documents audio-visuels authentiques est bien établie (Yoshimi, 2019), mais leur utilisation en contexte de classe reste difficile, car il est ardu de juger de l'adéquation de ce type de documents audios au niveau de compétence linguistique des apprenants. Une solution pour faire face à cette difficulté pratique pourrait consister à se reposer sur des modèles d'intelligence artificielle capables d'évaluer automatiquement le niveau de difficulté de documents oraux, sur le modèle de ce qui est fait en évaluation automatisé de la lisibilité des documents écrits. Si les travaux en lisibilité se sont imposés comme une des thématiques du TAL (François, 2011; Vajjala, 2021), ceux sur la facilité d'écoute (ou *listenability*, en anglais) sont nettement moins nombreux, en particulier en français.

Les objectifs de cette étude sont doubles. Tout d'abord, le principal objectif est de développer un modèle capable de prédire automatiquement le niveau de facilité d'écoute de documents audios en français. À notre connaissance, un tel modèle n'existant actuellement pas en français, il s'agit déjà d'une contribution significative. Dans cette optique, deux questions émergent. D'une part, quelle est la meilleure façon d'encoder les caractéristiques stylistiques des documents oraux ? Pour y répondre, nous comparerons l'apport de variables linguistiques définies par des experts et combinées à l'aide de machines à vecteur de support (SVM) à l'utilisation de l'architecture transformer et CamemBERT (Martin *et al.*, 2020) qui représente le contenu des documents audios à l'aide de plongements de mots. D'autre part, quel est l'apport de variables prosodiques ? En effet, la grande majorité des travaux en facilité d'écoute se limitent à prédire la difficulté sur la base de la transcription et de variables calculées sur le texte, sans prendre en compte les caractéristiques spécifiques à l'oralité. Nous comparerons donc ces deux approches au sein de nos modèles, encodant les caractéristiques prosodiques soit à l'aide de variables incluses dans les SVM, soit en tirant profit de l'architecture wav2vec (Baevski *et al.*, 2020). Le second objectif de cet article vise à évaluer l'impact du style de parole sur les performances des modèles précités. Nous disposons d'une variété de documents oraux que nous regroupons simplement selon deux styles de parole : le dialogue et le monologue. Nous examinons si les performances des modèles diffèrent en fonction de la présence ou de l'absence d'un interlocuteur dans le discours.

Cet article propose d'abord une vue d'ensemble de la notion centrale de la facilité d'écoute dans la section 2, suivie de la description de la méthodologie de recherche dans la section 3. Il présente et compare ensuite les résultats des modèles SVM, de wav2vec et de CamemBERT dans la section 4, avant de conclure.

2 Le domaine de la facilité d'écoute

Plusieurs définitions de la facilité d'écoute ont été proposées jusqu'à présent (Harwood, 1950; Harwood & Cartier, 1952; Cartier, 1952; Rubin, 2012). En tenant compte de ces définitions, dans cet article, le terme *facilité d'écoute* est défini comme la facilité ou la difficulté qu'un auditeur donné – avec son expérience spécifique – éprouve à comprendre un discours oral dans une situation de communication particulière, laquelle est fonction de l'effet des caractéristiques stylistiques de ce discours (ex. lexicale, syntaxique, prosodique, etc.) sur les processus cognitifs de l'auditeur.

En tant que domaine de recherche, la facilité d'écoute s'est principalement développée en langue anglaise. Dans les années 1940, ce champ a été initié par l'application des modèles issus des études de lisibilité à l'analyse des transcriptions de la langue orale (Flesch, 1943; Chall & Dial, 1948). Par

la suite, les travaux continuent à évaluer la facilité d'écoute sur la base de transcriptions et avec des formules de lisibilité développées pour l'écrit (Cartier, 1955; O'Keefe, 1971).

Toutefois, peu à peu, des formules spécifiques à l'oral apparaissent. Des recherches (Rogers, 1962; Fang, 1967; Kiyokawa, 1990) conçoivent des modèles directement sur des données orales, mais ils ne prennent toujours en compte que des variables liées au contenu (longueur des phrases, des mots, proportion de mots polylexicaux, etc.) et les observent sur des transcriptions. Au niveau de la facilité d'écoute dans le contexte spécifique des apprenants, il y a une recherche qui utilisent la régression multiple pour combiner une série de variables de contenu afin de prédire la difficulté subjective de compréhension orale des phrases anglaises, évaluée par 90 apprenants japonais de l'anglais (Ueda *et al.*, 2013). Le coefficient de corrélation multiple R de l'équation atteint seulement 0,54.

C'est avec les travaux de Kotani *et al.* (2014) que, en complément des variables de contenu, des caractéristiques phonologiques sont enfin considérées, à savoir le débit de parole, le taux d'élision [1], de réduction [2], de contraction [3], de liaison [4] et de déduction [5] dans une phrase. La présence de coefficients de régression négatifs pour le taux de contraction et de déduction suggère que des variations phonologiques peuvent accroître la facilité d'écoute pour les apprenants. Parallèlement, il y a une recherche qui examine dans quelle mesure le débit de parole, la complexité linguistique et l'explicitation (le degré d'expression explicite des idées) du texte influent sur la compréhension orale en L2 (Révész & Brunfaut, 2013). Les auteurs utilisent des analyses de Rasch et de régression pour estimer la difficulté de 18 tâches et sa relation avec les caractéristiques du texte. Les résultats démontrent que des indices de complexité lexicale et discursive expliquent une part significative de la difficulté des tâches de compréhension orale.

Enfin, plus récemment, les arbres de décision (Kotani & Yoshimi, 2017) ou les machines à vecteurs de support (SVM) (Yoshimi & Kotani, 2020) ont été utilisés. Néanmoins, toutes les recherches susmentionnées ont été menées en anglais. En ce qui concerne la langue française, les investigations sont très restreintes aux travaux de Ruggia (2019, 2020, 2021). De plus, un outil est aussi disponible pour l'évaluation automatique des textes oraux, DeepFLE [6], mais il reste encore limité à l'analyse des transcriptions et n'intègre pas de variables prosodiques. Enfin, l'utilisation de représentations latentes (ex. plongements de mots) et de wav2vec plutôt que des variables ne semble pas avoir encore été envisagé dans ce domaine.

3 Méthodologie

Dans cet article, nous cherchons à évaluer l'apport de variables linguistiques et prosodiques à un modèle de facilité d'écoute pour le français, mais posons également la question de savoir si les représentations latentes du contenu ne sont pas préférables à des variables définies par des experts, comme cela a déjà été établi en lisibilité (Martinc *et al.*, 2021; Yancey *et al.*, 2021). Cette section commence par décrire les données utilisées pour entraîner nos modèles, puis décrit les différentes variables « expert » considérées, avant de se clôturer avec une présentation des différents modèles évalués.

1. L'élision est l'élimination des phonèmes.
2. La réduction est l'affaiblissement du son en transformant une voyelle en schwa.
3. La contraction est la combinaison de deux mots.
4. La liaison consiste à relier le son final d'un mot au son initial du mot suivant.
5. La déduction est l'élimination des sons entre les mots.
6. http ://deeptext.unice.fr/FLE/.

3.1 Données

Les données sont constituées d'enregistrements audios accompagnés de leurs transcriptions que nous avons extraits de 25 manuels de FLE. Chaque enregistrement s'est vu attribué un niveau de compétence sur l'échelle du Cadre européen commun de référence (CECR), à savoir A1, A2, B1, B2 ou C (C1 et C2 sont regroupés pour notre étude). Le niveau attribué à un enregistrement donné est tout simplement celui du manuel dont il a été tiré.

Plusieurs prétraitements ont été effectués. Tout d'abord, le contenu d'une seule piste audio est en principe traité comme une seule donnée. Toutefois, si une piste contient, par exemple, plusieurs dialogues dans différents contextes, l'audio (et la transcription) est alors divisé manuellement. Ensuite, nous avons éliminé les éléments de soutien pédagogique (tels que les exercices de grammaire et les listes de mots), ainsi que les données provenant des premières et dernières unités de chaque manuel (qui sont trop proches du niveau précédent ou suivant). Dans notre corpus, les enregistrements comprennent des conversations quotidiennes, des annonces au public, des interviews médias et des monologues. Toutes ces données ont été classées manuellement en deux grands styles de parole : les dialogues et les monologues. Enfin, les données dont le contenu informatif est extrêmement élevé ou faible dans une certaine catégorie de niveau/style de parole peuvent réduire l'homogénéité lors de l'examen des caractéristiques de ce niveau/style de parole. Pour cette raison, les cinq données présentant des valeurs aberrantes extrêmes en termes de syllabes par minute sans les pauses ou/et de mots par minute sans les pauses ont été exclues. Ces cinq données se composaient d'un document B2, d'un document C1 et de trois documents C2. Les documents B1, C1 et un des C2 ne comportaient qu'une ou deux phrases. La petite taille des données par rapport au niveau/style pourrait être considérée comme une indication d'hétérogénéité. Les deux autres documents C2 étaient des données avec beaucoup de bruit et des phrases en langue étrangère non mentionnés dans les transcriptions. Ils ont été traitées comme des pauses pour des raisons de commodité et, par conséquent, les valeurs aberrantes ont montré que ces données étaient hétérogènes et ne convenaient pas comme données pour l'analyse. Le nombre total final de paires de données (audio et transcription) est de 1 323 documents (Table 1). Le nombre de mots ainsi que la durée d'enregistrement pour les niveaux C1 et C2 sont largement supérieurs à ceux des autres niveaux. Cependant, les variables que l'on utilise sont normalisées lorsque c'est nécessaire afin d'éviter les biais.

3.2 Variables

Dans cette étude, la difficulté de compréhension (comprenant nos cinq niveaux issus du CECR comme modalités) a été utilisée comme variable dépendante tandis que 68 variables capturant les caractéristiques prosodiques et le contenu textuel sont utilisées comme variables indépendantes. Les variables indépendantes sont divisées en deux catégories principales : les variables de fluidité (7 variables) et les variables de lisibilité (61 variables). Dans cet article, les variables liées à la prosodie se limitent à celles relatives au débit de parole, à la durée de parole et aux pauses.

La seule variable prosodique qui a été identifiée dans la littérature comme étant corrélée à la facilité d'écoute est le nombre de mots par minute (Harwood, 1955; Kotani *et al.*, 2014; Kotani & Yoshimi, 2017). Cependant, étant donné le nombre réduit de recherches sur cette question, nous avons décidé d'enrichir la liste des variables prises en compte en nous référant au concept de fluidité, qui est souvent utilisé dans les recherches sur la production orale. La fluidité se réfère à la capacité d'utiliser la langue en temps réel, de mettre l'accent sur le sens et d'utiliser les systèmes lexicaux (Skehan

	Nb de données (%)		Nb de mots (%)		Durée d'enregistrement (minutes) (%)	
	Dialogue	Monologue	Dialogue	Monologue	Dialogue	Monologue
A1	290 (100 %)		18 950 (100 %)		160 (100 %)	
	205 (71 %)	85 (29 %)	14 576 (77 %)	4 374 (23 %)	120 (75 %)	40 (25 %)
A2	309 (100 %)		33 755 (100 %)		230 (100 %)	
	180 (58 %)	129 (42 %)	25 665 (76 %)	8 090 (24 %)	180 (78 %)	50 (22 %)
B1	240 (100 %)		32 818 (100 %)		190 (100 %)	
	146 (61 %)	94 (39 %)	25 289 (77 %)	7 529 (23 %)	150 (79 %)	40 (21 %)
B2	241 (100 %)		51 920 (100 %)		290 (100 %)	
	131 (54 %)	110 (46 %)	30 825 (59 %)	21 095 (41 %)	170 (59 %)	120 (41 %)
C1/C2	243 (100 %)		127 396 (100 %)		710 (100 %)	
	121 (50 %)	122 (50 %)	95 934 (75 %)	31 462 (25 %)	520 (73 %)	190 (27 %)
Total	1 323 (100 %)		264 839 (100 %)		1 580 (100 %)	
	783 (59 %)	540 (41 %)	192 289 (73 %)	72 550 (27 %)	1 140 (72 %)	440 (28 %)

TABLE 1 – Description du jeu de données, qui précise le nombre de documents, de mots par document et la durée en minute, de façon globale et par style de parole.

& Foster, 1999). Étant donné que la compréhension orale, tout comme la production orale, est une aptitude qui requiert de telles compétences, il est raisonnable d'appliquer ce concept. Nous nous référons à la méthode de calcul utilisée dans les recherches précédentes (Cucchiarini *et al.*, 2002; Ginther *et al.*, 2010; Préfontaine, 2010; Peltonen, 2017; Segalowitz *et al.*, 2017) pour définir les variables utilisées dans cet article, à savoir : (1) le nombre de syllabes par minute, (2) le nombre de syllabes par minute sans les pauses, (3) le nombre de mots par minute, (4) le nombre de mots par minute sans les pauses, (5) la durée totale, (6) la durée moyenne des pauses, et (7) le nombre de pauses par minute.

Pour les variables de lisibilité, 61 variables de FABRA (Wilkens *et al.*, 2022) sont utilisées. FABRA est un outil en ligne permettant de calculer un large éventail de variables prédictives de lisibilité pour le français. Le domaine de la facilité d'écoute s'est développé par l'application des formules de lisibilité à l'analyse des transcriptions, et l'existence signalée de variables valides pour la facilité d'écoute ne peut être ignorée. Cet article se focalise sur les variables de lisibilité qui ont été considérées comme pertinentes dans les études précédentes sur la facilité d'écoute. Ainsi, bien que FABRA puisse fournir des informations sur 509 variables liées à la lisibilité, certaines d'entre elles sont moins pertinentes dans le contexte de la facilité d'écoute et ne font pas partie de notre analyse. Les variables liées au nombre de ponctuations et de guillemets dans FABRA, par exemple, ne sont pas directement pertinentes pour cette analyse. Bien que certaines variables relatives aux erreurs dans FABRA existent, il n'est pas non plus nécessaire de les considérer, étant donné que l'étude ne se concentre pas sur la production des apprenants. En outre, cet article se concentre sur les variables liées aux mots de contenu, nécessaires pour comprendre le sens du discours, plutôt qu'aux mots fonctionnels. De plus, certaines variables de FABRA sont relativement redondantes, mesurant le même phénomène à l'aide de variantes. Dans ce cas, seule l'une d'entre elles est prise en considération (par exemple, la proportion de mots A1 et A2 selon la ressource FLELex (François *et al.*, 2014) [7] est fort redondante. Dès lors, seule la proportion de mots A1 dans le texte selon la ressource a été prise en considération).

7. FLELex est un lexique pour le FLE qui donne les fréquences normalisées des lemmes à chaque niveau du CECRL (http://cental.uclouvain.be/flelex/).

Au terme de cette première sélection de variables, 61 variables de lisibilité ont été retenues parmi les variables de FABRA. Cette procédure de sélection est conforme à la première étape de sélection des variables, qui consiste à construire un meilleur ensemble de caractéristiques « ad hoc » à partir de la connaissance du domaine de la facilité d'écoute, selon Guyon & Elisseeff (2003). En résumé, les variables comprennent (1) 2 variables basées sur la longueur (ex. nombre de syllabes par mot), (2) 17 variables lexicales (ex. fréquence moyenne des lemmes dans FLELex pour les noms, les noms propres, les verbes, les adjectifs ou les adverbes, (3) 41 variables syntaxiques basées sur le formalisme UD et l'analyseur Stanza (Qi *et al.*, 2020) (ex. proportion d'indicatifs présent, mots identifiés comme adjectifs, etc.) et (4) le score de la formule de lisibilité de Kandel et Moles (Kandel & Moles, 1958) [8] [9].

3.3 Modèles

Nous comparons trois approches différentes pour entraîner notre modèle de facilité d'écoute. La première s'ancre dans les travaux de lisibilité computationnelle, combinant les 68 variables décrites à la section 3.2 à l'aide de SVM. Elle vise à reproduire une méthodologie fiable, qui a fait ses preuves en lisibilité et est relativement facile à entraîner. La seconde approche consiste simplement à affiner une architecture CamemBERT (Martin *et al.*, 2020).

Plus original, l'emploi du modèle wav2vec 2.0 (Baevski *et al.*, 2020) en facilité d'écoute vise à capturer au mieux la richesse du signal vocal. Le modèle traite la forme d'onde brute du signal vocal avec un réseau neuronal convolutionnel multicouches pour obtenir des représentations audio latentes. Ces représentations sont ensuite introduites dans un quantificateur et un transformateur. Le quantificateur choisit une unité vocale pour la représentation audio latente à partir d'un inventaire d'unités apprises. Environ la moitié des représentations audio sont masquées avant d'être introduites dans le transformateur. Le transformateur ajoute des informations provenant de l'ensemble de la séquence audio. Enfin, la sortie du transformateur est utilisée pour résoudre une tâche contrastive. Cette tâche exige que le modèle identifie les unités de parole quantifiées correctes pour les positions masquées. Pour la version française, nous utilisons le modèle *facebook/wav2vec2-large-xlsr-53-french* [10], qui est un modèle wav2vec 2.0 entraîné sur des données audio non annotées de 12 langues provenant du benchmark Common Voice. [11]

Nous avons également ajouté une tête de classification à wav2vec. Cette tête est composée d'un pooling de moyennes 1D sur le dernier état caché de wav2vec. Ensuite, les informations regroupées passent par un MLP composé de deux couches linéaires denses (la première avec une activation tanh et la seconde avec une activation softmax) chacune précédée d'une couche de Dropout avec des probabilités de 0,5 et 0,1.

Afin de tester nos deux hypothèses principales : l'effet des variables prosodiques et du style de parole, plusieurs modèles ont été entraînés sur la base des 3 architectures ci-dessus. En ce qui concerne l'effet des variables prosodiques, trois jeux de variables différents ont été utilisés pour les SVMs : un jeu ne comprenant que des variables de fluidité (SVM-F), un second jeu ne comprenant que des variables de lisibilité (SVM-L) et un dernier jeu incluant à la fois les variables de fluidité et celles de

8. Voir https://cental.uclouvain.be/fabra/docs.html pour le détail des variables.
9. Les détails de toutes les variables utilisées dans cette étude sont donnés en annexe.
10. https://huggingface.co/facebook/wav2vec2-large-xlsr-53-french
11. Pour plus d'informations, voir https://ai.meta.com/blog/wav2vec-20-learning-the-structure-of-speech-from-raw-audio/.

lisibilité (SVM-FL). Au niveau des représentations latentes, nous avons CamemBERT et wav2vec. En comparant CamemBERT avec SVM-L et wav2vec avec SVM-F, il est possible d'examiner s'il y a une contribution des modèles neuronaux et des représentations distribuées. En outre, en comparant SVM-FL, SVM-L et SVM-F, il est possible d'examiner si les variables de fluidité ont un impact sur la construction d'un modèle de facilité d'écoute. Par ailleurs, ces 5 modèles ont été déclinés sur trois ensembles de données différents : l'ensemble des données, les données correspondant au style de parole monologal et celles associées aux dialogues.

Enfin, en ce qui concerne la recherche des hyperparamètres, nous les avons explorés à l'aide d'une recherche par grille. Les performances de tous nos modèles ont été évaluées au moyen d'une procédure de validation croisée à 10 blocs et à l'aide des mesures d'évaluation de précision, de rappel et de F1. Dans la validation croisée à 10 blocs, 80 % des données servent à la phase d'entraînement, 10 % pour le développement et 10 % pour le test. Les moyennes des 10 plis obtenus pour l'exactitude et la F1 score sont rapportées.

4 Analyse des variables et des modèles

Dans cette section, nous présentons d'une part une analyse corrélationnelle des 68 variables à la section 4.1, puis les résultats des différents modèles (SVM-F, SVM-L, SVM-FL, CamemBERT et wav2vec) à la section 4.2.

4.1 Analyse corrélationnelle des variables

Tout d'abord, pour l'ensemble des données et chaque style de parole (dialogue, monologue), le coefficient de corrélation de Spearman a été utilisé pour examiner la corrélation entre chacune des 68 variables et le niveau. Seules les variables dont la valeur absolue du coefficient de corrélation est supérieure à 0,2 ($p < 0{,}05$) sont conservées. Lorsque plusieurs variables sont incluses dans une même catégorie, pour chaque catégorie, seules celles qui ont la plus forte corrélation avec le niveau sont conservées tandis que les autres sont exclues, ce qui permet de minimiser la colinéarité entre les variables lors de la construction du modèle. En suivant la procédure ci-dessus pour chaque ensemble de données (données entières, dialogue, monologue), les variables finales à incorporer dans nos modèles ont été déterminées.

En comparant les trois ensembles de variables, les différences dans les coefficients de corrélation sont particulièrement notables, surtout pour les variables liées à la fluidité (nombre de syllabes par minute, durée moyenne des pauses et nombre de pauses par minute). Toutes les variables sélectionnées sur l'ensemble des données se retrouvent dans les jeux de variables spécifiques au dialogue et au monologue, l'inverse n'étant pas vrai.

En confrontant les résultats obtenus sur les dialogues et sur les monologues, on constate que 17 variables sont communes aux deux ensembles. Cela représente 56 % des variables finales dans le dialogue et 80 % des variables finales dans le monologue, ce qui suggère que le niveau tend à être déterminé par davantage de variables dans le dialogue que dans le monologue. En particulier, dans le dialogue, les variables liées aux temps des verbes sont plus fréquemment mentionnées comme variables fortement corrélées avec le niveau que dans le monologue. Ainsi, malgré le fait que les variables finalement traitées ont été considérées dans les mêmes conditions dans les deux styles

de parole, il y a des différences dans le nombre de variables et les variables ne sont pas tout à fait identiques. En d'autres termes, cela montre déjà l'utilité de construire le modèle séparément pour chaque style de parole.

4.2 Comparaison des modèles

Il est intéressant d'analyser les performances des 5 modèles sur l'ensemble des données (dialogue et monologue) dont les résultats sont repris à la Table 2a, mais aussi leurs performances sur les deux styles de parole de façon isolée (Tables 2b et 2c).

En comparant SVM-L et CamemBERT pour examiner l'apport des variables linguistiques, CamemBERT apparait à première vue comme le meilleur modèle, aussi bien sur l'ensemble des données que sur les monologues. Pour chaque modèle, la F1 de CamemBERT se situe entre 0,5 et 0,7. Un tel résultat n'est pas surprenant, car les corpus issus de manuels de FLE se caractérisent généralement par une forte hétérogénéité au niveau des annotations de la difficulté entre manuels (François, 2014). Cependant, un problème important de CamemBERT est que son écart-type est particulièrement élevé, ce qui implique une instabilité dans le modèle et nous conduit à regarder avec prudence ces résultats.

Ensuite, SVM-F et wav2vec sont comparés afin d'examiner l'apport des variables prosodiques. Les résultats de wav2vec montrent que, pour chaque style de parole, la F1 est d'environ 0,2, ce qui est inférieur à celles de SVM-F. Cela semble indiquer que l'information audio identifiée par wav2vec ne constitue pas un signal assez fort pour identifier les niveaux de difficulté. Les performances de SVM-F, assez mauvaises également, confirment que les variables de fluidité envisagées dans cette étude ont une contribution limitée à la prédiction de la facilité d'écoute. Néanmoins, ces deux résultats de F1 montrent de manière générale des performances assez élevé sur le niveau A1 (0,44 à 0,53 pour wav2vec, 0,55 à 0,65 pour SVM-F) et, dans une moindre mesure, sur le niveau A2 (0,31 à 0,38 pour wav2vec, 0,28 à 0,31 pour SVM-F), pour tous les styles de parole. Or, ces niveaux sont notamment caractérisés par un débit plus lent : le nombre de syllabes par minute en A1 (181 syll./min.) est bien moindre que au niveau A2 (226 syll./min.)) et aux niveaux B1 (246 syll./min.) et B2 (255 syll./min.). Cela peut indiquer une tendance du modèle à utiliser des informations relatives au débit de parole pour évaluer la difficulté des discours. Toutefois, sur les niveaux supérieurs, le contenu devient plus critique et il semblerait que wav2vec ne capture pas suffisamment la teneur de celui-ci.

En ce qui concerne SVM-FL, il apparaît que, pour chaque style de parole et de façon générale, la F1 obtenue (0,49 à 0,54) est supérieure à celle de SVM-L. Cela signifie que les variables de fluidité apportent tout de même des informations utiles au modèle et différentes de celles des variables de lisibilité. Cependant, ces F1 sont inférieures à la F1 de CamemBERT, pour l'ensemble de données et pour les monologues. En d'autres termes, dans ces styles de parole, CamemBERT, qui est un modèle neuronal sans informations prosodiques, est plus performant.

Enfin, en examinant les résultats spécifiquement sur les dialogues et les monologues, ces deux styles de parole ne produisent pas systématiquement les mêmes résultats pour chaque modèle. Néanmoins, on peut observer une certaine régularité dans les résultats : la F1 du modèle de tous les SVMs et de wav2vec montre que ceux-ci sont presque toujours meilleurs sur les dialogues. Au contraire, CamemBERT se comporte toujours mieux sur les monologues. Cependant, comme mentionné ci-dessus, en raison des problèmes de variance dans les résultats de CamemBERT, les performances de ce modèle doivent être considérées avec précaution. Par conséquent, en se basant sur la tendance claire des résultats des SVMs et de wav2vec, on peut considérer que prédire la facilité d'écoute sur

	CamemBERT	wav2vec	SVM-FL	SVM-L	SVM-F
Exactitude (écart-type)	0,64 (0,36)	0,33 (0,03)	0,51 (0,05)	0,49 (0,02)	0,35 (0,03)
F1 (écart-type)	0,59 (0,42)	0,24 (0,04)	0,49 (0,05)	0,47 (0,02)	0,28 (0,04)
F1-A1 (écart-type)	0,50 (0,50)	0,53 (0,12)	0,72 (0,07)	0,69 (0,05)	0,59 (0,07)
F1-A2 (écart-type)	0,63 (0,40)	0,34 (0,05)	0,52 (0,09)	0,50 (0,04)	0,28 (0,09)
F1-B1 (écart-type)	0,59 (0,44)	0,11 (0,10)	0,30 (0,12)	0,24 (0,09)	0,06 (0,06)
F1-B2 (écart-type)	0,58 (0,44)	0,17 (0,09)	0,32 (0,09)	0,36 (0,11)	0,04 (0,05)
F1-C (écart-type)	0,63 (0,43)	0,05 (0,06)	0,60 (0,04)	0,58 (0,04)	0,44 (0,05)

(a) Résultat pour la tâche de prédiction du niveau de facilité d'écoute sur l'ensemble des données

	CamemBERT	wav2vec	SVM-FL	SVM-L	SVM-F
Exactitude (écart-type)	0,58 (0,36)	0,35 (0,04)	0,56 (0,04)	0,52 (0,04)	0,37 (0,04)
F1 (écart-type)	0,50 (0,43)	0,23 (0,07)	0,54 (0,04)	0,50 (0,06)	0,33 (0,03)
F1-A1 (écart-type)	0,52 (0,45)	0,53 (0,05)	0,74 (0,08)	0,70 (0,08)	0,65 (0,10)
F1-A2 (écart-type)	0,55 (0,43)	0,31 (0,08)	0,54 (0,11)	0,52 (0,10)	0,29 (0,07)
F1-B1 (écart-type)	0,48 (0,45)	0,11 (0,12)	0,39 (0,09)	0,25 (0,10)	0,27 (0,07)
F1-B2 (écart-type)	0,42 (0,48)	0,16 (0,14)	0,43 (0,17)	0,41 (0,15)	0,18 (0,12)
F1-C (écart-type)	0,53 (0,48)	0,06 (0,13)	0,60 (0,07)	0,65 (0,09)	0,26 (0,15)

(b) Résultat pour la tâche de prédiction du niveau de facilité d'écoute sur les données de style dialogue

	CamemBERT	wav2vec	SVM-FL	SVM-L	SVM-F
Exactitude (écart-type)	0,72 (0,36)	0,31 (0,07)	0,52 (0,10)	0,49 (0,08)	0,37 (0,04)
F1 (écart-type)	0,67 (0,43)	0,22 (0,07)	0,51 (0,09)	0,47 (0,08)	0,31 (0,03)
F1-A1 (écart-type)	0,63 (0,48)	0,44 (0,18)	0,68 (0,13)	0,56 (0,16)	0,55 (0,18)
F1-A2 (écart-type)	0,66 (0,45)	0,38 (0,07)	0,55 (0,12)	0,51 (0,11)	0,31 (0,13)
F1-B1 (écart-type)	0,65 (0,46)	0,00 (0,00)	0,36 (0,15)	0,27 (0,15)	0,09 (0,12)
F1-B2 (écart-type)	0,67 (0,44)	0,09 (0,10)	0,36 (0,22)	0,40 (0,19)	0,09 (0,11)
F1-C (écart-type)	0,75 (0,36)	0,18 (0,16)	0,59 (0,09)	0,59 (0,12)	0,50 (0,06)

(c) Résultat pour la tâche de prédiction du niveau de facilité d'écoute sur les données de style monologue

TABLE 2 – Résultat pour la tâche de prédiction du niveau de facilité d'écoute

les dialogues est plus aisée que sur les monologues. En ce qui concerne la nature des prédicteurs, les performances de CamemBERT et même des SVMs montrent que la transcription et les variables qui en sont dérivées contribuent plus aux performances que les informations prosodiques. Néanmoins, nous avons identifié que les caractéristiques du signal vocal ajoutent bien de l'information complémentaire pertinente pour les modèles.

5 Conclusion et perspectives

Cette étude a comparé divers modèles en vue de prédire automatiquement le niveau de facilité d'écoute de documents audios en français. Tout d'abord, aussi bien le SVM-L utilisant uniquement les variables de lisibilité que le CamemBERT affiné obtiennent une F1 relativement élevée. Sur l'ensemble de données et sur les monologues, la F1 de CamemBERT est légèrement supérieure à celle du SVM-L. Cependant, dans cette étude, CamemBERT souffre de problèmes d'échantillonnage

causant de grandes disparités d'une session d'entraînement à l'autre, ce que révèle l'écart-type élevé des métriques d'évaluation estimées pour CamemBERT. Ensuite, l'analyse des résultats de SVM-F, utilisant uniquement les variables de fluidité, et de wav2vec a révélé que prédire à partir des seules caractéristiques prosodiques ne permettait pas d'atteindre des résultats très élevé, surtout pour les niveaux plus élevés. Enfin, nous avons constaté que les variables de fluidité apportaient tout de même des informations utiles au modèle lorsqu'elles sont combinées avec les variables de lisibilité (SVM-FL). Il a également été constaté que les différents styles de parole produisent des résultats différents en termes de performance du modèle, mais que la plupart des modèles se comportent mieux sur les dialogues.

Enfin, il convient de noter que la F1 par niveau est très différente d'un niveau à l'autre, pour tous les modèles analysés dans cette étude. Dans l'ensemble, on constate que la F1 est faible pour les niveaux B1 et B2, qui sont des niveaux intermédiaires du CECR. L'examen minutieux des variables les plus aptes à discriminer les niveaux B est l'une des tâches futures. Une autre perspective serait de combiner wav2vec, qui a atteint une F1 d'environ 0,3 à 0,5 aux niveaux A1 et A2, et CamemBERT affiné, qui a eu une bonne performance de prédiction en général. L'analyse de ce modèle combiné, reposant sur des plongements, constitue également une perspective intéressante qui permettrait de déterminer dans quelle mesure ces deux modèles sont complémentaires en matière d'information. Dans le même ordre d'idée, il serait possible d'explorer des architectures hybrides combinant des plongements de mots avec des variables de lisibilité.

Références

BAEVSKI A., ZHOU Y., MOHAMED A. & AULI M. (2020). wav2vec 2.0 : A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*, **33**, 12449–12460.

CARTIER F. A. (1952). The social context of listenability research. *Journal of Communication*, **2 (1)**, 44–47. DOI : 10.1111/j.1460-2466.1952.tb00177.x.

CARTIER F. A. (1955). Ii. listenability and "human interest". *Communications Monographs*, **22 (1)**, 53–57. DOI : 10.1080/03637755509375134.

CHALL J. S. & DIAL H. E. (1948). Predicting listener understanding and interest in newscasts. *Educational Research Bulletin*, **27 (6)**, 141–168.

CUCCHIARINI C., STRIK H. & BOVES L. (2002). Quantitative assessment of second language learners' fluency : Comparisons between read and spontaneous speech. *The Journal of the Acoustical Society of America*, **111(6)**, 2862–2873.

DIAS G., Éd. (2015). *Actes de TALN 2015 (Traitement automatique des langues naturelles)*, Caen. ATALA, HULTECH.

FANG I. E. (1967). The "easy listening formula". *Journal of Broadcasting & Electronic Media*, **11 (1)**, 63–68.

FLESCH R. (1943). Marks of readable style, a study in adult education. *Teachers College Contributions to Education*.

FRANÇOIS T. (2014). An analysis of a french as a foreign language corpus for readability assessment. In *Proceedings of the 3rd workshop on NLP for Computer-assisted Language Learning, NEALT Proceedings Series Vol. 22, Linköping Electronic Conference Proceedings 107*, p. 13–32.

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

FRANÇOIS T., GALA N., WATRIN P. & FAIRON C. (2014). Flelex : a graded lexical resource for french foreign learners. *International conference on Language Resources and Evaluation*.

GINTHER A., DIMOVA S. & YANG R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral english proficiency with implications for automated scoring. *Language Testing*, **27**, 379–399.

GUYON I. & ELISSEEFF A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, **3**, 1157–1182.

HARWOOD K. A. (1950). A concept of listenability. *Western Speech*, **14 (2)**, 10–12. DOI : 10.1080/10570315009373409.

HARWOOD K. A. (1955). Iii. listenability and rate of presentation. *Communications Monographs*, **22(1)**, 57–59. DOI : 10.1080/03637755509375135.

HARWOOD K. A. & CARTIER F. (1952). On definition of listenability. *Southern Journal of Communication*, **18 (1)**, 20–23. DOI : 10.1080/10417945209371245.

KANDEL L. & MOLES A. (1958). Application de l'indice de Flesch à la langue française. *Cahiers Études de Radio-Télévision*, **19**, 253–274.

KIYOKAWA H. (1990). A formula for predicting listenability : The listenability of english language materials 2. *Wayo Women's University Language and Literature*, **24**, 57–74.

KOTANI K., UEDA S., YOSHIMI T. & NANJO H. (2014). A listenability measuring method for an adaptive computer-assisted language learningand teaching system. *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing*, p. 387–394.

KOTANI K. & YOSHIMI T. (2017). Effectiveness of linguistic and learner features for listenability measurement using a decision tree classifier. *The Journal of Information and Systems in Education*, **16 (1)**, 7–11.

KUMAI N. (1992). Teaching listening comprehension : What and how (in japanese). *English and American Studies*, **27**, 21–30.

MARTIN L., MULLER B., ORTIZ SUÁREZ P. J., DUPONT Y., ROMARY L., DE LA CLERGERIE É., SEDDAH D. & SAGOT B. (2020). CamemBERT : a tasty French language model. In D. JURAFSKY, J. CHAI, N. SCHLUTER & J. TETREAULT, Éds., *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, p. 7203–7219, Online : Association for Computational Linguistics. DOI : 10.18653/v1/2020.acl-main.645.

MARTINC M., POLLAK S. & ROBNIK-ŠIKONJA M. (2021). Supervised and unsupervised neural approaches to text readability. *Computational Linguistics*, **47**(1), 141–179.

O'KEEFE T. (1971). The comparative listenability of shortwave broadcasts. *Journalism Quarterly*, **48 (4)**, 744–748.

PELTONEN P. (2017). Temporal fluency and problem-solving in interaction : An exploratory study of fluency resources in l2 dialogue. *System*, **70(C)**, 1–13.

PRÉFONTAINE Y. (2010). Differences in perceived fluency and utterance fluency across speech elicitation tasks : a pilot study. *Papers from the Lancaster University Postgraduate Conference in Linguistics Language Teaching*, p. 134–154.

QI P., ZHANG Y., ZHANG Y., BOLTON J. & MANNING C. D. (2020). Stanza : A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics : System Demonstrations*, p. 101–108.

ROGERS J. R. (1962). A formula for predicting the comprehension level of material to be presented orally. *The journal of educational research*, **56 (4)**, 218–220. DOI : 10.1080/00220671.1962.10882926.

RUBIN D. L. (2012). Listenability as a tool for advancing health literacy. *Journal of health communication*, **17 (3)**, 176–190. DOI : 10.1080/10810730.2012.712622.

RUGGIA S. (2019). Le deep learning : un outil pour la didactique du fle ? *Dialettica pedagogica*, **1**, 79–106.

RUGGIA S. (2020). Caractériser un texte en français : les passages-clés des niveaux a1 et a2 du cecrl. *Actes des 15èmes Journées internationales d'Analyse statistique des Données Textuelles*, p. 1–11.

RUGGIA S. (2021). Deepfle : l'intelligence artificielle pour décrire et prédire le(s) niveau(x) du cecrl d'un texte. *Les cahiers de l'AREFLE*, **2 (1)**, 103–109.

RÉVÉSZ A. & BRUNFAUT T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition*, **35**, 31–65.

SEGALOWITZ N., FRENCH L. & GUAY J.-D. (2017). What features best characterize adult second language utterance fluency and what do they reveal about fluency gains in short-term immersion ? *Revue canadienne de linguistique appliquée*, **20(2)**, 90–116.

SKEHAN P. & FOSTER P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, **49**, 93–120.

UEDA S., NANJO H., YOSHIMI T. & KOTANI K. (2013). A listenability formula considering listening proficiency level information of learners of english as a foreign language (in japanese). *The Association for Natural Language Processing, NLP2013*, p. 410–413.

VAJJALA S. (2021). Trends, limitations and open challenges in automatic readability assessment research. *arXiv preprint arXiv :2105.00973*.

WILKENS R., ALFTER D., WANG X., PINTARD A., TACK A., YANCEY K. & FRANÇOIS T. (2022). Fabra : French aggregator-based readability assessment toolkit. *Proceedings of the 13th Language Resources and Evaluation Conference*, p. 1217–1233.

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*, **2021**(2), 229–258.

YOSHIMI K. (2019). The effectiveness and prospects of incorporating current topics into the learning of listening skills : A practical example of listening comprehension 1/2 in the school of contemporary international studies (in japanese). *Bulletin of Nagoya University of Foreign Studies*, **5**, 105–119.

YOSHIMI T. & KOTANI K. (2020). Non-linear regression analysis of the combined listening ease and accuracy index appropriate for learners' proficiency (in japanese). *Transactions of Japanese Society for Information and Systems in Education*, **37 (1)**, 44–49.

Annexes

Manuel	Éditeur	Année	Nb de données
A1			
Alter Ego	Hachette	2006	26
écho	CLE International	2011	63
Texto	Hachette	2016	19
Cosmopolite	Hachette	2017	104
Atelier	Didier	2019	78
A2			
Entre Nous	Maison des langues	2016	70
Tendances	CLE International	2016	81
Texto	Hachette	2016	16
Cosmopolite	Hachette	2017	72
Défi	Maison des langues	2018	35
Atelier	Didier	2019	35
B1			
Entre Nous	Maison des langues	2016	72
Tendances	CLE International	2016	90
Cosmopolite	Hachette	2018	51
Défi	Maison des langues	2019	27
B2			
Édito	Didier	2006	31
LE DELF B2	Didier	2016	60
Entre Nous	Maison des langues	2017	20
Tendances	CLE International	2017	50
Cosmopolite	Hachette	2019	80
C1/C2			
Réussir le DALF	Didier	2007	31
abc DALF	CLE International	2014	57
Le DALF 100 % réussite	Didier	2017	75
Édito	Didier	2018	70
Tendances	CLE International	2019	10

TABLE 3 – Liste des manuels

Variable	Méthode de calcul
Nombre de syllabes par minute	Nombre total de syllabes / durée totale d'enregistrement [minute].
Nombre de syllabes par minute sans les pauses	Nombre total de syllabes / (durée totale d'enregistrement [minute] - durée totale de pauses [minute]).
Nombre de mots par minute	Nombre total de mots / durée totale d'enregistrement [minute].
Nombre de mots par minute sans les pauses	Nombre total de mots / (durée totale d'enregistrement [minute] - durée totale de pauses [minute]).
Durée totale	Durée totale d'un enregistrement.

Durée moyenne des pauses	Durée totale de pauses [minute] / nombre total de pauses.
Nombre de pauses par minute	Nombre total de pauses / durée totale d'enregistrement [minute].

TABLE 4 – Liste des variables de fluidité

Famille	Variable	Description
Basé sur la longueur		
Longueur du mot	LENwrdSYL	Nombre de syllabes par mot.
Longueur de la phrase	LENsntWRD	Nombre de tokens par phrase, à l'exclusion de la ponctuation.
Variables lexicales		
Chevauchement de contenu	LEXcovLGAL	Tout lemme est partagé dans toutes les phrases.
Diversité lexicale	LEXdvrFLC	CTTR de tous les types de lemmes de noms, de noms propres, de verbes, d'adjectifs et d'adverbes dans le texte, en tenant compte de tous les tokens.
Fréquence lexicale	LEXfrqFCL	Fréquence de la forme du lemme de tous les noms, noms propres, verbes, adjectifs et adverbes en fonction de leur occurrence dans le corpus FLELex.
	LEXfrqLCL	Fréquence de la forme du lemme de tous les noms, noms propres, verbes, adjectifs et adverbes sur la base de l'occurrence dans le corpus Lexique3.
Lexiques gradués	LEXgrdBA1	Proportion de mots dans les descripteurs de niveau de référence du français de Beacco pour le niveau du CECR A1.
	LEXgrdFA1	Fréquence des mots dans la ressource FLELex pour le niveau du CECR A1.
	LEXgrdFFOA1	Proportion de lemmes de niveau A1 selon la méthode de la première occurrence (ni-veau du CECR où un mot est rencontré pour la première fois).
	LEXgrdFMLA1	Proportion de lemmes de niveau A1 selon la méthode d'apprentissage automatique de Pintard et François, 2020 (https ://aclanthology.org/2020.readi-1.13.pdf) entraînée sur les descripteurs de niveau de référence Beacco.
	LEXgrdFSOOUA1	Proportion de lemmes de niveau A1 selon la méthode Significant Onset of Use (Alfter et al., 2016; https ://aclanthology.org/W16-6501.pdf)
Voisins orthographiques	LEXnghPHO	Distance phonologique moyenne de Le-venstein, calculée sur Lexique3.
Normes lexicales	LEXnrmCNCR	Niveau de concrétion des mots.
	LEXnrmFAM	La familiarité des mots, également appe-lée fréquence subjective.
	LEXnrmIMG	Imageabilité de chaque mot.

Sophistication lexicale	LEXsopFK1	Nombre de lemmes dans les premières bandes de fréquences de 1000 mots de FLELex.
	LEXsopGK1	Nombre de mots dans les premières bandes de fréquence de 1000 mots de la liste de vocabulaire de Gougenheim.
	LEXsopLLK1	Nombre de lemmes dans les premières bandes de fréquence de 1000 mots de Lexique3.
Caractéristiques graduées	LEXgrdBLA1	Expressions lexicales de niveau A1 dans le chapitre 5 de Beacco.
Variables syntaxiques		
Développement du langage	SYNdevNPHRS	Nombre de constituants/phrases.
	SYNdevTUL	Longueur des T unités.
	SYNdevVG1	Nombre de verbes du 1er groupe français dans le texte.
Caractéristiques de la clause	SYNclsLEN	Longueur de la clause.
Caractéristiques des temps verbaux	SYNtnsfINDP	Indicatif présent.
	SYNtnsfINDI	Indicatif imparfait.
	SYNtnsfINDPS	Indicatif passé simple.
	SYNtnsfINDPC	Indicatif passé composé.
	SYNtnsfINDPQP	Indicatif plus-que-parfait.
	SYNtnsfINDPA	Indicatif passé antérieur.
	SYNtnsfINDFS	Indicatif futur simple.
	SYNtnsfINDFA	Indicatif futur antérieur.
	SYNtnsfCNDP	Conditionnel présent.
	SYNtnsfCNDPS	Conditionnel passé.
	SYNtnsfIMPP	Impératif présent.
	SYNtnsfIMPPS	Impératif passé.
	SYNtnsfSUBP	Subjonctif présent.
	SYNtnsfSUBPS	Subjonctif passé.
	SYNtnsfSUBI	Subjonctif imparfait.
	SYNtnsfSUBPQP	Subjonctif plus-que-parfait.
	SYNtnsfINF	Infinitif.
	SYNtnsfINFC	Infinitif composé.
	SYNtnsfPP	Participe présent.
	SYNtnsfPPS	Participe passé.
	SYNtnsfGERP	Gérondif.
	SYNtnsfGERPS	Gérondif passé.
	SYNtnsfTAP	Temps passif.
POS Tag	SYNposADJ	Mots identifiés comme ADJ (adjectif), suivant les étiquettes POS universelles et l'analyseur Stanza.
	SYNposADP	Mots identifiés comme ADP (adposi-tion), suivant les étiquettes POS univer-selles et l'analyseur Stanza.
	SYNposADV	Mots identifiés comme ADV (adverbe), suivant les étiquettes POS universelles et l'analyseur Stanza.

	SYNposAUX	Mots identifiés comme AUX (auxiliaire), suivant les étiquettes POS universelles et l'analyseur Stanza.
	SYNposCCONJ	Mots identifiés comme CCONJ (conjonc-tion de coordination), suivant les éti-quettes POS universelles et l'analyseur Stanza.
	SYNposINTJ	Mots identifiés comme INTJ (interjec-tion), suivant les étiquettes POS univer-selles et l'analyseur Stanza.
	SYNposNOUN	Mots identifiés comme NOUN (nom), suivant les étiquettes POS universelles et l'analyseur Stanza.
	SYNposNUM	Mots identifiés comme NUM (numéral), suivant les étiquettes POS universelles et l'analyseur Stanza.
	SYNposPART	Mots identifiés comme PART (particule), suivant les étiquettes POS universelles et l'analyseur Stanza.
	SYNposPRON	Mots identifiés comme PRON (pronom), suivant les étiquettes POS universelles et l'analyseur Stanza.
	SYNposPROPN	Mots identifiés comme PROPN (nom propre), suivant les étiquettes POS uni-verselles et l'analyseur Stanza.
	SYNposSCONJ	Mots identifiés comme SCONJ (conjonc-tion de subordination), suivant les éti-quettes POS universelles et l'analyseur Stanza.
	SYNposVERB	Mots identifiés comme VERB (verbe), suivant les étiquettes POS universelles et l'analyseur Stanza.
	SYNposX	Mots identifiés comme X (autre), suivant les étiquettes POS universelles et l'analyseur Stanza.
Variables de lisibilité		
Service de formule de lisibilité	REAfrmKM	Formule de lisibilité pour le français (Kandel-Moles).

TABLE 5 – Liste des variables de lisibilité

Variables	Corrélation Spearman	Variables	Corrélation Spearman
LENsntWRD	0,63	LEXgrdFSOOUA1	-0,61
syllables_per_minute	0,57	REAfrmKM	-0,56
LEXdvrFLC	0,47	LEXgrdBA1	-0,45
SYNtnsfPPS	0,41	pauses_per_minute	-0,35
SYNtnsfPP	0,37	length_pauses	-0,34
SYNtnsfTAP	0,36	SYNtnsfINDP	-0,30
SYNposADP	0,36	SYNposINTJ	-0,25
SYNposSCONJ	0,31	LEXnrmIMG	-0,24
SYNtnsfSUBP	0,30		
SYNposNOUN	0,29		
LEXnghPHO	0,27		
SYNtnsfINDI	0,26		
SYNposADJ	0,25		
SYNtnsfINDPQP	0,25		
SYNtnsfINDFS	0,23		
SYNtnsfINF	0,22		

TABLE 6 – 24 variables finales et leurs coefficients de corrélation de Spearman avec le niveau du CECR dans l'ensemble des données (p<,01)

Variables	Corrélation Spearman	Variables	Corrélation Spearman
LENsntWRD	0,65	LEXgrdFSOOUA1	-0,61
syllables_per_minute	0,64	REAfrmKM	-0,57
LEXdvrFLC	0,50	LEXgrdBA1	-0,45
SYNposADP	0,40	pauses_per_minute	-0,43
SYNposSCONJ	0,40	SYNtnsfINDP	-0,38
SYNtnsfTAP	0,40	length_pauses	-0,37
SYNtnsfPP	0,39	SYNposINTJ	-0,35
SYNtnsfPPS	0,37	LEXnrmIMG	-0,34
SYNtnsfSUBP	0,36		
SYNtnsfINDI	0,35		
SYNtnsfINDPQP	0,32		
SYNposNOUN	0,27		
SYNtnsfINDFS	0,26		
SYNposX	0,24		
SYNposADJ	0,23		
SYNtnsfCNDPS	0,23		
LEXnghPHO	0,22		
SYNtnsfINDPC	0,22		
SYNtnsfINF	0,21		
SYNtnsfINFC	0,21		
SYNtnsfINDPS	0,20		

TABLE 7 – 29 variables finales et leurs coefficients de corrélation de Spearman avec le niveau du CECR dans le dialogue (p<,01)

Variables	Corrélation Spearman	Variables	Corrélation Spearman
SYNdevTUL	0,57	LEXgrdFMLA1	-0,61
LEXdvrFLC	0,51	REAfrmKM	-0,53
SYNtnsfPPS	0,45	LEXgrdBA1	-0,42
syllables_per_minute	0,38	length_pauses	-0,26
SYNtnsfPP	0,36	SYNposPRON	-0,22
SYNtnsfTAP	0,31	pauses_per_minute	-0,22
LEXnghPHO	0,30		
SYNposNOUN	0,27		
SYNposADV	0,26		
SYNposADP	0,26		
SYNposSCONJ	0,24		
SYNtnsfSUBP	0,23		
SYNposADJ	0,22		
SYNtnsfINF	0,21		
SYNtnsfINDFS	0,21		

TABLE 8 – 21 variables finales et leurs coefficients de corrélation de Spearman avec le niveau du CECR dans le monologue (p<,01)