Thèse de doctorat

Un modèle de la facilité d'écoute des documents audios pour les apprenants du français langue étrangère

Université Seinan Gakuin

Université catholique de Louvain

Minami Ozawa

Table des matières

Remerciements

Dans le cadre de la rédaction de cette thèse, j'ai bénéficié du soutien et des conseils de nombreuses personnes. Je tiens tout d'abord à exprimer ma profonde gratitude à ma directrice de recherche, madame Kaori Sugiyama de l'université Seinan Gakuin, pour ses conseils avisés et constants tout au long de mes études. Je remercie également monsieur Thomas François de l'université catholique de Louvain, qui m'a particulièrement accompagnée dans l'analyse menée à l'aide du traitement automatique du langage. Je leur adresse à tous deux mes sincères remerciements.

Je remercie également les membres du jury, monsieur Sylvain Detey de l'université Waseda et monsieur Jean-Luc Azra de l'université Seinan Gakuin, pour leurs remarques précieuses et leurs suggestions constructives.

Cette thèse a été réalisée dans le cadre d'une cotutelle entre l'université Seinan Gakuin et l'université catholique de Louvain. Je souhaite remercier très sincèrement madame Yuko Takematsu, ainsi que toutes les personnes ayant contribué à la mise en place de cet accord.

Je suis également reconnaissante envers tous les professeurs de français de l'université Seinan Gakuin, au sein de laquelle j'ai étudié depuis 2013. Grâce à eux, j'ai découvert le plaisir d'apprendre le français et j'ai été guidée, au fil des années, vers la recherche en lien avec cette langue. Je leur adresse mes plus vifs remerciements.

Au sein du CENTAL de l'université catholique de Louvain, j'ai été accueillie chaleureusement et ai pu mener mes recherches dans un environnement enrichissant. Je remercie en particulier monsieur Rodrigo Souza Wilkens, avec qui j'ai collaboré dans le cadre d'un projet de recherche. Lors de la rédaction de cette thèse en français, langue qui n'est pas ma langue maternelle, j'ai reçu une aide précieuse de la part de monsieur Hubert Naets et de monsieur Julien Zakhia Doueihi. Je leur suis profondément reconnaissante.

Le travail de recherche mené au CENTAL et la rédaction de cette thèse n'auraient pas été possibles sans le soutien financier de la bourse de la Wallonie-Bruxelles International que j'ai reçue d'octobre 2019 à septembre 2020. Je tiens à exprimer ici ma reconnaissance pour cette aide déterminante.

Enfin, je remercie de tout cœur madame Nami Yamaguchi et tous les collègues en master et doctorat à l'université Seinan Gakuin, avec qui j'ai partagé des moments d'émulation et de soutien mutuel durant la rédaction de cette thèse, ainsi que mes amis qui m'ont toujours encouragée avec chaleur. Enfin, je souhaite également exprimer ma plus profonde gratitude à ma famille, qui m'a soutenue tout au long de mon long parcours universitaire.

Introduction

La compréhension orale constitue une compétence essentielle dans le cadre de la communication linguistique. Cependant, son importance a longtemps été sous-estimée. Bien que l'expression orale et la compréhension orale relèvent toutes les deux des compétences orales, la partie compréhension a souvent été considérée comme un élément secondaire de l'expression orale, plutôt que comme une base pour celle-ci (Newmark & Diller, 1964: 18).

Les recherches exposées dans cette thèse portent sur la compétence de compréhension orale. Quels sont les meilleurs moyens pour les apprenants, en particulier les apprenants du français, d'améliorer leur compétence en compréhension orale ? L'un des moyens consiste à utiliser un support authentique et à rendre l'apprentissage attractif. En effet, le support authentique permet aux apprenants d'imaginer une communication réelle. Il peut être utilisé non seulement pour communiquer, mais aussi pour s'entraîner à écouter et à comprendre par exemple des informations à la télévision ou à la radio. Cependant, l'utilisation du discours authentique se heurte au fait qu'il ne s'avère pas toujours évident de déterminer le type de discours le plus pertinent en vue de l'apprentissage. Divers facteurs peuvent être pris en compte dans les critères de sélection, tels que les intérêts personnels ou l'accessibilité de l'audio. Si la difficulté ou le niveau du Cadre européen commun de référence pour les langues (CECR) de l'audio authentique peut être prédit, il peut constituer l'un des critères de sélection. Afin de développer l'enseignement de la compréhension orale du français, cette thèse propose un modèle de prédiction automatique du niveau de difficulté des documents audios français, qui permet de déterminer la facilité d'écoute d'une production orale.

Cette thèse est organisée en sept chapitres. Tout d'abord, le chapitre 1 présentera la compréhension orale, y compris le processus et les méthodes d'enseignement. Puis, le chapitre 2 définira la facilité d'écoute et retracera son histoire, tandis que le chapitre 3 décrira la méthodologie de cette étude. Ensuite, les chapitres 4 et 5 traiteront de la conception des modèles d'évaluation de la facilité d'écoute, en utilisant deux méthodes différentes et en présentent les résultats. Enfin, après une analyse comparative de ces résultats au chapitre 6, cette thèse se termine au chapitre 7 par une discussion sur l'applicabilité à l'enseignement.

1. Compréhension orale

Les examens du diplôme d'études en langue française (DELF) et du diplôme approfondi de langue française (DALF) évaluent quatre compétences en ce qui concerne l'enseignement des langues étrangères. Ainsi, il est courant de distinguer quatre grandes compétences linguistiques : la compréhension orale, la compréhension écrite, la production orale et la production écrite. Cette thèse se concentre sur la compréhension orale qui constitue un élément central de la communication, représentant plus de 45 % de la communication langagière de la vie quotidienne et se trouvant au centre de l'activité langagière (Feyten, 1991: 174). L'acte de compréhension orale est ancré profondément dans notre vie langagière et communicative. Dès lors, il convient de s'interroger sur le processus de la compréhension orale.

Dans ce chapitre, nous définirons la compréhension orale dans la section 1.1., puis nous décrirons son processus dans la section 1.2. Ensuite, dans la section 1.3., nous détaillerons les éléments utilisés dans le processus de la compréhension orale. Après avoir ainsi présenté la structure de cette dernière, la section 1.4. mettra en évidence les difficultés que rencontrent les apprenants dans ce domaine, tandis que la section 1.5. présentera les méthodes d'enseignement et d'apprentissage de la compréhension orale expérimentés jusqu'à présent.

1.1. Définition de la compréhension orale

Avant de définir la compréhension orale, il convient d'abord de clarifier la différence entre « écouter » et « entendre ». Rost (2002: 12) indique que ce qui distingue le fait d'écouter (*listening*) de celui d'entendre (*hearing*) repose sur le degré d'intention. La compréhension orale étudiée dans cette thèse constitue un acte d'écoute qui requiert un haut niveau d'intention afin de mener à la compréhension.

Il existe différentes définitions de la compréhension orale. Call (1985: 766) l'explique comme une des propositions :

> *The act of listening to and understanding a spoken language can be described as a series of processes through which the sounds associate with a particular utterance are converted into meaning.*

Du point de vue des ressources nécessaires à la compréhension orale, O'Malley *et al.* (1989: 434) propose la définition suivante :

> *Listening comprehension is an active and conscious process in which the listener constructs meaning by using cues from contextual information and from existing knowledge, while relying upon multiple strategic resources to fulfill the task requirements.*

En tenant compte de ces définitions, le processus de compréhension orale débute par un processus actif d'écoute et va jusqu'à l'interprétation du sens des informations auditives. À la lumière de ce qui précède, nous utilisons le terme « compréhension orale » dans le sens de comprendre le langage en lui-même et le contenu d'un énoncé à travers le son, en tenant compte du contexte de l'énoncé, de l'intention derrière celui-ci et des implications qu'il comporte, avec volonté et en s'appuyant sur les connaissances préexistantes.

1.2. Processus de la compréhension orale

Dans la section précédente, nous avons indiqué que la compréhension orale constitue un processus actif ; cette section donnera un aperçu des processus spécifiques impliqués. Le processus change selon que la compréhension orale se fait dans la langue maternelle ou dans une langue étrangère. Cependant, certains aspects du processus dans la langue étrangère se trouvent fortement influencés par celui dans la langue maternelle. Dans cette section, l'accent sera mis principalement sur le processus dans la langue étrangère, en référence au processus de compréhension orale dans la langue maternelle.

1.2.1. Modèle théorique de la compréhension orale

Le processus de compréhension orale fait intervenir des opérations cognitives et des comportements internes de sorte qu'il ne s'avère pas facile de le mesurer directement (Rubin, 1994: 210). Ce processus se révélant si compliqué à observer, il fait l'objet de recherches dans divers domaines et selon diverses optiques – psychologie cognitive, psychologie linguistique, science cognitive du cerveau, physiologie du cerveau, psychologie auditive ainsi qu'acoustique (Akita, 2009: 99). De plus, deux modèles coexistent : le premier, lié au fait d'écouter, consiste à recevoir exactement ce que dit l'interlocuteur et à le comprendre ; le second modèle comportemental correspond à l'attitude à écouter (*ibid.*, 99). Dans cette thèse, nous traiterons le processus de compréhension dans une optique inspirée de la psychologie cognitive.

Par ailleurs, le modèle appliqué dans de nombreuses études sur la compréhension orale correspond à celui d'Anderson (1980). Ainsi, Anderson (1980: 401-402) propose un modèle de compréhension globale de la première langue qui tient compte de trois phases chronologiques : la perception, l'analyse et l'utilisation. Tout d'abord, le décodage du message se fait durant la première phase. Une représentation basée sur la signification de la séquence originale de mots peut être conservée dans la mémoire à court terme. Par exemple, avec une séquence de la forme « nom + verbe + nom », le locuteur veut dire qu'une instance du premier nom a la relation spécifiée (verbe) avec une instance du deuxième nom. En revanche, si la phrase adopte la forme « nom + verbe passif + par + nom », le locuteur exprime qu'un élément représenté par le second nom entretient la relation

spécifiée avec le premier nom. Ainsi, la connaissance de la structure permet d'apprécier la différence entre « *A doctor shot a lawyer* » et « *A doctor was shot by a lawyer* » en anglais (*ibid.*, 402).

Dans la deuxième phase, les mots du message se transforment en une représentation mentale de la signification combinée des mots du message. En utilisant des connaissances syntaxiques permettant de comprendre les énoncés, il s'avère également possible d'utiliser la signification des mots impliqués. Ainsi, la signification d'une chaîne de mots peut être déterminée simplement en réfléchissant à la manière dont ces mots peuvent être assemblés de manière à donner un sens. Même si la phrase « *Jane fruit eat* » en anglais ne correspond pas à la syntaxe de l'anglais, il est possible de comprendre sa signification (*ibid.*, 411). Le principal indice de la segmentation en compréhension orale correspond au sens, qui peut être représenté syntaxiquement, sémantiquement, phonologiquement ou par toute combinaison de ces éléments (O'Malley *et al.*, 1989: 420).

Dans la troisième phase, l'interprétation se fait en reliant la représentation en mémoire à court terme à la connaissance préalable conservée dans la mémoire à long terme. Fondamentalement, cette tâche consiste à relier les nouvelles informations aux anciennes. La plupart des phrases d'une communication compréhensible contiennent à la fois des informations nouvelles et anciennes, car le locuteur, en essayant d'affirmer de nouvelles informations, doit les relier aux anciennes informations connues de l'auditeur. Ainsi, le locuteur suppose les anciennes informations afin d'affirmer les nouvelles informations (Anderson, 1980: 416).

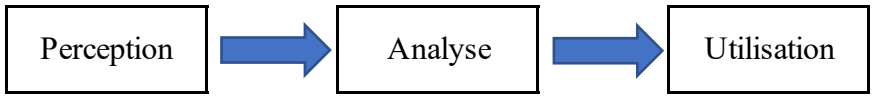

Figure 1 : Modèle de la compréhension d'après Anderson (1980)

Ces trois étapes – perception, analyse et utilisation – s'avèrent nécessairement partiellement ordonnées dans le temps ; cependant, elles se chevauchent aussi en partie. Les auditeurs peuvent faire des déductions à partir de la première partie d'une phrase tout en percevant une partie ultérieure (*ibid.*, 402).

Un modèle proposé après Anderson (1980) est celui de Cutler & Clifton (1999). Ce modèle comporte quatre phases : le décodage, la segmentation, la reconnaissance et l'interprétation. Lors de la phase de décodage, des informations phonétiques sont transformées en une représentation abstraite. Ainsi, l'auditeur doit d'abord séparer la parole de toute autre entrée auditive qui pourrait atteindre simultanément ses oreilles ; il doit ensuite la transformer en une représentation plus abstraite (*ibid.*, 123).

La phase suivante, la segmentation, consiste, quant à elle, à effectuer un traitement phonologique segmentaire ou supra-segmentaire. Elle ne constitue pas une phase distincte, comme le montrent les cases superposées de la figure 2, mais se déroule en même temps que d'autres processus.

Ensuite, la phase de reconnaissance est divisée en deux parties fortement reliées, l'une correspond à la reconnaissance de mot et l'autre à l'interprétation de l'énonciation. Enfin, lors de la dernière phase, comme le montre la figure 2, l'interprétation ne représente pas l'étape finale du processus d'extraction par l'auditeur du message du locuteur à partir de l'entrée auditive. En effet, l'énoncé doit être relié à son contexte discursif au sens large, au-delà des relations thématiques. En outre, les connaissances du locuteur doivent être prises en compte, de même qu'une série de facteurs sociolinguistiques impliqués dans l'interaction conversationnelle (*ibid.*, 125).

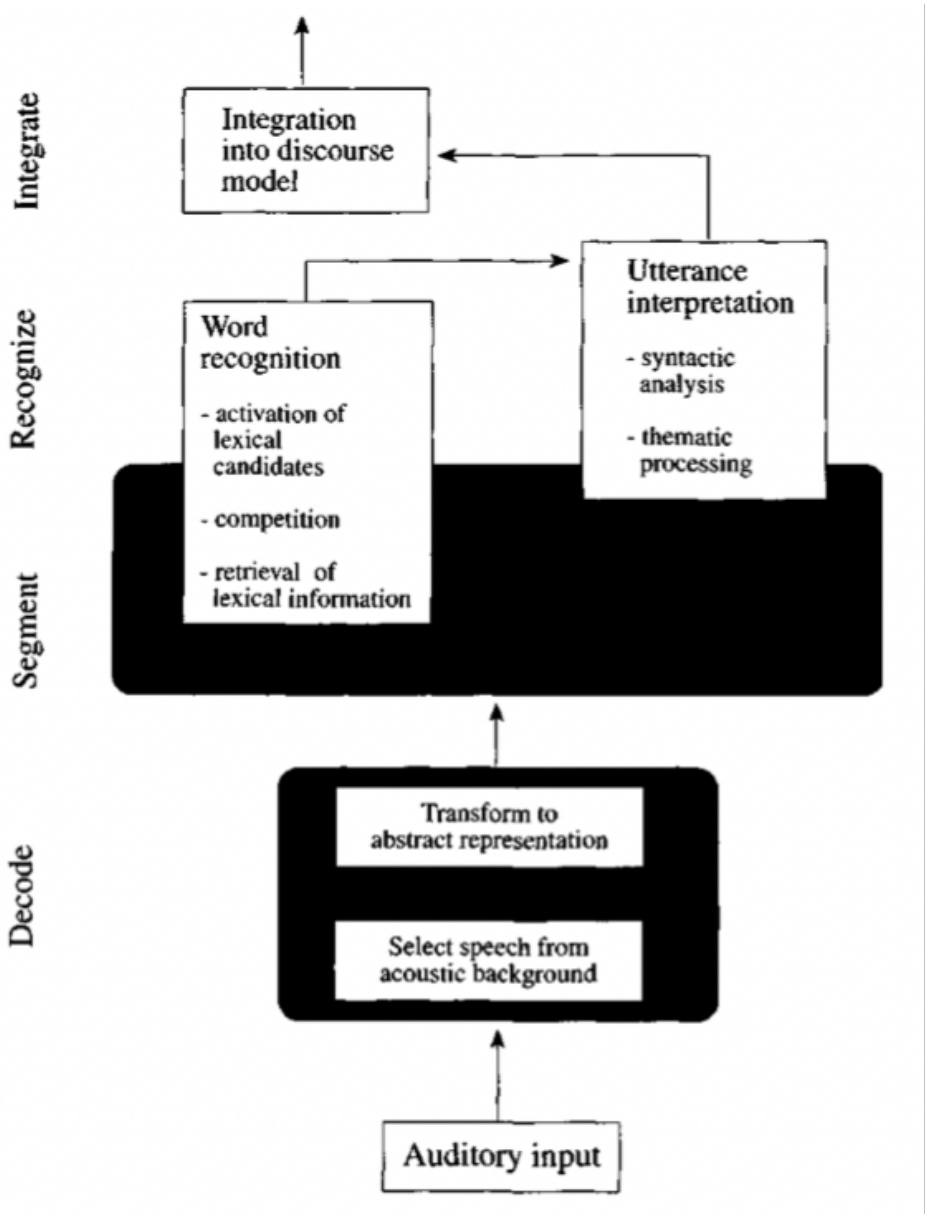

Figure 2 : Modèle de la compréhension orale d'après Cutler & Clifton (1999: 124)

Nous proposons d'examiner ce modèle plus en détail. Lors de l'étape de décodage, la première tâche de l'auditeur consiste à séparer la parole des autres signaux auditifs qui lui parviennent en même temps (*ibid.*, 125-126). Après avoir isolé la partie du signal entrant qui correspond à de la parole, l'auditeur peut alors commencer le décodage. Il s'agit de transformer un signal d'entrée continu, variable dans le temps, en une représentation composée d'éléments discrets (*ibid.*, 126).

Dans la phase de segmentation, le vocabulaire candidat se trouve activé, les candidats lexicaux sont mis en concurrence et les informations lexicales sont explorées. En d'autres termes, les auditeurs choisissent le vocabulaire approprié. L'activation des représentations lexicales constitue un processus continu, basé sur toutes les informations disponibles (*ibid.*, 136). Les auditeurs doivent faire attention au mot approprié puisque tout mot prononcé tend à ressembler à d'autres mots (*ibid.*, 133). Dès lors, ils doivent traiter la recherche d'informations lexicales.

Le processus de reconnaissance ne s'arrête pas à l'identification des mots et de leur signification. En effet, déterminer le message véhiculé par une séquence de mots implique bien plus qu'une simple addition des significations des mots. Tout d'abord, l'énoncé qui les contient doit être divisé en ses éléments constitutifs. Puis, les relations entre ces éléments doivent être déterminées et interprétées sémantiquement. De plus, la référence des éléments, leur relation avec le discours en cours, ainsi que la vérité ou la force communicative de la phrase ou du discours dans son ensemble doivent être établies. Ce processus repose sur la connaissance de l'utilisateur de la structure de sa langue et sur les informations spécifiques fournies par les mots d'une phrase (*ibid.*, 141). Ainsi, ce modèle vise à expliquer la manière dont une entrée vocale est convertie en une représentation du sens. Il est dessiné comme englobant différents niveaux de traitement, avec un flux d'informations unidirectionnel depuis l'entrée du son jusqu'à la sortie du sens de l'énoncé (*ibid.*, 151).

Parmi les modèles théoriques, celui d'Anderson (1980) a davantage influencé les recherches sur la compréhension orale en la langue étrangère. Notamment, O'Malley *et al.* (1989) mettent en évidence que le modèle d'Anderson (1980) semble susceptible d'être appliqué au processus de compréhension orale chez les apprenants d'une deuxième langue (L2). Ils ont ainsi examiné les processus mentaux utilisés par les apprenants d'une langue étrangère lors de la compréhension orale, les stratégies qu'ils utilisent au cours des différentes phases de la compréhension et les différences dans l'utilisation des stratégies entre les auditeurs efficaces et peu efficaces. Les participants à cette étude étaient onze lycéens inscrits à des cours d'anglais langue seconde dans deux lycées du nord-est des États-Unis. Tous les élèves étaient classés par le district scolaire au niveau intermédiaire de maîtrise de l'anglais, que le district définit comme une maîtrise limitée de la compréhension et de l'expression orale de l'anglais, et une maîtrise faible ou nulle de la lecture et de l'écriture de l'anglais. Tous les participants

étaient originaires de pays hispanophones d'Amérique centrale ou d'Amérique du Sud (*ibid.*, 424-425). Les élèves ont été classés comme des auditeurs efficaces ou peu efficaces. Les critères d'un auditeur efficace ont été déterminés par les enseignants et comprenaient l'attention en classe, la capacité à suivre les instructions sans demander d'éclaircissements, la capacité et la volonté de comprendre le sens général d'un passage difficile, la capacité à répondre de manière appropriée dans une conversation, et la capacité et la volonté de deviner le sens de mots et d'expressions non familiers. L'application de ces critères a permis de sélectionner huit auditeurs efficaces et trois peu efficaces, bien que les données ne soient rapportées que pour cinq auditeurs efficaces et les trois inefficaces (*ibid.*, 425).

Ces données ont été collectées à l'aide de procédures de *think-aloud* en se basant sur des tâches linguistiques académiques. Les procédures de *think-aloud* consistent à interrompre les participants pendant une activité de compréhension orale et à leur demander d'indiquer ce à quoi ils pensent (*ibid.*, 418). Lors de l'activité sur la langue anglaise, les élèves ont été invités à écouter un passage enregistré en anglais qui contenait huit pauses. Après chaque pause, ils devaient réfléchir à haute voix à la manière dont ils avaient compris ce qu'ils avaient entendu : la présence de mots inconnus, ce qu'ils n'avaient pas compris, la manière dont ils l'avaient compris, et si des souvenirs ou des images leur étaient venus à l'esprit pendant qu'ils écoutaient. Les élèves avaient la possibilité de réfléchir à haute voix en espagnol, leur langue maternelle, ou en anglais (*ibid.*, 425).

Chaque session de collecte de données comprenait trois phases : un échauffement, une transition et un rapport verbal. Lors de l'échauffement, le chercheur posait des questions générales sur le pays d'origine de l'élève, la durée de sa résidence aux États-Unis, son niveau d'éducation dans sa langue maternelle ou d'autres questions générales appropriées pour les introductions de chaque session. Lors de la transition, le chercheur a présenté un problème mathématique ou logique à l'élève, lui a demandé de réaliser le *think-aloud* pendant qu'il travaillait à la solution, et l'a ensuite prié d'évaluer dans quelle mesure ce qu'il avait rapporté reflétait son processus de réflexion réel. Au stade du rapport verbal, les élèves se voyaient proposer trois activités d'écoute par session, choisies parmi les types d'activités suivants : un cours d'histoire, une expérience scientifique, un cours de science, une histoire courte et une dictée (*ibid.*, 425-426). À la fin du processus de *think-aloud*, le chercheur a posé des questions de compréhension sur

l'idée principale du passage ou sur la signification de termes spécifiques (ibid., 426).

À l'aide de ces données, les commentaires des élèves pendant les séances de *think-aloud* ont été examinés afin de déterminer les distinctions entre les différentes phases de la compréhension orale, les types de stratégies utilisées dans chaque phase et les différences d'utilisation des stratégies entre les auditeurs efficaces et inefficaces (*ibid.*, 427). Les auteurs ont constaté que les processus mentaux utilisés par les élèves lors de la compréhension orale s'avéraient parallèles aux trois phases du processus de compréhension dérivées de la théorie : le traitement perceptif, l'analyse syntaxique et l'utilisation. Ainsi, le modèle d'Anderson (1980) paraît approprié comme modèle de la compréhension orale d'une langue étrangère.

Goh (2000) a, quant à lui, identifié les difficultés de la compréhension orale d'un point de vue cognitif, en utilisant aussi le modèle en trois étapes d'Anderson (1980) avec 40 apprenants d'anglais d'origine chinoise. Les données ont été collectées à partir de trois sources. La principale source était constituée de journaux hebdomadaires tenus par 40 étudiants dans le cadre de leur cours de compréhension orale. Dans ces journaux, les étudiants ont écrit sur des événements de la compréhension orale réels et ont décrit la manière dont ils ont essayé de comprendre ce qu'ils ont entendu et les problèmes rencontrés. De plus, 17 étudiants sur ces 40 ont participé à des entretiens en petits groupes. Il s'agissait d'entretiens semi-structurés visant à découvrir ce que les étudiants savaient sur la tâche consistant à apprendre à écouter l'anglais. Parmi eux, 23 étudiants ont également fait une verbalisation rétrospective immédiate (Ericsson & Simon, 1984, 1987) concernant la collecte de données verbales, dès qu'ils ont terminé leurs tâches. Même si l'objectif principal de ces séances consistait à examiner les stratégies de traitement utilisées, les protocoles de rappel des étudiants incluaient souvent des descriptions de difficultés liées à la compréhension orale. Suite à l'analyse de ces données, dix problèmes de traitement survenus lors de la phase de compréhension orale des étudiants ont été mis en évidence :

- phase de perception :
  – **ne pas reconnaître des mots qu'ils connaissent pourtant,**
  – **négliger la partie suivante lorsqu'ils réfléchissent à la signification,**
  – ne pas pouvoir fragmenter les flux de parole,
  – manquer le début des audios,
  – se concentrer trop fort ou, au contraire, se révéler incapable de se concentrer ;
- phase d'analyse :
  – **oublier rapidement ce qui est entendu,**
  – **ne pas être capable de former une représentation mentale à partir des mots entendus,**
  – mal comprendre des parties suivantes en raison de problèmes antérieurs ;
- phase d'utilisation :
  – **comprendre les mots indépendamment, mais pas le message global,**
  – faire des confusions au sujet des idées clés du message[1].

Les problèmes rencontrés au stade de la perception concernaient principalement la reconnaissance des sons en tant que mots ou groupes de mots distincts. Ils comprenaient également des difficultés d'attention. Les problèmes d'analyse, quant à eux regroupaient diverses difficultés liées au développement d'une représentation mentale cohérente des mots entendus. Lors de la phase d'utilisation, certains apprenants ont éprouvé des difficultés à comprendre le message voulu par le locuteur. Des difficultés sont également apparues à ce stade lorsque l'auditeur s'avérait incapable de traiter davantage le texte oral en raison soit d'un manque de connaissances préalables, soit d'une application inappropriée de ces dernières.

Nous avons présenté ci-dessus un aperçu des modèles de la compréhension orale, en nous appuyant notamment sur Anderson (1980). Comme le souligne Goh (2000), il faut reconnaître que la compréhension orale des langues étrangères s'avère difficile pour les apprenants. Dès lors, après avoir compris le mécanisme par lequel la compréhension orale est effectuée, la section suivante fournira un aperçu

[1] Ce résumé est basé sur le tableau 1 (*Problems related to different phrases of listening comprehension*) et la figure 1 (*Five common listening comprehension problems*) dans Goh (2000: 59-60). Les problèmes particulièrement signalés par les apprenants sont mis en gras par l'auteur.

du flux d'informations et de la direction du traitement au cours du processus de compréhension oral.

1.2.2. Modèle de traitement des informations de la compréhension orale

Après avoir décrit le modèle théorique dans les sections précédentes, il convient de s'interroger sur la manière dont le traitement de l'information s'effectue dans la pratique à la lumière de la théorie. Les auditeurs utilisent deux types de connaissances afin d'identifier les propositions : la connaissance de la syntaxe de la langue cible et celle du monde réel (Richards, 1983: 220). En ce qui concerne le modèle connu de traitement, il convient de distinguer les différents modèles : ascendant, descendant et interactif. Ce dernier tire profit de deux différents sens de modèles. Par ailleurs, ces trois modèles de traitement de l'information ne concernent pas l'ordre chronologique (Fukuda, 2002: 369).

Tout d'abord, le modèle ascendant constitue un processus dans lequel les auditeurs comprennent en commençant par la plus petite unité de message acoustique, à savoir les phonèmes et les syllabes. Par la suite, ces unités sont combinées en mots, qui, à leur tour, forment des propositions et des phrases (Field, 1999: 338). Ainsi, il s'agit du processus d'utilisation des connaissances de la syntaxe. Selon ce modèle, la communication peut avoir lieu sans aucune référence au locuteur, à l'auditeur ou au contexte plus large (Flowerdew & Miller, 2005: 25). Les auditeurs qui utilisent des connaissances de la syntaxe emploient également des propositions et des schémas[2] en mémoire à long terme, mais les informations utilisées sont basées sur les règles grammaticales ou syntaxiques. Les auditeurs qui interprètent le sens en fonction des caractéristiques linguistiques du texte oral utilisent un traitement « ascendant » et se trouvent obligés de déterminer le sens de mots individuels, ensuite de les agréger vers le haut en unités de sens plus grandes (O'Malley *et al.*, 1989: 421). Cette approche se révèle problématique puisqu'elle est influencée par l'interférence de la langue maternelle (Byrnes, 1984: 323).

[2] Le schéma est défini comme une structure mentale, composée de connaissances individuelles pertinentes, de mémoire et d'expérience, qui permet d'incorporer ce que est appris dans ce que est su (Anderson & Lynch, 1988: 14).

De plus, il existe un argument selon lequel le modèle ascendant se révèle incorrect pour les trois raisons suivantes. (1) Il n'existe pas de correspondance simple entre les segments du signal vocal et les sons perçus. En effet, la suppression d'une partie du signal de la parole ne signifie pas que le reste du son n'est pas entendu comme faisant partie du mot d'origine. (2) Pour de nombreux phonèmes, il n'existe pas de caractéristiques distinctives constantes qui les distinguent absolument de tous les autres. Le contexte du mot environnant affecte en effet les caractéristiques acoustiques du phonème. (3) Même en se plaçant au niveau des mots, contrairement au niveau des phonèmes, seulement environ la moitié des mots peuvent être reconnus isolément. Cela est vrai lorsque des mots individuels se trouvent extraits d'enregistrements de conversations et présentés aux auditeurs pour identification (Anderson & Lynch, 1988: 22-23).

Au contraire du modèle ascendant, le modèle descendant consiste à comprendre le sens en faisant appel à des connaissances et à l'expérience sur le monde réel que l'auditeur possède (Yin, 2002: 280). Ainsi, ce processus se réalise à partir de la plus grande unité, définissant le schéma comme point de départ (Field, 1999: 338). Il s'agit du processus d'utilisation des connaissances du monde réel.

Selon Anderson (1984: 248), la fonction de schéma est résumée en six parties :

1. fournir un échafaudage idéationnel pour l'assimilation des informations textuelles ;
2. faciliter l'attribution sélective de l'attention ;
3. faire une élaboration inférentielle ;
4. rechercher la mémoire par ordre ;
5. faciliter l'édition et la synthèse ;
6. encourager une reconstruction inférentielle.

Les connaissances du monde réel sont stockées soit dans des propositions, soit dans des schémas. Puis, les auditeurs peuvent augmenter les propositions ou les schémas existants avec de nouvelles informations au lieu d'être obligés de construire des structures de connaissances entièrement nouvelles. Dès lors, l'auditeur élabore de nouvelles informations en utilisant ou en associant des connaissances connues. Dans certains cas, les connaissances existantes se trouvent stockées sous forme de scripts, qui correspondent à des schémas spéciaux contenant des informations spécifiques dans des situations réelles. Elles peuvent

aussi être stockées dans des grammaires d'histoires, qui consistent en des schémas spéciaux représentant l'organisation du discours des fables, des histoires, des récits, etc. Les avantages de ces schémas sont qu'ils permettent à l'auditeur d'anticiper ce qui va se passer ensuite, de prédire les conclusions et de déduire le sens d'une partie du texte oral. Ainsi, les auditeurs qui emploient efficacement les connaissances schématiques utilisent un traitement « descendant » parce qu'ils s'appuient sur des informations en mémoire ou sur l'analyse de la signification du texte oral pour la compréhension. Il ne s'avère pas surprenant que le temps de réponse au niveau de la phrase pour le traitement syntaxique soit plus long que pour le traitement basé sur le sens (O'Malley *et al.*, 1989: 421).

L'étude de O'Malley *et al.* (1989) fait également état de résultats expérimentaux concernant ces processus de traitement de la compréhension orale utilisés par les apprenants. Concernant la différence entre les auditeurs efficaces et inefficaces, les auteurs indiquent que, en général, les auditeurs efficaces écoutaient de plus gros morceaux, ne reportant leur attention sur des mots individuels qu'en cas de défaillance de la compréhension. Les auditeurs inefficaces, quant à eux, abordaient l'écoute comme une tâche exigeant principalement une compréhension mot par mot. Ainsi, l'approche générale des apprenants les plus efficaces consistait à utiliser le traitement descendant et à ne recourir au traitement ascendant qu'en cas de besoin. En revanche, l'approche des auditeurs moins efficaces reposait systématiquement sur une stratégie ascendante. Les étudiants qui écoutaient efficacement déduisaient le sens des mots nouveaux, importants pour la compréhension du texte oral, en utilisant le contexte de la phrase ou du paragraphe dans lequel le mot non familier apparaissait (O'Malley *et al.*, 1989: 429). Par ailleurs, la mesure dans laquelle les auditeurs utilisent un processus ascendant ou descendant dépend du but de la compréhension orale, des caractéristiques de l'apprenant, telles que le niveau de compétence linguistique, et du contexte de l'événement de la compréhension orale. Un auditeur qui a besoin de vérifier un détail spécifique, par exemple, s'engage dans un traitement plus ascendant qu'un auditeur qui souhaite comprendre l'essentiel d'un texte (Vandergrift, 2007: 193).

Cependant, il a été rapporté que les apprenants se révèlent incapables de faire face à l'approche descendante dans la compréhension orale au niveau des phrases courtes. Kita (2005) a mené une expérience avec des apprenants japonais de l'anglais en utilisant dix exemples de phrases courtes en anglais et en les écoutant phrase par phrase. Les sujets ont d'abord écrit le sens des phrases courtes qu'ils

ont écoutées, ils ont ensuite réécouté l'audio et dicté les phrases. L'analyse des résultats a montré que, même avec des phrases anglaises de trois à quatre mots, il s'avérait extrêmement difficile de saisir le sens de la phrase en se basant uniquement sur la relation contextuelle si les sujets ne se trouvaient pas en mesure de reconnaître phonétiquement les mots importants de la phrase. En d'autres termes, cela montre qu'il semble difficile d'utiliser une stratégie descendante lorsque les apprenants rencontrent des mots inconnus. Dans certains cas, les participants ont été incapables de traiter le sens de la phrase en écoutant le vocabulaire en fragments, ce qui confirme la difficulté d'utiliser la stratégie descendante au niveau de la phrase courte.

Dès lors, il existe également une autre idée selon laquelle les modèles ascendant et descendant décrits ci-dessus fonctionnent simultanément et se complètent l'un l'autre (Field, 1999: 338). Les auditeurs s'efforcent à comprendre grâce à l'avantage des deux modèles (Cook, 1996: 78). Il s'agit du modèle interactif (figure 3).

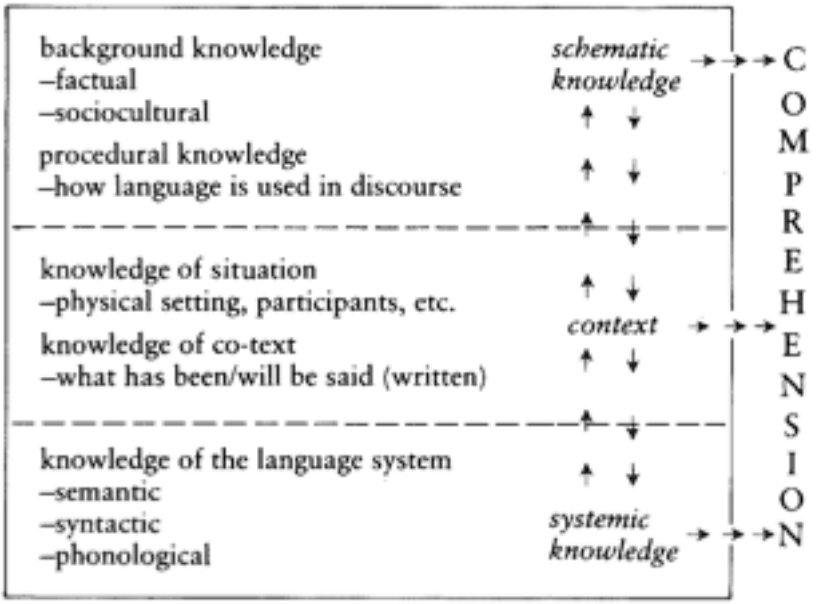

Figure 3 : Sources d'information dans la compréhension (Anderson & Lynch, 1988: 13)

En effet, il paraît souvent difficile d'établir la différence entre ce qui a été réellement dit et ce qui a été construit en intégrant les paroles prononcées à des connaissances et des expériences propres (Anderson & Lynch, 1988: 13). Par ailleurs, ce qui semble appréciable dans ce modèle est qu'il permet de développer

la compréhension d'une manière qui convient aux préférences et compétences individuelles (Flowerdew & Miller, 2005: 27). En particulier, en ce qui concerne les apprenants avancés, ils possèdent déjà une connaissance grammaticale suffisante de la langue étrangère en question. Ainsi, ils peuvent facilement mobiliser ces connaissances, tout en s'appuyant sur le sens. Le modèle interactif fonctionne donc bien pour eux (Cornaire, 1998: 47).

Outre les trois modèles de traitement des informations mentionnés jusqu'à présent, Gremmo & Holec (1990) proposent deux autres modèles, qui mettent particulièrement l'accent sur le sens et la forme : le modèle sémasiologique et le modèle onomasiologique. Dans le modèle sémasiologique, l'auditeur commence par isoler la chaîne phonique du message et identifier les sons qui la composent. Ensuite, il segmente les mots, groupes de mots et phrases que ces sons forment, et il leur associe un sens. Enfin, il construit la signification globale du message en combinant ces sens. En revanche, dans le modèle onomasiologique, l'auditeur formule d'abord des hypothèses sur le contenu du message en s'appuyant sur ses connaissances préalables ainsi que sur les informations extraites du message au fur et à mesure de son déroulement. Parallèlement, il élabore des hypothèses formelles basées sur sa connaissance des structures des signifiants de la langue dans laquelle le message est encodé. Ensuite, il vérifie ces hypothèses (*ibid.*, 31-33). Quel que soit le modèle utilisé, la mémoire joue un rôle étroitement lié au traitement de l'information.

1.2.3. Compréhension orale et mémoire

En décrivant la compréhension orale jusqu'à présent, nous avons mentionné plusieurs fois la mémoire à court terme et la mémoire à long terme. Le processus de traitement cognitif de la compréhension orale suppose en effet la fonction de mémoire. La compréhension orale requiert la conservation temporelle puisqu'il est impossible de traiter des informations qui seraient oubliées aussitôt après les avoir écoutées. En plus de cela, sans connaissances préalables, il s'avère impossible de comprendre le sens de ce qui est écouté (Fukuda, 2002: 370). Dans cette section, nous décrirons la mémoire à court terme, la mémoire à long terme et une autre

forme mémoire importante, la mémoire de travail, qui sont toutes bien concernées par la compréhension orale.

Tout d'abord, la mémoire à court terme correspond littéralement à la mémoire qui est conservée pendant de courtes périodes. Dans la mémoire marginale à court terme, seuls 7 ± 2 chunks sont retenus (Miller, 1956). Cook (1977) a mené un test sur la mémoire à court terme en demandant à des apprenants de l'anglais de mémoriser des chiffres à l'oral. Les apprenants ont été divisés en débutants et avancés avant d'effectuer le test. Les résultats ont montré que les débutants possédaient une capacité maximale moyenne de 5,9 chiffres et le groupe avancé de 6,7. Les résultats des locuteurs natifs s'élevaient à 8 tandis que ceux des non-natifs se sont révélés toujours légèrement inférieurs à ceux des locuteurs natifs, même à des stades avancés. Cependant, cette étude montre qu'une meilleure maîtrise de la langue peut étendre la portée de la mémoire à court terme dans une langue étrangère (Cook, 1996: 66-67). Par ailleurs, la mémoire à court terme se perd avec le temps et l'arrivée de nouvelles informations, tandis que la mémoire qui est remémorée à plusieurs reprises devient la mémoire à long terme et peut être manipulée.

Ainsi, la mémoire à long terme, par opposition à celle à court terme, désigne la mémoire qui est conservée pendant une longue période. Les schémas y sont stockés et sont activés lorsque cela s'avère nécessaire afin de récupérer les éléments requis. Par conséquent, même si le vocabulaire qui n'est pas stocké dans la mémoire à long terme est traité dans la phase de perception du modèle d'Anderson (1980), décrit dans la section 1.2.1., il ne peut pas atteindre la phase d'utilisation des connaissances existantes, et ne peut pas être compris (Fukuda, 2002: 371).

Enfin, la mémoire de travail pour les tâches cognitives complexes constitue une mémoire active qui permet la conservation temporelle et le traitement de calcul lors de la tâche cognitive avancée, à savoir la compréhension linguistique, le calcul et l'inférence (Baddeley & Hitch, 1974). En ce qui concerne le mécanisme complexe de la mémoire de travail, il se compose de trois systèmes subordonnés et d'un système de mémoire qui contrôle les fonctions centrales. Le premier système subordonné, appelé « boucle phonologique », contient le concept de la mémoire à court terme ; le deuxième système, appelé « calepin visuospatial », retient et traite les informations visuelles et spatiales ; le troisième système constitue un nouveau système de buffer sémantique proposé par Baddeley (2000). Le système qui contrôle ces systèmes subordonnés, appelé « administrateur

central », leur distribue, quant à lui, les informations d'entrée. Il extrait également les informations de la mémoire à long terme et les intègre avec les nouvelles informations qui viennent juste d'arriver du monde extérieur afin de travailler (Baddeley, 2003: 191-203).

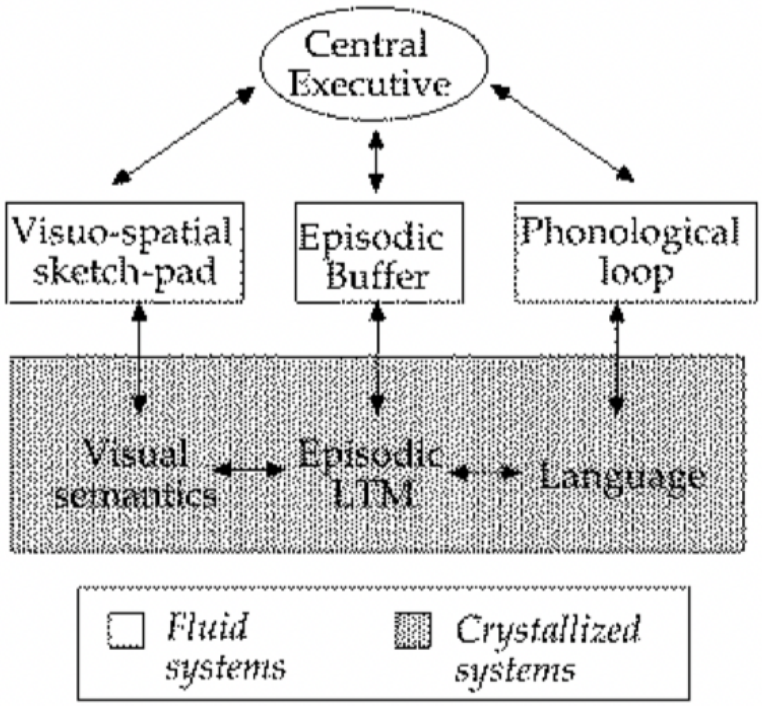

Figure 4 : Système de la mémoire de travail (Baddeley, 2003: 196)

Parmi les systèmes de la mémoire de travail, la boucle phonologique concerne le plus le langage (Sugai, 2020: 4). Elle se compose d'un stock phonologique et d'un processus de récapitulation articulatoire. Bien que le stock phonologique reçoive directement des informations auditives, une répétition s'avère nécessaire dans le processus de récapitulation articulatoire afin de les retenir longtemps (Baddeley, 2003: 191).

Ensuite, le calepin visuospatial incarne un autre système subordonné de la mémoire de travail. Il retient temporairement et traite des images visuelles non verbales et des informations spatiales (*ibid.*, 200). Par ailleurs, il est positionné dans le cerveau droit, qui contrôle le non-verbal, par opposition à la boucle phonologique, positionnée dans le cerveau gauche, qui contrôle le langage (Onaha, 2014: 11).

Baddeley (2000) a proposé un buffer sémantique comme quatrième système dans lequel les informations peuvent être temporairement conservées, sans faire de distinction entre les informations auditives et visuelles. Par conséquent, il semble avoir pour fonction de retenir simultanément les informations de la boucle phonologique et du calepin visuospatial et de les intégrer à la mémoire à long terme. En d'autres termes, la mémoire de travail joue un rôle dans l'établissement d'un lien entre les nouvelles informations, qui entrent dans la mémoire à court terme, et la mémoire à long terme ou dans la mise à jour de la mémoire à long terme.

L'administrateur central peut distribuer les informations d'entrée aux trois systèmes subordonnés. De plus, il extrait les informations de la mémoire à long terme et intègre de nouvelles informations provenant du monde extérieur qui viennent d'entrer dans la mémoire à court terme afin d'effectuer des activités cognitives de haut niveau telles que la planification, la pensée, la supposition et le jugement dans un temps limité. Il joue également un rôle dans la détermination de l'activité de l'ensemble de la mémoire de travail (Onaha, 2014: 12).

Une étude examinant l'influence de la mémoire de travail sur les compétences de la compréhension orale des apprenants a été menée par Satori (2021). Elle a cherché à clarifier si la capacité de la mémoire de travail influençait la compréhension orale en L2 indépendamment des connaissances linguistiques et des compétences de traitement en L2, et si ces effets opéraient différemment selon les niveaux de compétence en L2. Pour ce faire, les participants étaient composés de 150 apprenants japonais de langue anglaise et étaient classés en trois niveaux de compétence sur la base du test de compréhension orale du TOEIC[3]. Ils ont passé des tests d'empan numérique[4], un test de dictée, un test auditif de vocabulaire, un test auditif de grammaire et un test de sensibilité à l'accent. Seuls les tests de l'empan numérique ont été effectués pour la langue maternelle (L1), le japonais, et la L2.

L'analyse des résultats a révélé que la capacité de la mémoire de travail était plus fortement associée, en L2, à la compréhension orale, au traitement perceptif et au traitement syntaxique dans le groupe de faible compétence par rapport au groupe de haute compétence. Cela signifie que plus la maîtrise de la L2 s'avère

[3] Test of English for International Communication.
[4] Les participants écoutent les cinq à neuf chiffres et reproduisent la séquence de chiffres sur la feuille de réponse immédiatement.

faible, plus l'impact de la mémoire de travail semble important. Il est également apparu que la capacité de la mémoire de travail en L1 représente une variance significative dans la compréhension orale en L2. Ainsi, les résultats suggèrent que la capacité de la mémoire de travail en L1 pourrait être liée à la compréhension orale en L2 indépendamment de la connaissance linguistique en L2 dans le groupe de faible compétence. Il ne fait presque aucun doute que la mémoire décrite jusqu'ici est fortement impliquée dans le traitement du langage.

1.3. Composantes de la compréhension orale

Il est important de noter les éléments nécessaires au processus de compréhension orale que nous venons d'examiner. Cela facilite la compréhension des difficultés des apprenants en matière de compréhension orale. En rappelant le processus de la compréhension orale selon une optique psychologique, Rost (1991) explique les composants répartis en sept catégories pour trois compétences : la compétence perceptive, analytique et synthétique (*ibid.*, 3-4). Tout d'abord, la compétence perceptive contient les composants de discrimination entre les sons (1) et de reconnaissance des mots (2). Ensuite, la compétence analytique comprend l'identification de groupes grammaticaux de mots (3) et l'identification d'unités pratiques (4). Enfin, les capacités à relier des informations linguistiques à des informations paralinguistiques (intonation et accent) et à celles non linguistiques (gestes, objectif pertinent dans la situation) (5), l'utilisation de connaissances et du contexte (6) et le rappel de mots et idées importants constituent la compétence synthétique (7).

Bachman & Palmer (1996) expliquent plus en détail les composants nécessaires. Ils définissent la connaissance linguistique comme un domaine d'information en mémoire qui peut être utilisé par les stratégies métacognitives afin de créer et interpréter un discours sur l'utilisation du langage (*ibid.*, 67). En outre, ils expliquent les composants nécessaires en divisant la connaissance linguistique par la connaissance organisationnelle et pragmatique (*ibid.*, 67-70). Ces deux connaissances possèdent des sous-connaissances.

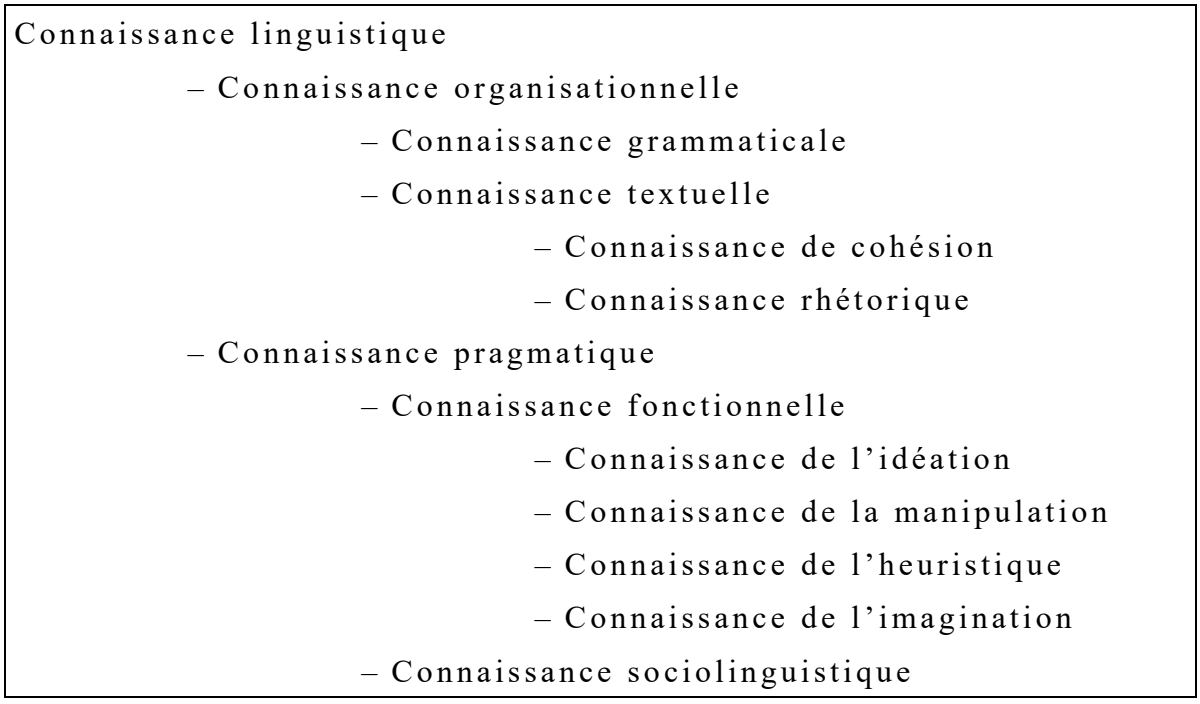

Figure 5 : Classification de Bachman & Palmer (1996)

Tout d'abord, la connaissance organisationnelle est impliquée dans le contrôle de la structure formelle du langage afin de comprendre des énoncés ou des phrases grammaticalement acceptables, de les organiser et ainsi de former des textes oraux. Elle comporte aussi deux sous-connaissances : la connaissance grammaticale et la connaissance textuelle. La première est impliquée dans la compréhension d'énoncés ou de phrases formellement exacts. Elle comprend la connaissance du vocabulaire, de la syntaxe et de la phonologie. La seconde est impliquée dans la compréhension des textes, qui constituent des unités de langue parlées qui consistent en deux ou plusieurs énoncés ou phrases. Cette connaissance textuelle comprend également deux sous-connaissances : la connaissance de cohésion et la connaissance rhétorique. La première est impliquée dans la compréhension des relations explicitement marquées entre les phrases parmi les énoncés dans les conversations tandis que la seconde l'est dans la compréhension du développement organisationnel au cours des conversations.

La connaissance pragmatique, quant à elle, permet de créer ou d'interpréter un discours en reliant des énoncés ou des phrases et des textes à leur signification, aux intentions des utilisateurs de la langue et aux caractéristiques pertinentes du contexte d'utilisation de la langue. De même, elle possède deux sous-connaissances : la connaissance fonctionnelle et la connaissance sociolinguistique. La première permet d'interpréter les relations entre des énoncés ou des phrases et

des textes et les intentions des utilisateurs du langage tandis que la seconde permet d'interpréter un langage adapté à un contexte particulier d'utilisation des langues. La connaissance sociolinguistique comprend la connaissance des conventions qui déterminent l'utilisation appropriée des dialectes ou des variétés, des registres, naturels ou idiomatiques, des expressions, des références culturelles et des figures de style. Par ailleurs, la connaissance fonctionnelle est divisée en quatre sous-connaissances : la connaissance de l'idéation, de la manipulation, de l'heuristique et de l'imagination.

Premièrement, la connaissance de l'idéation permet d'interpréter le sens selon son expérience du monde réel. Cette fonction comprend l'utilisation du langage de manière à interpréter des informations sur des idées, des connaissances ou des sentiments. Deuxièmement, la connaissance de la manipulation permet d'utiliser le langage afin d'affecter le monde qui nous entoure. Elle comprend la connaissance des éléments suivants :

- les fonctions instrumentales, qui sont exécutées de manière à amener d'autres personnes à faire des choses pour nous (les exemples incluent les demandes, suggestions, commandes et avertissements) ;
- les fonctions de réglementation, utilisées afin de contrôler ce que font les autres (par exemple, les règles, règlements et lois) ;
- les fonctions interpersonnelles, qui permettent d'établir, maintenir et changer les relations interpersonnelles (les exemples incluent les salutations et les congés, les compliments, les insultes et les excuses).

Troisièmement, la connaissance de l'heuristique permet d'utiliser le langage pour étendre notre connaissance du monde qui nous entoure, par exemple lors de l'utilisation du langage pour l'enseignement et l'apprentissage, la résolution de problèmes et la conservation d'informations. Quatrièmement, la connaissance de l'imagination permet d'utiliser le langage afin de créer un monde imaginaire ou étendre le monde qui nous entoure à des fins humoristiques ou esthétiques (les exemples incluent les blagues et l'utilisation du langage figuratif et de la poésie).

En résumé, afin de réaliser une compréhension orale, il faut d'abord disposer d'informations qui arrivent sous forme de sons, non seulement du vocabulaire et de la syntaxe, mais aussi de la connaissance pragmatique. Cela permet de mieux comprendre le contenu et l'intention du discours. Les composants mentionnés ci-

dessus sont conformes aux descriptions du modèle théorique et du modèle de traitement des informations.

1.4. Difficultés de la compréhension orale pour les apprenants de langue étrangère

Dans cette section, nous explorerons les points qui empêchent le traitement de la compréhension orale et l'amélioration de cette compétence chez les apprenants. Lorsqu'ils écoutent, ces derniers sont confrontés à un certain nombre de problèmes.

En ce qui concerne le processus de compréhension orale décrit à la section 1.2., les auditeurs d'une langue étrangère peuvent éprouver des difficultés à comprendre une langue parlée à un rythme conversationnel typique par des locuteurs natifs. Cela se révèle particulièrement vrai s'ils ne connaissent pas les règles de segmentation. Même s'ils peuvent comprendre des mots individuels, lorsque ces derniers sont entendus séparément, la compréhension globale reste difficile. Par ailleurs, ils doivent concaténer ou combiner les segments analysés au cours de la compréhension afin de comprendre les textes en langue étrangère. Cette tâche impose une charge supplémentaire pour la mémoire à court terme. Or, cette dernière peut déjà être surchargée d'éléments non codés de la nouvelle langue (O'Malley *et al.*, 1989: 420-421).

De plus, retenir des informations peut être plus difficile pour la compréhension orale chez les apprenants que chez les natifs. Oishi & Kinoshita (2008: 151), qui ont mené des recherches dans le domaine des neurosciences, rapportent que le débit sanguin cérébral s'avère plus important lors d'une tâche de compréhension orale en langue étrangère que lors d'une tâche en langue maternelle. Même lorsque les participants disposent de connaissances sur le texte en langue étrangère, le niveau de l'activité cérébrale se révèle plus grand que lors de l'écoute d'un texte sans connaissances préalables, mais en langue maternelle. Ainsi, les auteurs suggèrent que la compréhension orale en langue étrangère demande plus d'attention en raison de cette difficulté.

Nous nous interrogeons sur les difficultés pour les apprenants en ce qui concerne les composantes de la compréhension orale présentées à la section 1.3. Nous nous fixons particulièrement sur celles des auditeurs, alors que les difficultés fondamentales, telles que la manque de clarté et de structure logique à l'oral, bloquent également la compréhension orale. En synthèse des recherches portant sur des difficultés pour les apprenants, les difficultés liées à la connaissance grammaticale et à la connaissance pragmatique indiquées à la section 1.3.

empêchent la compréhension orale des apprenants. Plus précisément, il s'agit des éléments suivants :

- l'aspect phonétique (connaissance grammaticale de la figure 5) ;
- l'aspect grammatical (connaissance grammaticale de la figure 5) ;
- le manque de connaissances préalables (connaissance pragmatique de la figure 5) ;
- le facteur individuel des auditeurs.

Toutefois, il y a certains aspects qui interagissent.

Concernant l'aspect phonétique, les environnements phonétiques, caractérisés par les facteurs extérieurs tels que le bruit, le son flou et le volume, peuvent impacter la compréhension (Kobayashi, 2008: 3). Il existe également des problèmes à propos du manque de compétence permettant de distinguer le changement sonore. Cette difficulté correspond, par exemple, à l'incapacité de reconnaître les délimiteurs de mots, propositions ou bien phrases, et l'incapacité de comprendre les sons raccourcis et assimilés, et les sons omis et faiblement prononcés (*ibid.*, 3 ; Ikeda, 2003: 74). À cause de ces difficultés, même les mots que les apprenants peuvent comprendre lors de la lecture sont perçus comme des sons inconnus, ce qui les rend difficile à comprendre. Par ailleurs, les problèmes liés au débit et aux pauses s'ajoutent à cela.

En réalité, Griffiths (1992) a révélé une relation entre le débit de parole et la compréhension orale. L'étude est basée sur un total de neuf ensembles de données issus de trois histoires en anglais, chacune enregistrée avec trois débits de parole différents. Au total, 24 apprenants adultes de l'anglais, répartis en groupes, ont été invités à écouter une fois chacune des trois histoires. Puis, quinze questions de type « vrai ou faux » ont été posées pour chaque histoire ; la capacité de compréhension orale des apprenants a été déterminée à partir des résultats des réponses aux questions. La moyenne de réponses correctes aux questions a été examinée afin de déterminer une éventuelle différence statistiquement significative entre la moyenne des réponses correctes aux questions à différents débits. Les résultats ont montré que les scores obtenus pour le débit lent sont significativement différents des scores obtenus pour les débit moyen et rapide.

En effet, en augmentant le débit de parole, le nombre de pauses devient nécessairement plus faible et la durée des pauses plus courte. Dès lors, le

traitement devient plus difficile car la pause représente un temps qui permet de traiter collectivement des informations qui sont entrées jusqu'à un moment donné (Ikeda, 2003: 74).

Blau (1990) a comparé l'effet du débit de parole et des pauses. Dans chacun des trois monologues, les données sont disponibles pour trois débits de parole différents : débit normal, débit lent et débit pour lequel des pauses de trois secondes ont été insérées aux limites des phrases, des clauses et d'expressions sélectionnées. L'étude a été menée sur des échantillons de deux populations : 36 étudiants d'anglais d'une université polonaise et 70 d'une université de Porto Rico. Les étudiants ont été répartis en trois groupes pour l'expérience. La compréhension de chaque monologue a été mesurée à l'aide de questions posées immédiatement après leur écoute. Les participants ont également été invités à indiquer le pourcentage des monologues qu'ils pensaient avoir compris.

Ensuite, les données ont été analysées à l'aide d'une approche de régression de l'analyse de la covariance. Les résultats ont montré que la compréhension se révélait meilleure lors de l'écoute de la version comprenant des pauses. En outre, ces scores significativement plus élevés pour cette version correspondent à l'évaluation en pourcentage réalisée par les participants à propos de leur perception de leur compréhension.

Par ailleurs, il convient de questionner la difficulté grammaticale telle que le manque de connaissance de vocabulaire et syntaxe. Cette difficulté risque d'empêcher l'interprétation correcte du sens en coupant les groupes de sens (Ikeda, 2003: 74). Dans le cas particulier de la syntaxe, il s'avère nécessaire de comprendre non seulement la grammaire écrite, mais aussi les caractéristiques de la grammaire parlée dans la conversation quotidienne (Kobayashi, 2008: 5). Or, l'oral présente des caractéristiques différentes de l'écrit, puisque le langage parlé se révèle souvent moins grammaticalement complet que l'écrit. En lien avec cela, on observe également des hésitations.

Voss (1979) a identifié l'importance des phénomènes d'hésitation comme source d'erreurs de perception chez les apprenants de l'anglais. Les trois phénomènes d'hésitation étudiés dans le cadre de cette étude sont les suivants :

1. les répétitions qui couvrent toutes les répétitions sémantiquement non significatives ;
2. les faux départs, qui constituent des énoncés incomplets ou auto-interrompus pouvant être retracés (c'est-à-dire corrigés) ou non retracés ;
3. les pauses remplies qui couvrent toutes les occurrences de dispositifs d'hésitation tels que [ɜ, ə, m], etc.

Les participants de l'étude étaient 22 étudiants allemands. La tâche consistait à transcrire le plus fidèlement possible une partie d'interview spontanée. Les participants pouvaient écouter l'audio autant de fois qu'ils le souhaitaient. L'analyse de leurs transcriptions a montré des erreurs de perception dans le contexte des phénomènes d'hésitation :

– les cas dans lesquels des répétitions ou des pauses remplies ont été prises pour des mots ou parties de mot ;
– les cas dans lesquels des mots ou parties de mot ont été pris pour des répétitions ou des pauses remplies et ont donc été ignorés.

Par conséquent, l'auteur a précisé qu'environ un tiers des erreurs perceptives se trouvait lié au phénomène des hésitations. Les malentendus ont été dus soit à une mauvaise interprétation des phénomènes d'hésitation en tant que mots ou parties de mot, soit à une mauvaise interprétation de mots ou parties de mot en tant qu'hésitations à ne pas consigner par écrit.

Quant au manque de connaissances, d'une manière générale, si les auditeurs rencontrent la difficulté sur la compréhension, ils cherchent à compléter naturellement la compréhension à l'aide de leur connaissance préalable et leur sens commun. Cependant, cela ne s'avère pas si facile pour les apprenants qui possèdent une culture différente (O'ki *et al.*, 2016: 520 ; Ikeda, 2003: 75).

La dernière difficulté concerne les facteurs individuels des auditeurs. Tout d'abord, dans le cadre de la communication quotidienne, la mémoire paraît assez sollicitée afin de retenir des termes utilisés par un locuteur et le contenu dans une certaine mesure. Cependant, sauf dans le cas d'une conférence ou d'une réunion, il est rare d'écouter en prenant des notes, et mémoriser ce que les gens disent semble d'autant plus difficile dans le cas de langue étrangère (Yonezu & Ogura, 2017: 18). Par ailleurs, le facteur mental de la part des apprenants affecte aussi la

compréhension. Lors de la compréhension orale en cours ou encore en avant, les auditeurs peuvent être stressés et s'ils rencontrent des parties incompréhensibles, leur anxiété se révèle élevée et cela peut bloquer leur compréhension (*ibid.*, 18-19 ; Ikeda, 2003: 75). Or, certains apprenants éprouvent originairement des difficultés à trier les informations importantes et à saisir l'ensemble ou le récapitulatif de choses. Cette compétence peut découler du niveau de compréhension en langue maternelle, et si cette compétence ne se situe pas au bon niveau dans la langue maternelle, il paraît assez logiquement plus difficile pour ces apprenants d'avoir une bonne compétence en langue étrangère (*ibid.*, 75). Les difficultés traitées ci-dessus sont considérées comme les difficultés principales au sein de la compréhension orale chez les apprenants de langue étrangère.

1.5. Méthodes et approches d'enseignement et d'apprentissage

Il convient de s'interroger sur la manière dont la compréhension orale en langue étrangère, constituant un processus complexe et posant de nombreuses difficultés, a été enseignée en classe. Pour ce faire, les approches de l'enseignement de la compréhension orale seront présentées à la section 1.5.1. Puis, la section suivante 1.5.2. donnera un aperçu des méthodes d'enseignement spécifiques utilisées aujourd'hui tandis que la section 1.5.3. se penchera sur les supports d'enseignement, et en particulier sur l'importance de l'utilisation de supports authentiques pour l'enseignement et l'apprentissage de la compréhension orale.

1.5.1. Méthodes et approches d'enseignement et d'apprentissage de la compréhension orale

En ce qui concerne les approches de l'enseignement de la compréhension orale, Flowerdew & Miller (2005) présente neuf catégories : la méthode grammaire-traduction, la méthode directe, l'approche grammaire, la méthode audio-orale, l'approche par éléments distincts, l'approche communicative, l'approche par tâches, l'approche stratégie de l'apprenant et l'approche intégrée. En ce qui concerne les méthodes et les approches d'enseignement du français, Nakamura & Hasegawa (1995) en privilégient cinq : la méthode grammaire-traduction, la méthode directe, la méthode audio-orale, la méthode structuro-globale audiovisuelle et l'approche communicative. Après l'approche communicative, l'approche actionnelle se développe (Conseil de l'Europe, 2001 ; 15-16).

Tout d'abord, la méthode grammaire-traduction, en vigueur des années 1840 aux années 1940, mettait l'accent sur la lecture et l'écriture et n'incluait pas l'apprentissage par la compréhension orale. La raison en était que les élèves apprenaient des langues « mortes » (le latin et le grec), des langues qu'ils n'auraient pas l'occasion d'écouter (Flowerdew & Miller, 2005: 4). Au milieu et à la fin du XIX^{e} siècle, une opposition à la méthode grammaire-traduction s'est progressivement développée dans plusieurs pays européens. Ce mouvement de réforme a jeté les bases du développement de nouvelles méthodes d'enseignement des langues (Richards & Rogers, 1986: 5).

Ainsi, cette section retracera l'évolution de l'enseignement et de l'apprentissage à la compréhension orale de la langue française selon cinq méthodes et approches depuis la méthode directe : la méthode directe, la méthode audio-orale, la méthode structuro-globale audiovisuelle, l'approche communicative et l'approche actionnelle.

1.5.1.1. Méthode directe

La méthode directe est née en réaction à la méthode grammaire-traduction (Flowerdew & Miller, 2005: 4-5). Les principes de cette méthode sont résumés comme suit (Richards & Rogers, 1986: 9-10) :

1. L'enseignement en classe est dispensé exclusivement dans la langue cible.
2. Seul le vocabulaire et les phrases de tous les jours sont enseignés.
3. Les compétences en communication orale ont été développées dans le cadre d'une progression soigneusement graduée, organisée autour d'échanges de questions-réponses entre les enseignants et les apprenants dans de petites classes intensives.
4. La grammaire est enseignée de manière inductive.
5. Les nouveaux points d'enseignement sont introduits oralement.
6. Le vocabulaire concret est enseigné au moyen de démonstrations, d'objets et d'images ; le vocabulaire abstrait est enseigné par association d'idées.
7. La production orale et la compréhension orale ont toutes deux étés enseignées.
8. La prononciation et la grammaire correctes se trouvent mises en avant.

Comme indiqué ci-dessus, la langue cible constitue la seule langue utilisée en classe. Cependant, il existe des inconvénients à cet égard.

Tout d'abord, la méthode directe exige des enseignants qu'ils soient des locuteurs natifs ou qu'ils aient une aisance comparable à celle d'un natif dans la langue étrangère. Dès lors, cette méthode dépend largement des compétences de l'enseignant, plutôt que d'un manuel. Or, tous les enseignants ne s'avèrent pas suffisamment compétents dans la langue étrangère de manière à adhérer aux principes de la méthode. Ainsi, les critiques ont souligné que l'adhésion stricte aux principes de la méthode directe se révélait souvent contre-productive, car les

enseignants doivent se donner beaucoup de mal afin d'éviter d'utiliser la langue maternelle, alors que parfois une simple explication brève dans la langue maternelle de l'apprenant aurait été plus efficace pour la compréhension (Richards & Rogers, 1986: 10-11).

De plus, bien que la langue cible soit utilisée à toutes fins dans la classe, il n'existe aucune tentative systématique d'enseigner la compréhension orale ou de développer ses stratégies chez les apprenants. L'enseignant part du principe que les apprenants peuvent entendre ce qui est dit et que la compréhension suivrait plus tard (Flowerdew & Miller, 2005: 6).

1.5.1.2. Méthode audio-orale

L'émergence de la méthode audio-orale résulte de l'attention accrue portée à l'enseignement des langues étrangères aux États-Unis vers la fin des années 1950 (Richards & Rogers, 1986: 47-48). Il s'agit de programmes linguistiques des forces de défense américaines pendant et après la Seconde Guerre mondiale (Flowerdew & Miller, 2005: 8). Cette méthode a atteint sa période d'utilisation la plus répandue dans les années 1960 (Richards & Rogers, 1986: 59).

La méthodologie centrale de la méthode audio-orale se caractérise comme suit (Richards & Rogers, 1986: 51) :

1. L'apprentissage d'une langue étrangère constitue essentiellement un processus de formation mécanique d'habitudes.
2. Les compétences linguistiques sont apprises plus efficacement si les éléments à apprendre dans la langue cible sont présentés sous forme orale avant d'être vus sous forme écrite.
3. L'analogie constitue une meilleure base pour l'apprentissage des langues que l'analyse.

Concernant la compréhension orale, cette méthode met l'accent sur l'écoute de la prononciation et des formes grammaticales, ensuite sur l'imitation de ces formes à l'aide d'exercices (Flowerdew & Miller, 2005: 8). L'inconvénient de cette méthode est qu'il s'avère difficile de voir le dialogue s'étendre au-delà de ce modèle en deux phases. Les apprenants se sont souvent révélés incapables de transférer les compétences acquises grâce à la méthode audio-orale à la

communication réelle en dehors de la salle de classe. En outre, beaucoup ont trouvé l'expérience de l'étude par l'intermédiaire de procédures audio-orales ennuyeuses et insatisfaisantes (Richards & Rogers, 1986: 59). De plus, puisqu'il n'existe pas de tentative d'enseigner le lexique ou de contextualiser les phrases, le développement des compétences de la compréhension orale ne constitue pas l'objectif principal de cette méthode contrairement à la manipulation des structures (Flowerdew & Miller, 2005: 10).

1.5.1.3. Méthode structuro-globale audiovisuelle

Cette méthode a été développée en France à la fin des années 1950, intégrant l'image et le son dans le processus d'enseignement et d'apprentissage des langues étrangères (Ivan, 2006: 16). Les similitudes de la méthode structuro-globale audiovisuelle avec la méthode audio-orale, présentée dans la section précédente, résident dans l'apprentissage de la langue parlée par le dialogue, la pratique répétitive et la non-utilisation de la langue maternelle (Nakamura & Hasegawa, 1995: 154). À l'inverse, trois différences peuvent être identifiées (*ibid.*, 155) :

1. Les moyens visuels s'avèrent les plus efficaces, notamment parce que la langue parlée est perçue par les oreilles et les yeux.
2. La langue parlée dispose pour composantes principales de la parole, de l'intonation, du sens et de la situation. Dès lors, ces éléments ne doivent pas être séparés, mais constamment appris ensemble dans leur structure complète.
3. L'objectif ne consiste pas à automatiser la langue en pratiquant des modèles de phrases, mais à être capable de la réutiliser dans des situations différentes.

Dans cette méthode, l'accent a été mis sur les activités de compréhension orale ainsi que sur les moyens visuels. Les compétences sont développées de manière progressive : écouter et regarder, parler, ensuite lire, et enfin écrire (Ivan, 2006: 17). Les enseignants ont donc essayé d'utiliser l'image et le son pour présenter le dialogue dans une situation et faire comprendre le sens aux apprenants sans utiliser leur langue maternelle (Nakamura & Hasegawa, 1995: 155). Cependant, le dialogue restait plat et les apprenants s'avéraient incapables de faire

face à des dialogues de la vie réelle une fois qu'ils quittaient la situation du matériel (*ibid.*, 157).

1.5.1.4. Approche communicative

L'approche communicative a été introduite dans les années 1970. Elle s'intéresse à ce que les gens font avec le langage et à la manière dont ils réagissent à ce qu'ils entendent (Flowerdew & Miller, 2005: 12). Dans cette approche, les caractéristiques de la vision communicative de la langue comptent les éléments suivant (Richards & Rogers, 1986: 71) :

1. La langue constitue un système d'expression du sens.
2. Les fonctions premières du langage comptent l'interaction et la communication.
3. La structure de la langue reflète ses utilisations fonctionnelles et communicatives.
4. Les unités primaires de la langue ne comprennent pas simplement ses caractéristiques grammaticales et structurelles, mais également des catégories de signification fonctionnelle et communicative telles qu'elles sont illustrées dans le discours.

Pour cette approche, la langue est apprise en demandant aux apprenants d'écouter des informations telles que les prix et les emplacements de plusieurs restaurants, par exemple, dans le but de discuter des restaurants qu'ils souhaitent visiter avec des amis. Il ne suffit plus de maîtriser uniquement les aspects linguistiques (sons, structures, lexique, etc.) d'une langue étrangère afin de communiquer efficacement ; il devient également nécessaire d'en connaître les règles d'emploi (Cornaire, 1998: 21).

1.5.1.5. Approche actionnelle

La perspective privilégiée ici est, de manière générale, de type actionnel. Elle considère avant tout l'usager et l'apprenant d'une langue comme un acteur social, chargé d'accomplir des tâches (qui ne sont pas exclusivement langagières) dans des circonstances et un environnement donnés, au sein d'un domaine d'action

particulier. Les tâches existent dans la mesure où l'action constitue le fait des sujets. Ces derniers mobilisent stratégiquement les compétences dont ils disposent afin d'atteindre un objectif déterminé. La perspective actionnelle prend ainsi en compte non seulement les ressources cognitives, mais aussi les ressources affectives, volitives et l'ensemble des capacités que possède et met en œuvre l'acteur social (Conseil de l'Europe, 2001: 15).

Dans le cadre de l'approche actionnelle, les apprenants sont invités à écouter ce qui est décrit comme des situations « authentiques » et à « faire quelque chose » avec les informations. Il peut s'agir, par exemple, de compléter un diagramme ou un graphique, de remplir un tableau ou de faire un dessin. L'information se trouve généralement transférée d'un texte oral à une forme graphique. Comme il s'agit de textes authentiques (généralement semi-scénarisés), les apprenants doivent faire face à une langue parlée à un débit normal et à des caractéristiques telles que les accents, les hésitations, les mots de remplissage et les ellipses (Flowerdew & Miller, 2005: 14).

En résumé, on peut constater qu'à l'origine, les méthodes d'enseignement des langues ne reconnaissaient pas la nécessité d'enseigner la compréhension orale. Cependant, les approches ultérieures ont utilisé une variété de techniques permettant de développer des compétences de la compréhension orale spécifiques ou générales. De tels changements reflètent trois évolutions dans la manière dont la compréhension orale est considérée (Field, 1998: 110-11). Premièrement, un changement de perspective s'est produit, de sorte que la compréhension orale en tant que compétence prime désormais sur les détails du contenu linguistique. Deuxièmement, il a semblé souhaitable de relier la nature de la compréhension orale pratiquée en classe au type de compréhension orale qui a lieu dans la vie réelle. Cela se traduit par la mise en place d'un cadre contextuel, la reconnaissance de l'importance de l'inférence du sens des mots nouveaux, l'utilisation d'enregistrements d'origine authentique, l'inclusion de matériel présentant des caractéristiques de conversation et l'utilisation de tâches simulées plutôt que d'exercices formels. Troisièmement, l'importance de motiver et d'orienter la compréhension orale a été observée : l'auditeur est encouragé à développer des attentes quant à ce qu'il écoute dans le texte, ensuite à les vérifier par rapport à ce qui est réellement dit. Lors de la préparation des questions et des tâches, il

devient essentiel que les apprenants connaissent dès le départ l'objectif de l'exercice de compréhension orale et qu'ils n'aient pas à faire appel à leur mémoire.

En appliquant les cinq approches décrites ci-dessus à ces changements, les premier et deuxième développements se trouvent dans l'approche communicative. En ce qui concerne plus particulièrement le premier développement, l'intérêt pour la compréhension orale s'est manifesté avant cela, mais l'enseignement pratique de la compréhension orale ne s'est réellement ancré dans les pratiques des enseignants que dans le cadre de l'approche communicative. La troisième évolution se situe après l'approche communicative. Cela signifie que cette dernière joue un rôle très important dans l'enseignement et l'apprentissage à la compréhension orale. En effet, l'accent a été mis sur une approche communicative et la tendance croissante consistait à incorporer la compréhension orale comme une composante essentielle. Des manuels concernant la compréhension orale ainsi que des cassettes audio ont été développés grâce aux avancées techniques, ce qui a entraîné un intérêt croissant pour la compréhension orale (Morley, 1990: 321).

1.5.2. Méthodes d'enseignement spécifiques utilisées aujourd'hui

Avec le développement des technologies de l'information, l'environnement d'apprentissage a également changé : des laboratoires de langues ont été mis en place et l'utilisation des ordinateurs s'est généralisée. Parallèlement, les méthodes d'apprentissage et d'enseignement de la compréhension orale ont également évolué. Malgré les différentes approches concernant la compréhension orale, la dictée et le *shadowing* s'avèrent encore employés aujourd'hui comme principales méthodes d'enseignement et d'apprentissage. En effet, ces méthodes sont faciles à utiliser pour les apprenants.

Parmi les trois types de traitement de l'information dans la compréhension orale (ascendant, descendant et interactif), la dictée constitue une méthode d'enseignement de la compréhension orale qui utilise plutôt le traitement ascendant (O'ki & Izumi, 2015: 228). Elle permet de s'entraîner à écouter un discours et à le transcrire. Elle constitue également une activité à faible charge cognitive car les apprenants disposent de temps afin d'écrire les phrases qu'ils écoutent (*ibid.*, 230). Les avantages en matière d'enseignement comprennent la facilité de préparation (les enseignants peuvent préparer le matériel en remplissant simplement les blancs dans le texte), la facilité d'instruction en écrivant les

énoncés écoutés, et l'existence d'une réponse unique, de sorte que l'appariement peut être fait de manière fiable (Sugiura *et al.*, 2002).

Par ailleurs, le *shadowing* consiste en un autre moyen d'utiliser le traitement ascendant. Il se trouvait à l'origine utilisé comme une forme de formation de base pour les aspirants interprètes de conférence. Il désigne une méthode d'entraînement consistant à suivre une voix modèle émise par un enseignant ou un appareil tel qu'un lecteur CD et à la réciter à haute voix sans attendre la fin d'une seule phrase (Ogawa, 2018: 79). Il désigne également une activité cognitivement exigeante qui implique d'écouter et de parler en même temps (O'ki & Izumi, 2015: 230). L'efficacité de ces deux méthodes d'enseignement dans le but d'améliorer la compétence de la compréhension orale a été rapportée par Sugiura *et al.* (2002), Kiany & Shiramiry (2002) et Onaha (2014).

1.5.3. Avantages et inconvénients de l'utilisation de matériels authentiques

Alors que nous nous sommes concentrés sur les approches de l'enseignement de la compréhension orale et les méthodes d'enseignement, cette section fera référence au matériel pédagogique. Il s'avère d'abord nécessaire, avant de pratiquer la dictée ou le *shadowing* mentionnés dans la section précédente, de réfléchir d'abord au choix du matériel pédagogique à utiliser pour ces activités. Les approches actuelles de l'enseignement des langues, telles que l'enseignement communicatif des langues, attribuent un rôle primordial au matériel pédagogique. Ce dernier constitue un élément essentiel de la conception pédagogique et un moyen d'influencer la qualité de l'interaction en classe et de l'utilisation de la langue (Richards, 1993: 4). Par conséquent, le matériel pédagogique utilisé doit être sélectionné avec soin. À la section 1.5.1., nous avons constaté que l'approche communicative constitue un tournant important dans l'enseignement à la compréhension orale et que, dans le cadre de cette approche, l'utilisation de matériels « authentiques » et « tirés de la vie » en classe sont revendiqués (Richards & Rogers, 1986: 80).

Tout matériel qui n'a pas été spécifiquement produit à des fins d'enseignement des langues renvoie à l'authenticité dans l'enseignement des langues étrangères (Nunan, 1989: 54). L'authenticité des matériels sonores comprend les caractéristiques suivantes (Rost, 2002: 166) :

- un débit naturel : parler en courtes rafales, timing irrégulier ;
- un phénomène phonologique naturel : des pauses et une intonation naturelles, l'utilisation de la réduction, de l'assimilation et de l'élision ;
- un vocabulaire de haute fréquence en fonction des limites de la mémoire à court terme lors de la planification du discours parlé ;
- un langage familier, tel que des formules courtes, de l'argot courant, qui montrent une sensibilité à l'auditeur ;
- des hésitations, des faux départs, des autocorrections, en tant qu'indicateurs des processus cognitifs en temps réel du locuteur ;
- une orientation du discours vers un auditeur « en direct », y compris les pauses naturelles permettant à l'auditeur de fournir des informations en retour ou des réponses.

L'efficacité de cette authenticité, ou des matériels authentiques, dans l'apprentissage et l'enseignement de la compréhension orale a été rapportée (Field, 2002 ; Rost, 2002). Un avantage pour les matériels authentiques généralement cité est qu'ils fournissent des exemples d'hésitations, de faux départs, de pauses pleines et vides, etc., qui caractérisent le discours naturel. Cela aide les apprenants à se familiariser avec les cadences réelles de la langue cible (Field, 1998: 114). En outre, l'une des principales raisons de cette efficacité repose sur l'augmentation de la motivation des apprenants. En effet, le fait d'écouter un discours authentique représente une forte motivation pour la poursuite de l'apprentissage (Stempleski, 1987: 5). De plus, lorsque leur motivation s'avère supérieure, les apprenants se trouvent en mesure de mieux progresser au niveau de leur compétence de compréhension orale (Yoshimi, 2019: 116).

Cependant, il existe aussi des préoccupations concernant l'utilisation de matériels authentiques, qui sont liées aux niveaux de compétence des apprenants. Larsen-Freeman (2000: 132-133) note que pour les étudiants disposant d'une faible maîtrise de la langue cible, des matériels authentiques plus simples (par exemple, des prévisions utilisant des bulletins météorologiques), ou au moins des supports pédagogiques réalistes s'avèrent plus souhaitables. Ainsi, lors de l'introduction de supports authentiques, le niveau de compétence de l'apprenant et le contenu du support doivent être pris en compte.

1.6. Conclusion

Après un aperçu des modèles et des processus de la compréhension orale, ce chapitre s'est également penché sur l'enseignement de la compréhension orale dans les langues étrangères. En examinant les modèles théoriques, on a constaté que le modèle d'Anderson (1980) influence davantage la recherche sur la langue étrangère. Ce modèle explique que la compréhension orale s'effectue en trois étapes : la perception, l'analyse et l'utilisation. La mémoire joue un rôle majeur dans le traitement de l'information lors de la compréhension orale, avec les modèles ascendant et descendant, ainsi qu'avec le modèle interactif qui mêle les deux en même temps. De plus, pour une compréhension orale efficace, le vocabulaire, la syntaxe, mais aussi une connaissance pragmatique se révèlent nécessaires. Cependant, les apprenants de langues étrangères se heurtent à des obstacles dispersés dans la mise en pratique de chaque type de connaissance, ce qui entraîne des difficultés de la compréhension orale pour les apprenants.

Bien qu'aujourd'hui la compréhension orale soit considérée à juste titre comme l'une des compétences linguistiques utiles, il a fallu beaucoup de temps afin d'y parvenir. L'introduction de l'approche communicative, encore largement utilisée aujourd'hui, a constitué un tournant majeur concernant l'enseignement de la compréhension orale. Dans cette approche, la compréhension orale est enseignée d'une manière qui tient compte du fait qu'il ne s'agit pas simplement d'une compétence apprise en classe, mais également d'une compétence pratique ayant des liens étroits avec la vie réelle. Diverses méthodes et supports d'enseignement ont été testés et éprouvés de manière à garantir un enseignement et un apprentissage efficaces de la compréhension orale. Les supports authentiques s'avèrent particulièrement suggérés comme étant efficaces dans l'enseignement de la compréhension orale. Cependant, il existe des obstacles majeurs à leur utilisation efficace. L'un de ces obstacles réside dans la difficulté de choisir les supports appropriés. Afin de surmonter cette difficulté, l'étude de la *listenability,* qui sera décrite dans le chapitre suivant, peut être une solution efficace.

2. Facilité d'écoute

L'exploration des obstacles à la compréhension des matériels oraux constitue la pierre angulaire de la recherche sur la *listenability*, dont cette thèse traite. À l'origine, le domaine de la *listenability* partage les mêmes origines que celui de la lisibilité[5] : le souhait des éducateurs de classer le matériel de lecture pour les écoliers (Harwood, 1950: 11). Le terme *listenability* a ensuite été utilisé afin de se référer à la facilité ou la difficulté de compréhension du langage parlé (*ibid.*, 11). Bien que les travaux sur la langue première soient à l'origine du domaine, les études actuelles sont réalisées davantage en lien avec la langue seconde ou étrangère. À l'instar de ce qui s'est fait pour la langue première, leur principal objectif de recherche consiste à sélectionner le matériel approprié aux apprenants.

De plus, si les recherches sur la *listenability* ont été principalement menées en éducation et en linguistique, elles l'ont aussi été du point de vue du droit. Par exemple, Charrow & Charrow (1979) ont mené des recherches en psycholinguistique sur la compréhensibilité des instructions au jury. Eastwood & Snook (2012), quant à eux, ont conduit des recherches sur la compréhension des mises en garde de la police tandis que Snook *et al.* (2016) se sont intéressés à la compréhension des droits d'interrogation. Ces études ont été conçues de manière à faciliter la compréhension d'explications de nature juridique. Ainsi, en raison du caractère commun de l'importance de la langue, la *listenability* a également été étudiée dans le domaine du droit, mais ce domaine académique ne sera pas abordé dans cette thèse.

À l'aube du XXI^{e} siècle, avec l'utilisation généralisée des ordinateurs, des technologies permettant de traiter facilement de grands ensembles de données se sont progressivement développées. Les recherches en *listenability* ont progressé parallèlement à cette évolution. Ce troisième chapitre donnera une vue d'ensemble de la recherche sur la *listenability* à ce jour. Comme le thème de cette thèse est la *listenability* dans une langue seconde ou étrangère, notre état des recherches se concentrera sur cette question.

Tout d'abord, il convient de noter que le terme « facilité d'écoute », qui est déjà apparu dans le titre de ce chapitre, sera utilisé dans cette thèse comme la traduction française du concept de *listenability*. La définition du terme lui-même et le

[5] Plus en détail au point 2.2.1.

contexte de l'établissement de la traduction française ci-dessus seront discutés à la section 2.1. avant de retracer les évolutions de la recherche au sein de ce champ à la section 2.2., puis de finir avec des conclusions à la section 2.3.

2.1. Définition de la facilité d'écoute

Afin d'appréhender les recherches sur la facilité d'écoute, nous commençons par définir ce terme. Tout d'abord, nous examinerons les définitions de la *listenability* qui ont été proposées jusqu'à présent.

Harwood (1950: 11), qui a mené les premières recherches sur les études de la *listenability*, la définit ainsi :

> *Listenability is the effect of style, i. e., the number, kind, and arrangement of words of the message in the signal, upon the person receiving.*

Cette définition suggère que, plus la charge d'effet s'avère faible, autrement dit plus l'impact du nombre et du type de mots, etc., l'est, plus l'auditeur peut traiter le contenu reçu facilement. Dès lors, cela rend le contenu plus facile à écouter, c'est-à-dire plus facile à comprendre.

En 1952, Harwood, en collaboration avec Cartier, a publié un article intitulé *« On the definition of listenability »* dans lequel ces deux auteurs ont défini ce champ comme suit (Harwood & Cartier, 1952: 22) :

> *" "Listenability" is thus defined as "capability of being comprehended;" and, of course, there can be degrees of listenability. It is a general concept and depends on the audibility and recognizability of speech transmission, on the vocabulary and the grammatical structure of the language, the ability of the speaker to use helpfully meaningful pauses, stress, intonation, etc., and a multitude of other factors. For example, the listenability of a radio news program is determined by all those factors which contribute to the success of our listening to it, that is the success of our ability to understand it."*

Le terme « *listenability* » est également utilisé afin de désigner des domaines de recherche au niveau desquels Cartier a défini la *listenability* la même année comme suit (Cartier, 1952: 47) :

> *"Listenability is a field of experimental research dealing in part with choice of language for clarity and comprehensibility in speech. As such, it is an important sub-area of rhetoric."*

Par ailleurs, dans le contexte de la *Health Literacy* – un concept qui recouvre la capacité à obtenir, comprendre et utiliser des informations correctes sur la santé et les soins médicaux –, Rubin (2012: 176) détermine ainsi la notion de *listenability* :

> *"Listenability is the quality of discourse that eases the cognitive burden that aural processing imposes. Listenability is a function of oral-based language plus "considerate" rhetorical structures."*

Le terme *« cognitive burden »* que l'on retrouve dans la définition de Rubin (2012) inclurait également la charge due au *« effect of style »* dans la définition de Harwood (1950) et les facteurs tels que le vocabulaire et la structure grammaticale mentionnés dans Harwood & Cartier (1952). Aucune de ces définitions ne traite de *« cognitive burden »*, de *« effect of style »* ou de *« factors »* de manière exhaustive, mais leur clarification devient l'essence même de l'étude de la *listenablity*. La clé de la recherche sur cette dernière consiste à clarifier ce qui se trouve inclus dans *« cognitive burden »*, *« effect of style »* et *« factors »*. Les définitions présentées par Harwood et Cartier, pionniers de la recherche sur la *listenability*, sont suivies dans cette thèse.

En matière de terminologie, comme vu précédemment, les efforts définitoires ont été principalement réalisés en anglais. Bien que le terme n'ait pas encore de traduction ferme en français, il se retrouve néanmoins dans certaines publications, par exemple dans Cornaire (1998: 126). Cependant, il paraît nécessaire de fournir et présenter clairement une traduction en français de la notion de *listenability*.

Sur la base des concepts ci-dessus, dans cette thèse, le terme « facilité d'écoute » se définit 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 (par exemple, lexicale, syntaxique, prosodique, etc.) sur les processus cognitifs de l'auditeur. Nous utiliserons ce terme de manière à indiquer également son domaine de recherche. La section suivante retracera l'histoire de la recherche sur la facilité d'écoute, confirmera ce qui a été rapporté jusqu'à présent, et explorera et clarifiera les défis à venir.

2.2. Histoire des recherches sur la facilité d'écoute

La compréhension orale avait tendance à être négligé puisqu'elle était considérée comme une compétence passive apprise naturellement au cours de l'apprentissage de la langue (Kikuchi & Nakayama, 2006: 254). Comme mentionné dans le chapitre précédent, vers 1970, les programmes d'enseignement ont enfin commencé à s'y intéresser (Morley, 1990: 321). Ainsi, l'Association internationale d'écoute (International Listening Association, ILA) a été créée en 1979 afin de promouvoir toutes les activités liées à la compréhension orale, ce qui inclut non seulement l'éducation et la recherche dans les langues étrangères, mais aussi les affaires, les soins de santé, l'hospitalité, la spiritualité et les arts. Ses activités ne disposent pas d'une très longue histoire[6]. Par conséquent, il n'existe pas non plus de longue histoire des recherches sur la facilité d'écoute, qui se concentrent sur la compréhension orale parmi les quatre compétences langagières (compréhension orale/écrite, production orale/écrite). En revanche, la compréhension écrite fait plus activement l'objet de recherches sur la lisibilité. Celles-ci visent à analyser la difficulté de compréhension de textes et présentent une histoire plus longue que celles sur la facilité d'écoute. Dès lors, il s'avère impossible d'ignorer l'impact des travaux concernant la lisibilité sur les recherches en matière de la facilité d'écoute.

Ainsi, cette section retracera l'histoire de la recherche sur la facilité d'écoute, en présentant ses sous-sections séparément pour chaque étude, délimitant les points de transition en tant qu'histoire de la recherche. Tout d'abord, dans la section suivante 2.2.1., nous donnerons un bref aperçu de la recherche sur la lisibilité. Nous y expliquerons la relation entre la facilité d'écoute et la lisibilité, et nous vérifierons aussi la pertinence d'appliquer des recherches sur la lisibilité à celles sur la facilité d'écoute. Ensuite, les sections 2.2.2. à 2.2.5. détailleront le développement de la recherche sur la facilité d'écoute.

2.2.1. Lisibilité

La lisibilité constitue un domaine qui cherche à évaluer la difficulté de matériels à la lecture, en se reposant sur diverses propriétés des textes : des aspects lexicaux,

[6] https://www.listen.org/.

syntaxiques, de cohérence et de cohésion (François, 2011: 97). La recherche sur la lisibilité est née afin de classer les manuels scolaires et autres documents destinés aux classes élémentaires. Par la suite, les recherches ont été étendues de manière à suggérer comment préparer des matériels plus adéquats aux élèves. Elles se sont concentrées sur la recherche d'une relation entre les éléments structurels du texte et une certaine mesure du succès de ce texte par de grands groupes de lecteurs. Ainsi, les études sur la lisibilité s'intéressent aux critères représentés par le niveau ainsi qu'aux prédicteurs de lisibilité (Lorge, 1944: 404).

Parmi les formules de lisibilité, celles de Lorge (1944), Flesch (1948) et de Dale & Chall (1948) incarnent trois des formules bien connues et appliquées assez souvent afin de prédire la facilité d'écoute.

2.2.1.1. Lorge (1944)

Lorge a proposé une formule de prédiction dans l'échantillon de passages. Les passages choisis correspondent aux 376 extraits des quatre livres du *Standard Test Lessons in Reading* de McCall & Crabbs (1926). Les prédicteurs étudiés par Lorge étaient (1) les cinq utilisés par Gray & Leary (1935) (nombre de mots différents, pourcentage de mots peu communs, nombre relatif de pronoms personnels, nombre relatif de phrases prépositionnelles et longueur moyenne des phrases), (2) un score pondéré pour le vocabulaire basé sur la liste de 20 000 mots de Thorndike (1932), (3) quatre éléments utilisés par Morriss & Holversen (1938) (pourcentage de mots élémentaires, pourcentage de localismes simples, pourcentage de mots concrets et pourcentage de mots abstraits), et (4) les deux variables de Flesch (morphèmes affixes et intérêt humain)[7].

Par conséquent, Lorge (1944) propose la formule suivante :

$$Y = 1{,}6126 + 0{,}07\,X_1 + 0{,}1301\,X_2 + 0{,}1073\,X_3$$

Y : une estimation du niveau scolaire correspondant à la réussite des trois quarts des questions d'un passage donné
X_1 : le nombre moyen de mots par phrase
X_2 : le nombre de phrases prépositionnelles pour 100 mots

[7] Voir le point 2.2.1.2.

$\mathbf{X_3}$: le nombre de mots difficiles ne figurant pas sur la liste Dale (1941) de 769 mots[8]

Cette formule de Lorge constitue un moyen de juger de la difficulté relative ou de la lisibilité de passages lus ou parlés. La lisibilité est basée sur la compréhension de passages par des écoliers.

2.2.1.2. Flesch (1948)

Flesch (1948) propose une paire de formules : la première indique la facilité de lecture tandis que la seconde établit un indice d'intérêt humain. Celle-ci revisite la formule de lisibilité de Flesch (1943) qui est basée sur la longueur moyenne des phrases en mots, le nombre d'affixes et le nombre de références personnelles (Flesch, 1948: 221).

Plus précisément, les inconvénients de la structure de base de la formule et les difficultés à utiliser la formule ont été réexaminés. Le premier problème reposait sur le fait qu'elle ne capturait pas toujours la grande lisibilité d'une écriture conversationnelle. Concernant le second, les utilisateurs de la formule ont trouvé le décompte du nombre d'affixes particulièrement fastidieux et ont admis leur incertitude quant à leur repérage. De plus, les références personnelles de la formule ont également parfois été ressenties comme arbitraire tandis que le principe sous-jacent a souvent été mal compris. Pour les raisons susmentionnées, la formule reconsidérée et nouvellement présentée est la suivante (Flesch, 1948: 221-222).

Pour la première des formules avec la facilité de lecture comme variable dépendante, son coefficient de corrélation multiple[9] s'élève à 0,7.

R.E. ("reading ease") = 206,835 – 0,846 wl – 1,015 sl

[8] La liste Dale est composée de mots communs aux 1 000 premiers mots anglais les plus fréquents de Thorndike et aux 1 000 premiers mots les plus fréquents connus par les enfants entrant en première année (Lorge, 1944: 413).

[9] Il s'agit du coefficient de corrélation entre la valeur réelle observée de la variable dépendante et l'estimation calculée en ajustant une équation de régression multiple.

R.E. : 0 (pratiquement illisible) et 100 (facile pour toute personne alphabétisée)
wl : nombre de syllabes par 100 mots
sl : nombre moyen de mots par phrase

Pour la formule qui considère l'intérêt humain comme variable dépendante, son coefficient de corrélation multiple s'élève à 0,43. Cette formule est considérée comme ayant plus d'intérêt humain si elle comporte plus des mots et des phrases personnels.

H.I. ("human interest") = 3,635 pw + 0,314 ps

H.I. : 0 (aucun intérêt humain) et 100 (intérêt humain élevé)
pw : pourcentage des « mots personnels »[10]
ps : pourcentage des « phrases personnelles »[11]

Ces deux formules de Flesch (1948) ont souvent été appliqués à la facilité d'écoute plus tard, mais en 1943, lorsque Flesch a publié un modèle antérieur à ceux-ci, il avait ceci à dire sur la facilité d'écoute (1943: 43) :

> *Writers for a radio audience may well supplement this method [the use of his formula] with certain recent findings about the relationship between reading and listening comprehension. Goldstein [...] has found significant differences in favor of listening for third grade reading, but no reliable [...] differences for seventh grade material. Larsen and Feder [...] have found listening superior for easy passages, reading superior for passages of medium difficulty, and a very pronounced difference in favor of reading for difficult material. In other words, if material is put on the air rather than on the printed page, easy matter*

[10] Tous les noms en genre naturel, tous les pronoms sauf les pronoms neutres, et les mots *« people »* (utilisés avec le verbe pluriel) et *« folks »* (Flesch, 1948: 223).
[11] Phrases parlées, marquées par des guillemets ou autrement ; questions, ordres, demandes et autres phrases directement adressées au lecteur, exclamations ; et phrases grammaticalement incomplètes dont le sens doit être déduit du contexte (Flesch, 1948: 223).

becomes easier but difficult matter more difficult. Radio will therefore magnify differentials in difficulty by the writer's formula.

Il n'est pas exagéré de dire que la recherche sur la facilité d'écoute a commencé par cette description. Dans la section 2.2.2. ci-dessous, nous examinerons les recherches menées aux débuts de la recherche sur la facilité d'écoute, qui se trouvent également liées à la citation de Flesch ci-dessus.

2.2.1.3. Dale & Chall (1948)

La formule de Dale & Chall (1948) s'appuie, quant à elle, sur le pourcentage de mots absents de la liste de 3 000 mots de Dale et le nombre moyen de mots par phrase. Son coefficient de corrélation multiple s'élève à 0,7.

$$Y = 0{,}1579\ X_1 + 0{,}0496\ X_2 + 36365$$

Y : le score en lecture d'un élève qui pourrait répondre correctement à la moitié des questions du test
X_1 : le score de Dale (nombre relatif de mots en dehors de la liste de Dale de 3 000 mots)
X_2 : le nombre moyen de mots par phrase

Ce modèle constitue également la base de l'analyse de la transcription de matières sonores. Les recherches sur la facilité d'écoute ont commencé par l'application de ces modèles.

2.2.2. Lisibilité et facilité d'écoute

À partir des quatre formules (celles de Lorge, les deux de Flesch et celle de Dale & Chall) présentées précédemment, destinées à l'anglais, des premières recherches sur la facilité d'écoute se développent dans le monde anglo-saxon. Dans cette section, nous présenterons sept travaux situés dans cette lignée. Ceux-ci correspondent à des études qui utilisent la formule de lisibilité comme base. L'objectif de ces études consiste à analyser si une formule de lisibilité, développée

sur le langage écrit, peut être appliquée de manière à prédire la difficulté du langage parlé.

Suite à la description de Flesch (voir 2.2.1.2.), l'application de la recherche sur la lisibilité a commencé à être expérimentée à la recherche sur la facilité d'écoute. L'étude de transcription a constitué l'un des premiers objets de recherche. En d'autres termes, il s'agit d'une facilité d'écoute, qui ne se révèle pas basée sur le son, mais sur la langue écrite.

2.2.2.1. Chall & Dial (1948)

Chall & Dial (1948) analysent si deux formules de lisibilité, à savoir celle de Flesch et celle de Dale & Chall, s'avèrent pertinentes afin de prédire le niveau de difficulté de transcriptions de nouvelles radiophoniques. Les données sont basées sur 18 transcriptions d'actualités enregistrés à la station WOSU[12]. Les sujets de cette recherche comptent 100 étudiants de première année des universités de Franklin et de l'Ohio. Les auteurs ont réalisé les deux mesures subjectives avec une échelle de Likert à cinq points et les questions objectives à choix multiples. Les premières sont réalisées afin de déterminer le niveau de compréhension (1, pas du tout compris, à 5, parfaitement compris) et d'intérêt (1, pas intéressant, à 5, très intéressant) des participants. Les dernières sont conçues de manière à mesurer la compréhension réelle des étudiants[13]. Les auteurs ont examiné la corrélation entre les valeurs provenant de mesures subjectives et de questions objectives, et les scores de difficulté des manuscrits prédites par les formules. Ils ont ainsi montré une relation certaine entre la compréhension et l'intérêt, puisque tant la corrélation entre le jugement de compréhension et celui d'intérêt (0,93) que celle entre le nombre de questions objectives correctes et le jugement d'intérêt (0,82) se révèlent élevées. Cependant, les transcriptions classées en fonction de la difficulté de lecture prédite par les formules de lisibilité n'ont pas montré une relation pertinente avec les jugements de compréhension des auditeurs. Il existait un accord uniquement pour les manuscrits les plus faciles. Ainsi, la formule de lisibilité ne semble pas capable de prédire parfaitement la difficulté des transcriptions. Malgré tout, cette analyse montre qu'avec des ajustements, une

[12] La station WOSU est une station de radio qui se trouve à Columbus, dans l'Ohio.
[13] Les questions objectives à choix multiples ne sont réalisées que pour 9 des 18 transcriptions.

formule de lisibilité pourrait devenir un outil utile permettant de prédire la compréhension et l'intérêt de l'auditeur.

À la suite de cette étude, dans les années 1950, les chercheurs ont commencé à s'intéresser à la facilité d'écoute des documents présentés oralement (Carroll, 1971: 104). En effet, plusieurs recherches ont été réalisées concernant la relation entre la facilité d'écoute et la lisibilité.

2.2.2.2. Allen (1952)

Allen (1952) analyse si les formules de lisibilité de Flesch (1948), de Dale & Chall (1948) et de Lorge (1944) s'avèrent efficaces afin de prédire la difficulté des matériels oraux. Il examine également les facteurs des meilleurs prédicteurs de la difficulté. Pour ce faire, les données sont basées sur quatre commentaires écrits à propos de deux types de films. Ceux-ci sont écrits selon les formules de la facilité de lecture et de l'intérêt humain de Flesch, répartis selon les deux niveaux de difficultés et les deux niveaux d'intérêt humain suivants : 5e primaire [14] & très intéressant, 5e primaire & monotone, 1e secondaire & très intéressant et 1e secondaire & monotone. Les groupes expérimentaux comprenaient 668 élèves de 6e primaire dans quatre écoles primaires du comté de San Bernardino, en Californie.

Tout d'abord, les films ont été montrés aux groupes expérimentaux ; la bande sonore d'origine n'a pas été diffusée, mais a été remplacée par l'enregistrement audio de commentaires écrits. Puis, la compréhension des participants a été mesurée par des questions à choix multiples et des questions à compléter. Il est apparu que les formules de Flesch, Dale & Chall et Lorge se révèlent à peu près égales concernant la prédiction de la compréhension des commentaires de films. Cependant, celle de Lorge prédisait systématiquement un niveau de lecture inférieur à celui proposé par la formule de la facilité de lecture de Flesch, tandis que celle de Dale & Call variait quant à ses prédictions. Par ailleurs, le facteur qui sert le plus à la prédiction constitue la longueur moyenne des phrases, dont la corrélation s'avère la plus élevée (0,61) avec la mesure de la compréhension. Ce facteur est inclus dans les trois formules de lisibilité.

---

[14] Dans cette thèse, toutes les années dans le système scolaire sont ajustés à celles de la Belgique.

2.2.2.3. Cartier (1955)

Cartier (1955) examine l'applicabilité de la formule de Flesch au langage parlé. Il existe trois types de données enregistrées, chacun avec un HI (*humain interest*) de Flesch, pour sept histoires de niveau différent de RE (*reading ease*), soit 21 données au total. Les sujets de recherches constituent un groupe de plus de 100 élèves de 4e secondaire à Compton, en Californie. Leur niveau de compréhension est mesuré par un test de compréhension de quinze questions à choix multiple, proposant cinq possibilités. Les résultats montrent que la formule de HI de Flesch ne constitue pas un indicateur utile de la facilité d'écoute. En revanche, chaque prédiction de difficulté de la formule de RE semble assez cohérente pour toutes les données.

2.2.2.4. Harwood (1955a)

La même année, Harwood (1955a) examine les relations entre la langue écrite et la langue parlée de différents niveaux de difficulté. Les données sont sélectionnées à l'aide des deux formules de Flesch. Les données utilisées s'appuient sur des articles de United Press Radio, des transcriptions de commentaires et de reportages radio du Columbia Broadcasting System, et des transcriptions de discours publics, pour un total de sept interventions. Chaque objet de l'étude présente un niveau différent de lisibilité prédite par la formule de Flesch, et se voit associé quinze questions à choix multiple. Toutes les histoires ont été présentées de manière orale à un groupe de sujets et de manière visuelle (à la lecture) à un autre groupe similaire. Les sujets de la recherche comptent 240 élèves de 4e secondaire à Compton, en Californie, 120 ayant passé les tests de lecture, tandis que les 120 autres ont réalisé les tests d'écoute. Les mêmes questions ont été posées à chaque groupe. Le nombre moyen de réponses correctes après la présentation sonore a été comparé à celui suivant la présentation visuelle. Les résultats montrent que dans l'ensemble, les données s'avèrent plus compréhensibles lorsqu'elles sont présentées à la lecture qu'à l'audition, et que la formule de lisibilité peut en général prédire la compréhension de l'oral à peu près aussi bien que celle de la lecture. Cependant, les données prédites comme difficiles à lire ont, en particulier, tendance à présenter des différences significatives en

matière de compréhension entre la lecture et l'audition. Malgré tout, l'auteur conclut que le développement d'une formule de la facilité d'écoute semble possible.

2.2.2.5. Harwood (1955b)

Harwood (1955b) a réalisé une autre recherche la même année. Le but de celle-ci consistait à examiner la relation entre le niveau de lisibilité prédit par la formule de lisibilité de Flesch et le débit de parole. Les données utilisées comptaient les mêmes que celles de Harwood (1955a), avec sept niveaux différents de lisibilité prévus. Chaque niveau comporte quatre débits de parole différents, pour un total de 28 ensembles de données. Ces dernières ont toutes été enregistrées et comprenaient quinze questions à choix multiple concernant le contenu. Les sujets comptaient 487 élèves de 4e secondaire à Compton, en Californie. Les résultats de cette recherche indiquent que la facilité d'écoute diminue avec l'augmentation du débit de parole. Cependant, il n'apparaît pas statistiquement de différences significatives entre la facilité d'écoute moyenne pour chacun des quatre débits de parole. En ce qui concerne la relation entre la lisibilité et la facilité d'écoute, l'auteur conclut que, selon les limites et les conditions examinées, la lisibilité peut être utilisée comme prédicteur brut de la facilité d'écoute. Il faut également souligner que cette étude constitue la première à prendre en compte les paramètres prosodiques.

2.2.2.6. O'Keefe (1971)

L'un des objectifs du travail d'O'Keefe consiste à analyser les transcriptions d'émissions sur des actualités de Voice of America (VOA) en fonction de la formule de Flesch. Les données sont basées sur deux types de transcriptions en anglais en Asie du Sud-Est et des émissions de trois autres pays (Russie, Grande-Bretagne et Allemagne de l'Ouest). L'autre but de cette recherche consiste à comparer les productions de ces quatre pays en matière de facilité d'écoute.

Afin d'expliquer plus précisément la nature des données, il faut mentionner que la moitié provient de l'émission régulière, qui cible des anglophones natifs ou des auditeurs familiers avec l'anglais. L'autre moitié est tirée de l'émission spéciale destinée à des apprenants et aux personnes non familières avec l'anglais.

Les résultats montrent que, d'après les scores de lisibilité de Flesch, les deux types de matériel de VOA s'avèrent destinés à un public très instruit. Même lors des émissions spéciales, la moyenne générale ne dépasse pas 50,61, ce qui se situe entre « difficile » et « assez difficile ». Les comparaisons entre les quatre pays ont permis de constater que les différences de difficulté ne s'avèrent pas très marquées. Sur la base de ces résultats, l'auteur conclut que, selon l'échelle éducative de Flesch, la plupart des auditeurs des émissions spéciales devraient avoir l'équivalent de la première secondaire au moins afin de comprendre même les émissions les plus simples. En outre, la VOA peut également limiter inutilement son audience avec ses émissions régulières.

2.2.2.7. Denbow (1975)

Denbow (1975) a tenté de clarifier la relation entre les matériels écrits et oraux, en utilisant les données des articles de presse extraites à l'aide de la formule de Dale & Chall (1948). Deux articles ont été utilisés ; chacun est décliné en une version de bas niveau et une autre de plus haut niveau, ce qui donne, au total, quatre documents, en version écrite et orale. Leur niveau de lisibilité est défini par la formule de Dale & Chall (1948). Les 140 sujets de cette expérience ont été sélectionnés dans sept classes d'expression orale et deux classes de journalisme de l'université Marshall. Ils ont passé un test préalable et un test postérieur, selon le format de tests à trous, sur deux des articles. Par ailleurs, ils étaient divisés en trois groupes expérimentaux et un groupe de contrôle : (1) un groupe de lecture normale, (2) un groupe de lecture rythmée, (3) un groupe oral, et (4) un groupe témoin. Les sujets du premier groupe lisaient chaque article de la manière habituelle tandis que ceux du deuxième groupe recevaient des articles, une phrase à la fois, par l'intermédiaire d'un projecteur de diapositives, pour une durée totale d'une minute et demie par article. Ceux du troisième groupe, quant à eux, les écoutaient pendant une minute et demie par article. Enfin, le groupe de contrôle lisaient des articles non pertinents.

Les résultats montrent que les articles faciles ont entraîné un niveau de compréhension significativement plus important que les articles difficiles, alors qu'il n'est apparu aucune interaction significative entre la modalité et le niveau de lisibilité. Dès lors, pour ces deux articles de contenu, la formule de

Dale & Chall prédit avec précision la difficulté relative des augmentations de la compréhension orale et écrite.

Comme nous l'avons déterminé précédemment, l'intérêt central de la recherche au cours de cette période a consisté à déterminer si les formules de lisibilité pouvaient également être appliquées à des transcriptions de documents oraux de manière à en prédire leur difficulté de compréhension à l'oral. Parmi les formules de lisibilité existantes, la question de recenser les formules les plus valables a également été examinée. Aucune étude n'a conclu que les formules de lisibilité s'avèrent totalement inutiles afin de prédire la facilité d'écoute ; au contraire, certaines se distinguent en démontrant leur validité. Pourtant, comme Carroll (1971: 104) le remarque, bien que diverses études aient été menées, les preuves de validité générale de l'application de formules de lisibilité se révèlent très rares afin de prédire la facilité d'écoute. Comme la formule de lisibilité s'applique à la langue écrite, ce résultat pourrait découler des différences fondamentales entre la langue parlée et la langue écrite. Par exemple, le vocabulaire utilisé et la quantité d'informations utilisées peuvent différer lors de la description d'un même contenu. Dans la recherche présentée dans cette section, l'applicabilité de la formule de lisibilité ne reste qu'une piste attendue. Par conséquent, plutôt que de l'appliquer directement à l'étude de la facilité d'écoute, il peut être plus pertinent d'incorporer des morceaux de la formule de lisibilité ou d'établir une formule spécifique à la langue parlée.

2.2.3. Développement de formules pour la langue parlée

Dans les années 1960, l'époque au cours de laquelle l'application de formules de lisibilité à la recherche de facilité d'écoute a réalisé des progrès, un mouvement de production des formules spécifiques à la langue parlée a été observé, et pas seulement dans le but d'appliquer des formules de lisibilité conçues pour l'écrit. Cette section donnera un aperçu des recherches à l'origine des formules de prédiction de la facilité d'écoute au sens propre.

2.2.3.1. Rogers (1962)

Rogers (1962) décrit une formule permettant de prédire le niveau de compréhension des matériels sonores. Il est fait référence aux études de lisibilité lors de la création de sa formule.

Cet auteur initie ses recherches concernant les formules de lisibilité, qui montrent que les huit variables suivantes s'avèrent significativement associées à la lisibilité :

- la longueur moyenne des phrases ;
- l'indice de subordination ;
- la longueur moyenne des clauses ;
- le nombre moyen de phrases prépositives par 100 mots ;
- le nombre moyen de phrases infinitives par 100 mots ;
- le nombre moyen de phrases infinitives et prépositives par 100 mots ;
- le nombre moyen de mots par 100 mots ne figurant pas sur la liste longue de Dale & Chall (1948)[15] ;
- le nombre moyen de mots ne figurant pas sur la liste plus courte de Dale (1941).

Cependant, la première variable, la longueur moyenne des phrases, pose un problème lors de l'étude de la langue parlée. En effet, les phrases ne sont pas toujours parfaitement séparées par des points ou d'autres marques ou inflexions. Par conséquent, au lieu de la longueur, une variable reposant sur les « unités d'idées » a été utilisée. Sa longueur est déterminée en divisant le nombre de mots de la transcription par le nombre de clauses indépendantes.

L'auteur a ensuite collecté des données en partant du principe que les enfants comprennent la langue qu'ils parlent et la langue parlée par les enfants de leur âge. Pour les données, des échantillons de langage spontanés et non répétés ont été enregistrés auprès d'au moins huit enfants de chaque niveau de cinq écoles

[15] La liste Dale consiste en environ 3 000 mots familiers. Ces mots sont connus en lecture par au moins 80 % des enfants de la 4e primaire. En outre, la liste présente une corrélation significative avec la difficulté de lecture. En revanche, elle n'est pas conçue comme une liste des mots les plus importants pour les enfants ou les adultes (Dale & Chall, 1948: 44).

primaires, cinq collèges et cinq lycées, soit 480 échantillons au total. L'équation de régression qui en résulte afin de prédire le niveau scolaire est la suivante.

$$G = .669\ I + .4981\ LD - 2.0625$$

G : le niveau scolaire
I : la longueur moyenne de l'unité d'idée (le nombre de mots de la transcription ou le nombre de clauses indépendantes)
LD : le nombre moyen de mots dans un échantillonnage de 100 mots qui n'apparaissent pas sur la longue liste de Dale

Le coefficient de corrélation multiple de *G* avec *I* et *LD* s'élève à 0,73. Cependant, cette formule est présentée comme non testée et non prouvée. Par conséquent, nous pouvons nous demander si elle peut être utilisée efficacement dans la pratique pour la population cible. Malgré tout, en matière de création d'une formule appliquée à la langue parlée, cette étude s'avère unique par rapport aux recherches précédentes sur la facilité d'écoute. De plus, cette formule apporte une nouvelle vision de variable dépendante : l'unité d'idée.

2.2.3.2. Fang (1967)

Fang (1967) affirme que la formule de la facilité d'écoute doit satisfaire à trois critères et aisni être :

- simple à utiliser ;
- orientée vers la diffusion ;
- fortement corrélée avec les facteurs de la facilité d'écoute déjà testés.

Concernant le premier critère, il ne faut pas oublier qu'une formule compliquée ne s'avère pas très utile, surtout à une époque au cours de laquelle l'informatisation demeure encore embryonnaire. À propos du deuxième critère, la comparaison des articles d'un journal et des scripts d'émissions radios devrait, par exemple, faire apparaître des différences entre eux. Même si le contenu reste le même, le texte peut se révéler plus long lorsqu'il est écrit pour les journaux et plus court lorsqu'il l'est pour la radio. Par conséquent, le support de texte doit

également être pris en compte. En ce qui concerne le dernier critère, une formule de la facilité d'écoute doit présenter une forte corrélation avec les facteurs.

En tenant compte de ces points, Fang (1967) propose une formule comptant uniquement le nombre de mots polysyllabiques.

> *ELF (Easy Listening Formula) = nombre de syllabes au-dessus d'une par mot*

Plus le score *ELF* est élevé, plus l'écriture à présenter oralement s'avère difficile à comprendre. Pour un texte, la facilité d'écoute est calculée avec la moyenne des *ELF* par phrase.

La corrélation entre les comptages de cette formule et la formule RE de Flesch (1948) s'élève à 0,96, ce qui semble extrêmement élevé, pour 36 scripts de télévision et 36 échantillons de journaux. L'auteur conclut que les rédacteurs de journaux télévisés ayant une grande réputation utilisent un style dont le score moyen selon la formule *ELF* se révèle inférieur à 12.

Cependant, Glasser (1975: 139-140) souligne les problèmes de la formule de Fang afin de mesurer la facilité d'écoute. Tout d'abord, cette formule se compose uniquement d'informations sur le nombre de syllabes et ne tient donc absolument pas compte de la syntaxe ou du sens. Ainsi, par exemple, « *Big(0) he(0) saw(0) never(1) awe(0) ohm(0) latest(1)*[16] » est évalué comme facile à écouter, bien qu'il s'agisse d'une construction dénuée de sens. Le deuxième problème concerne l'hypothèse selon laquelle les mots les plus fréquemment utilisés sont les plus courts. Cependant, le mot « *plane* », par exemple, est bien plus ambigu que le mot « *airplane* ». Or, ces deux points importants ne sont pas pris en compte dans la formule de Fang.

2.2.3.3. Kiyokawa (1990)

Si l'étude de Kiyokawa (1990) est placée dans cette section, elle a été réalisée environ 20 ans après l'étude précédente de Fang (1967). Cependant, cela ne signifie pas qu'aucune recherche sur la facilité d'écoute n'ait été menée pendant

[16] Il s'agit d'un exemple tiré de Glasser (1975: 139), où les chiffres entre parenthèses proviennent également du texte original. Les chiffres indiquent le nombre de syllabes de chaque mot, conformément à la formule de Fang.

cette période. Dans les années 1970-1980, Kiyokawa (1976, 1977, 1985, 1987, 1989), dont le travail est présenté dans cette section, a mené des recherches énergiques concernant la facilité d'écoute. Le point culminant de celles-ci est Kiyokawa (1990), qui est discuté dans cette section.

Héritant des méthodes de Rogers (1962), Kiyokawa (1990) essaye de développer une formule facilement utilisable et largement applicable afin de déterminer la facilité d'écoute des textes oraux. Les données utilisées représentent 50 échantillons de discours choisis au hasard parmi un total de 502 textes oraux réalisés par des élèves de la 1e primaire à la 6e secondaire, et recueillis par Rogers au Texas. La formule de Kiyokawa (1990) consiste en deux variables : la difficulté des mots, soit le pourcentage de mots ne figurant pas dans la liste des mots proposée par Kiyokawa (1987), et longueur de la phrase, soit le nombre moyen de mots par phrase.

$$Y = 0.6446\ X_1 + 0.2211\ X_2 + -4.1217$$

Y : les niveaux scolaires de 50 échantillons utilisés par Rogers (1962)
X_1 : le pourcentage de mots ne figurant pas dans la liste des mots proposée par Kiyokawa (1987)
X_2 : le nombre moyen de mots par phrase

La corrélation multiple de cette formule avec les niveaux scolaires s'élève à 0,47. L'important est qu'environ 40 % de la difficulté des échantillons peut être expliquée par la difficulté des mots. Afin de vérifier l'efficacité de cette formule pour l'enseignement, trois manuels d'anglais ont été analysés en utilisant la formule créée. En conséquence, les scores moyens calculés à l'aide de cette formule de facilité d'écoute ont généralement augmenté avec les niveaux scolaires indiqués. Cependant, le problème est que cette formule est basée sur les seuls dialogues, ce qui aurait pu affecter le résultat.

Comme nous l'avons mentionné précédemment, le principal objectif de recherche de cette période consistait à développer une formule originale permettant de prédire la facilité d'écoute. Cependant, certains défis subsistent ; par exemple, la validité des formules n'a pas été testée de façon satisfaisante et les coefficients de corrélation ne s'avèrent pas très élevés. De plus, comme le

montrent bien les variables de chaque formule, les travaux se limitaient à l'analyse de transcriptions. Les éléments spécifiques à l'oral – comme la prosodie, la modification phonologique, les pauses et le débit de parole – s'avéraient, quant à eux, ignorés.

2.2.4. Modèles spécialement conçus pour les apprenants

Une série d'études menées principalement par Kotani et Yoshimi depuis les années 2010 s'avère particulièrement remarquable. Ces auteurs ont réalisé un certain nombre d'études concernant la facilité d'écoute, en prenant l'anglais comme langue cible. Les sujets de recherche regroupent des locuteurs natifs du japonais. Nous examinerons leurs résultats dans les sections 2.2.4.1. à 2.2.4.6.

2.2.4.1. Ueda *et al.* (2013)

Ueda *et al.* (2013) précisent que les modèles doivent être adaptés aux niveaux de compétence des apprenants, car ceux-ci varient d'une personne à l'autre. En outre, ils soulignent également les problèmes rencontrés à mesurer la difficulté de compréhension orale des matières d'apprentissage, car les modèles précédents de la facilité d'écoute n'avaient pas été créés en fonction de la compétence des apprenants.

Les auteurs ont eu ainsi pour objectif de construire une formule de prédiction de la difficulté de compréhension orale, phrase par phrase, en tenant compte de la compétence en anglais des apprenants. Pour ce faire, quatre articles de Voice of America (VOA) ont été utilisés comme données, dont deux constituent des articles simples recueillis auprès de VOA Special English et les deux autres des articles difficiles recueillis auprès de VOA Editorials[17]. Une analyse de régression

[17] Comme mentionné à la section 2.2.2.6, VOA Special English est l'émission spéciale destinée à des apprenants et autres personnes non familières avec l'anglais. Elle se compose de phrases courtes et simples et n'utilise qu'un vocabulaire de base de 1 500 mots. En revanche, elle ne contient pas de phrases idiomatiques. Elle est lue clairement et lentement, à environ deux tiers du débit normal des locuteurs natifs (environ 250 syllabes par minute) (Ueda *et al.*, 2013: 411). VOA Editorials est l'émission régulière qui cible des anglophones natifs et ceux qui sont familiers avec l'anglais. Elle inclut des éditoriaux qui présentent différents arguments sur un large éventail de sujets. En outre, elle est lue au débit typique des locuteurs natifs (*ibid.*, 411).

multiple a été utilisée afin de construire la formule. La variable dépendante représente la difficulté subjective de compréhension orale des phrases en anglais estimées par 90 apprenants. Les variables indépendantes, quant à elles, comprennent : le nombre moyen de mots dans une phrase, le nombre moyen de syllabes dans un mot, le nombre de mots polysyllabiques (le nombre de mots ayant plus de deux syllabes dans une phrase), le taux de mots difficiles dans une phrase (défini comme le taux de mots ne figurant pas sur la liste de mots de Kiyokawa (1990) dans la phrase)[18], le score de compréhension orale du TOEIC (les 90 scores des apprenants de langue anglaise à la section de la compréhension orale du TOEIC). Le coefficient de corrélation multiple R de l'équation incluant toutes les variables ci-dessus s'élevait à 0,54, avec un ratio de contribution R^2 de 0,29. Le score de compréhension orale du TOEIC (coefficient de régression partielle standardisé -0,43) s'est avéré exercer la plus grande influence sur les valeurs prédites. Cela suggère la validité de l'ajout du score de compréhension orale du TOEIC aux prédispositions afin de mesurer la difficulté de compréhension orale, qui prend en compte le niveau de compétence des apprenants. Ainsi, cela confirme également l'hypothèse de l'auteur, qui estimait que la variation interpersonnelle constitue un élément important à prendre en compte dans le domaine de la facilité d'écoute.

2.2.4.2. Ueda *et al.* (2014)

Dans Ueda *et al.* (2014), l'objectif consiste également à construire une formule de prédiction (équation de régression multiple) de la difficulté de compréhension orale en fonction du niveau de compétence des apprenants. Ce niveau se trouve à nouveau opérationnalisé, comme le score de compréhension orale du TOEIC qui est inclus comme variable indépendante dans la formule. La variable dépendante correspond au jugement subjectif des apprenants sur la difficulté de compréhension orale des phrases. Ces variables sont collectées à l'aide de 90 apprenants d'anglais dont la langue maternelle est le japonais. Les données sont constituées d'un total de 80 phrases comprenant deux articles simples (25 phrases

[18] Toutes les références au taux de mots difficiles dans une phrase apparaissant jusqu'à la section 2.2.4.6. font référence au taux des mots ne figurant pas sur la liste de mots de Kiyokawa (1990) dans la phrase, avec la liste de mots de Kiyokawa (1990) comme référence.

par article) et deux articles difficiles (15 phrases par article) recueillis auprès de VOA, comme dans Ueda *et al.* (2013). Les apprenants ont écouté un article lu par un locuteur natif et ont jugé subjectivement la difficulté d'écoute d'une phrase à chaque fois qu'ils ont fini de l'écouter. Cette difficulté a été évaluée sur une échelle de cinq points : 1, facile ; 2, légèrement facile ; 3, normal ; 4, légèrement difficile ; 5, difficile. En outre, un test de compréhension comportant cinq questions à quatre choix par article a également été réalisé. En sus du score de compréhension orale du TOEIC, les variables indépendantes de l'analyse de régression multiple comprennent le nombre moyen de mots dans une phrase, le nombre moyen de syllabes dans un mot, le nombre total de syllabes moins un pour chaque mot d'une phrase et le taux de mots difficiles dans une phrase. Les auteurs ont construit et comparé trois formules :

- une formule utilisant toutes les variables indépendantes ci-dessus ;
- une formule excluant le score de compréhension orale du TOEIC ;
- une formule utilisant uniquement le score de compréhension orale du TOEIC.

Le coefficient de corrélation multiple avec la difficulté de compréhension orale s'est révélé plus élevé pour la première formule. Par conséquent, les résultats suggèrent que la perception subjective de la difficulté de la compréhension orale par les apprenants est influencée à la fois par les caractéristiques linguistiques de la phrase et par la capacité de compréhension orale des apprenants.

2.2.4.3. Kotani *et al.* (2014)

Kotani *et al.* (2014) ont également développé une méthode de mesure de la facilité d'écoute. La variable dépendante correspond à la difficulté subjective de compréhension orale des phrases en anglais. La facilité d'écoute est évaluée sur une échelle de Likert en cinq points, en fonction des jugements des apprenants. Les variables indépendantes constituent les caractéristiques de l'apprenant et les caractéristiques linguistiques du texte. Les premières mesurent la compétence en compréhension orale à partir du score de compréhension orale du TOEIC, tandis

que les secondes capturent les complexités lexicales, syntaxiques et phonologiques d'une phrase au travers des variables suivantes :

- le nombre moyen de syllabes dans un mot ;
- le nombre moyen de mots dans une phrase ;
- le taux de mots difficiles dans une phrase ;
- le nombre de mots polysyllabiques dans une phrase ;
- le débit de parole ;
- le taux d'élision dans une phrase (nombre d'élision/nombre total de mots dans une phrase)[19] ;
- le taux de réduction dans une phrase (nombre de réductions/nombre total de mots dans une phrase)[20] ;
- le taux de contraction dans une phrase (nombre de contraction/nombre total de mots dans une phrase)[21] ;
- le taux de liaison dans une phrase (nombre de liaison/nombre total de mots dans une phrase)[22] ;
- le taux de déduction dans une phrase (nombre de déduction/nombre total de mots dans une phrase)[23].

Pour cette étude, 90 apprenants universitaires de l'anglais langue étrangère ont participé à la collecte de données. Les données utilisées dans Kotani *et al.* (2014) ont aussi été collectées auprès de deux types de sections de VOA : les sections facile et difficile. Elles comprennent cinq questions de compréhension à choix multiple. Les résultats d'analyse suggèrent que le score de compréhension orale du TOEIC s'avère utile afin de mesurer la facilité d'écoute des apprenants. Parmi les cinq caractéristiques permettant la modification phonologique (élision, réduction, contraction, liaison, déduction), le taux d'élision et de réduction n'a pas entraîné d'effet statistiquement significatif. Parmi le taux de contractions, de liaisons et de déductions engendrant un effet significatif, le taux de liaison a eu un effet positif et celui de contraction et de déduction, un effet négatif. Or, la

[19] L'élision est l'élimination des phonèmes.
[20] La réduction est l'affaiblissement du son en transformant une voyelle en schwa.
[21] La contraction est la combinaison de deux mots.
[22] La liaison consiste à relier le son final d'un mot au son initial du mot suivant.
[23] La déduction est l'élimination des sons entre les mots.

présence d'un effet négatif des taux de contraction et de déduction suggère que la modification phonologique peut augmenter la facilité d'écoute pour les apprenants. Il existe également l'effet de la longueur des mots. Comme le montre la valeur négative du coefficient de régression partielle standardisé, les mots plus longs augmentent la facilité d'écoute. Cette étude paraît remarquable dans la mesure où elle constitue la première à utiliser des variables de prosodie.

2.2.4.4. Kotani & Yoshimi (2017)

Kotani & Yoshimi (2017) développent une méthode de mesure de la facilité d'écoute à l'aide d'un algorithme de classification par arbre de décision qui classe les phrases en cinq niveaux de facilité d'écoute. Les auteurs déterminent la précision de la mesure de la facilité d'écoute et la discriminabilité des caractéristiques linguistiques et des apprenants dans cette classification.

La facilité d'écoute est jugée par les apprenants d'anglais langue étrangère à l'aide de notes sur une échelle de Likert en cinq points (1, facile ; 2, assez facile ; 3, moyen ; 4, assez difficile ; ou 5, difficile). Les scores sont évalués phrase par phrase, chaque apprenant écoutant et attribuant des scores pour 80 phrases de quatre données sélectionnées dans les sections facile et difficile sur le site VOA. Cette difficulté subjective de compréhension orale des phrases est fixée comme variable dépendante. Les variables indépendantes suivantes ont été utilisées : le nombre moyen de mots dans une phrase, le nombre moyen de syllabes dans un mot, le nombre de mots polysyllabiques dans une phrase, le taux de mots difficiles dans une phrase, le débit de parole, le taux de modification phonologique comme caractéristiques linguistiques. De plus, le score de compréhension orale du TOEIC, l'expérience d'apprentissage (le nombre de mois pendant lesquels les apprenants ont étudié l'anglais), l'expérience de séjour dans une région anglophone (le nombre de mois passés par les apprenants dans les pays anglophones), et la fréquence d'écoute sont considérés comme caractéristiques des apprenants (le score sur une échelle de Likert en cinq points concernant la fréquence de l'utilisation de l'anglais).

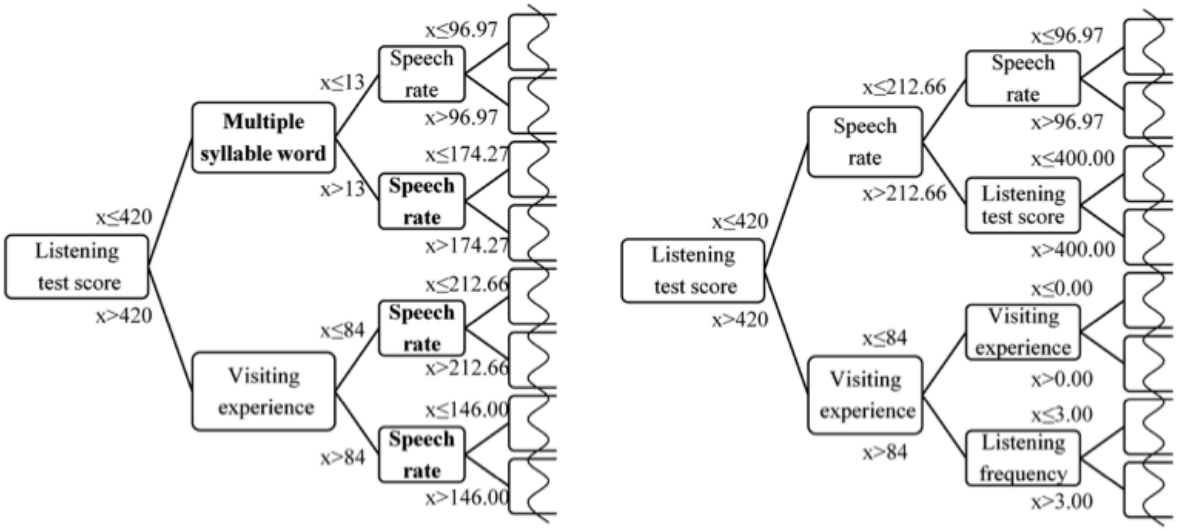

Figure 6 : Arbre de décision (I-IV à gauche et V à droite)[24]

La figure 6 montre les arbres obtenus à l'aide d'une procédure de validation croisée à cinq plis, ce qui permet d'observer les variables les plus pertinentes, même s'il faut rester prudent au vu de la grande sensibilité des arbres de décision à la variance des données. Par ailleurs, le score de compréhension orale du TOEIC étant attribué au nœud racine des cinq arbres de décision, il est considéré comme le plus discriminant. L'expérience dans un pays anglophone, quant à elle, est allouée au nœud enfant de premier niveau des arbres de décision, et constitue donc la deuxième caractéristique la plus discriminante. La troisième est incarnée par le débit de parole, qui est attribué au nœud enfant de premier ou de deuxième niveau dans chaque arbre. Globalement, bien que le niveau de précision (47 %) ne soit pas élevé, l'analyse montre que cette méthode surpasse la précision de base (25,2 %). En outre, les auteurs ont constaté que les caractéristiques de l'apprenant (score de compréhension orale du TOEIC, expérience dans un pays anglophone) se révélaient discriminantes concernant la précision de la mesure. Ce résultat suggère que ces caractéristiques devraient être prises en compte lors de la mesure de la compréhension orale des apprenants anglais.

[24] Voici une explication de la manière de lire la figure en se référant à l'exemple donné à gauche. *Listening test score* à gauche constitue le point de départ de l'analyse de l'arbre de décision. S'il est de 420 ou plus, les données se trouvent ensuite subdivisées en fonction du fait que *Multiple syllable word* dépasse 13 ou non. De cette manière, l'arbre est divisé en branches en fonction de l'importance des critères.

2.2.4.5. Kotani & Yoshimi (2019)

Dans Kotani & Yoshimi (2019), avant de déterminer subjectivement la facilité d'écoute de la phrase sur une échelle de Likert à cinq points, les apprenants ont écouté des phrases lues à haute voix par un locuteur natif de l'anglais, et les ont transcrites phrase par phrase. La différence entre cette transcription et l'originale s'appelle la distance d'édition normalisée. Les auteurs proposent de l'utiliser afin d'examiner si elle capture correctement la facilité d'écoute, en la mettant comme variable dépendante dans une analyse de régression multiple. Les variables indépendantes constituent les caractéristiques linguistiques et celle de l'apprenant. Plus concrètement, les caractéristiques linguistiques comptent le nombre moyen de mots dans une phrase, le nombre moyen de syllabes dans un mot, le taux de mots difficiles dans une phrase et le débit de parole, tandis que la caractéristique de l'apprenant correspond au score total du TOEIC. Les données ont été choisies parmi les textes distribués par l'International Phonetic Association. Les sujets de recherche regroupent 50 apprenants d'anglais comme langue étrangère à l'université. L'analyse statistique a montré que les variables indépendantes significatives comportaient le nombre moyen de mots dans une phrase, le nombre moyen de syllabes dans un mot, le taux de mots difficiles dans une phrase et le score total du TOEIC, mais pas le débit de parole. La différence entre la transcription des apprenants et la transcription originale constitue un indice de facilité d'écoute plus fiable et valide qu'un jugement subjectif.

2.2.4.6. Yoshimi & Kotani (2020)

En utilisant la même méthode d'évaluation de la facilité d'écoute que Kotani & Yoshimi (2019), Yoshimi & Kotani (2020) ont effectué une régression non linéaire avec des machines à vecteurs de support[25], en considérant la facilité d'écoute comme variable dépendante et les caractéristiques linguistiques et de l'apprenant comme variables indépendantes. La facilité d'écoute consiste en la difficulté subjective de compréhension orale des phrases découlant de l'analyse de la différence entre la transcription des apprenants et l'originale. Les

[25] Une machine à vecteurs de support (SVM) est un algorithme informatique qui apprend par l'exemple à attribuer des étiquettes aux objets (Noble, 2006: 1565).

caractéristiques linguistiques, quant à elle, comportent le nombre moyen de mots dans une phrase, le nombre moyen de syllabes dans un mot, le taux de mots difficiles dans une phrase, le nombre de mots polysyllabiques dans une phrase et le débit de parole. Enfin, la caractéristique de l'apprenant correspond au score total du TOEIC. Les données et les sujets utilisés sont les mêmes que dans Kotani & Yoshimi (2019).

Afin d'évaluer la validité de la facilité d'écoute, les auteurs ont examiné si les apprenants peuvent être différenciés en fonction de leur niveau de compétence, en se basant sur la facilité d'écoute. Après une analyse de variance grâce au test de Welch, des comparaisons multiples ont été effectuées à l'aide de la méthode de Tukey, révélant des différences significatives entre les niveaux débutant et intermédiaire, débutant et avancé, ainsi qu'intermédiaire et avancé. Par conséquent, la validité de la facilité d'écoute a été confirmée.

Il a également été constaté que la régression non linéaire tendait à obtenir de meilleures performances d'estimation de la facilité d'écoute que la régression multiple linéaire, comme cela a été montré aussi en lisibilité (François & Miltsakaki, 2012) et qu'un facteur utile pour la prédiction de la facilité d'écoute était le score total du TOEIC.

Comme écrit ci-dessus, il s'agit des études de la facilité d'écoute menées par une équipe de recherche japonaise. Diverses méthodes ont été utilisées afin de rechercher des variables efficaces concernant la facilité d'écoute. L'objectif consiste à créer un modèle basé sur les compétences individuelles, de sorte que l'inclusion des caractéristiques de l'apprenant, telles que les scores au TOEIC, en tant que variables, constitue une rupture par rapport aux recherches antérieures concernant la facilité d'écoute. En outre, il s'agit d'une étude pionnière dans la mesure où elle examine également les effets de variables phonétiques qui n'ont pas été prises en compte jusqu'à présent, à savoir le débit de parole et la modification phonologique, sur la facilité d'écoute.

2.2.5. Développement de formules avec la technologie computationnelle

Le développement de la technologie computationnelle – surtout avec le développement du traitement automatique du langage (TAL) –, permet de traiter

efficacement une grande quantité de données. Avec ce flux, les recherches à propos de l'élaboration de formule de la facilité d'écoute se développent de plus en plus. Dans cette section, nous présenterons des études menées par d'autres équipes de recherche que celle présentée dans la section précédente, et en particulier celles qui ont recours de façon importante à des technologies TAL dans leur recherche.

2.2.5.1. Révész & Brunfaut (2013)

Cette recherche examine dans quelle mesure le débit de parole, la complexité linguistique et l'explicitation (la mesure dans laquelle les idées sont explicitement exprimées) du texte influencent la compréhension orale en L2. Les textes pour la compréhension orale sont analysés en utilisant des outils analytiques automatisés. En outre, les auteurs utilisent des analyses de Rasch et de régression afin d'estimer la difficulté de la tâche et sa relation avec les caractéristiques du texte. Les participants de recherche comptent 77 étudiants d'un programme d'été d'anglais à des fins académiques dans une université britannique. Parmi eux, 68 ont effectué 18 versions d'une tâche de compréhension orale, chaque tâche étant suivie d'un questionnaire de perception. Les neuf autres étudiants ont participé à un rappel stimulé. Les langues maternelles des participants varient.

Les tâches de compréhension orale sont adoptées à partir des Graded Examinations in Spoken English du Trinity College de Londres et le niveau s'élève à C1. Chaque passage de la compréhension orale comprenait un bref récit d'environ 50 mots, tandis que la réponse attendue de la tâche consistait en un mot ou une phrase de cinq mots maximums. Par ailleurs, la moitié des tâches demandait aux participants d'écrire leurs réponses non seulement dans la langue cible, mais aussi dans leur L1, afin d'éliminer le risque que les apprenants puissent comprendre le texte mais aient des problèmes à fournir une réponse adéquate dans la L2. De plus, le questionnaire de perception comprenait huit affirmations que les participants devaient juger sur une échelle de Likert en cinq points. Ainsi, il évaluait les perceptions des participants sur :

- la difficulté globale de la tâche ;
- leur capacité à effectuer la tâche ;
- la complexité linguistique du passage à écouter ;

– la vitesse du passage ;
– l'explicitation des idées.

Le protocole de rappel stimulé comprenait trois étapes :

1. Les étudiants ont écouté une version de la tâche et ont écrit une réponse dans la langue cible.
2. Ils ont réécouté le même passage. Au cours de cette deuxième écoute, ils pouvaient mettre la cassette en pause à tout moment afin de décrire leurs pensées à un moment précis de l'écoute initiale.
3. Ils ont écouté le même passage une troisième fois. Au cours de cette dernière étape, chaque passage a été divisé en deux sous-sections plus courtes. Après chaque sous-section, les participants ont été invités à répondre à quelques questions prédéterminées (exemple : des éléments ont-ils rendu la compréhension du passage difficile ?).

L'analyse de Rasch et de régression de 18 passages de la compréhension orale en matière de débit de parole, de complexité linguistique (phonologique, lexicale, syntactique et discursive) et d'explicitation du texte montre que des indices de complexité lexicale et discursive expliquent une part importante de la variance de la difficulté des passages de la compréhension orale.

2.2.5.2. Loukina *et al.* (2016)

Cette recherche explore dans quelle mesure la difficulté des questions de compréhension orale dans un test de compétence de l'anglais peut être prédite par les propriétés textuelles du texte de la question.

Les données utilisées consistent en une grande collection d'items de la compréhension orale provenant d'un test international de compétence en langue anglaise. Le corpus comprenait quatre types d'items à choix multiple : la description d'image, la complétion de dialogue, la conversation, le monologue. Afin d'obtenir des caractéristiques textuelles, les auteurs ont utilisé un système automatisé de prédiction de la complexité des textes, TextEvaluator[26]. Ce dernier

[26] https://textevaluator.ets.org/textevaluator/.

génère 339 caractéristiques brutes basées sur des listes de vocabulaire et diverses technologies de TAL telles que le balisage et l'analyse syntaxique automatique. Les caractéristiques ont été conçues afin de mesurer le degré de complexité selon les quatre dimensions suivantes : vocabulaire, syntaxe, cohésion et discours.

Les résultats montrent que la complexité du texte de l'item constitue un prédicteur plus fort de la difficulté de l'item concernant la description d'image et la complétion de dialogue. Le niveau de difficulté de la description d'image et la complétion de dialogue, en tant que question, est dit inférieur à celui de la conversation et du monologue. Cependant, cette analyse suggère que la difficulté du passage seule peut ne pas constituer un prédicteur fort de la difficulté de l'item. Les auteurs concluent qu'une prédiction précise de la difficulté des questions nécessite un nouvel ensemble de caractéristiques qui capture l'interaction entre les passages et les questions.

2.2.5.3. Yoon *et al.* (2016)

Le but de la recherche de Yoon *et al.* (2016) consiste à proposer une méthode automatisée d'estimation de la difficulté de textes parlés pour les apprenants. Finalement, un modèle de régression linéaire multiple a été développé.

Les sujets comptent quinze personnes dont l'anglais n'est pas la langue maternelle et qui représentent diverses langues maternelles, notamment le chinois, le japonais, le coréen, le thaï et le turc. En ce qui concerne les données, les auteurs utilisent des passages tirés de tests de compréhension orale en anglais portant sur des sujets académiques ou professionnels, ainsi que des extraits de l'actualité et d'interviews, pour un total de 200 extraits. Une fois les données d'entraînement des modèles statistiques collectées, les auteurs ont défini un ensemble de variables textuelles et auditives susceptibles de prédire la facilité d'écoute des documents du corpus d'entraînement. Ensuite, les participants ont répondu à un questionnaire à propos des perceptions à la difficulté globale des textes parlés. Celui-ci était composé de cinq questions de type Likert combinées en un seul score composite lors de l'analyse. Ce score de la facilité d'écoute est utilisé comme variable dépendante pendant la construction du modèle. Il représente la moyenne des scores agrégés de trois auditeurs non natifs, un auditeur de chaque niveau de compétence (débutant, intermédiaire et avancé) dans un peu plus de détails. En effet, l'objectif

consiste à prédire la difficulté moyenne perçue des échantillons de parole chez les apprenants d'anglais aux trois niveaux différents.

Les variables utilisées à partir de trois dimensions comme variables indépendantes regroupent :

- les variables acoustiques : débit de parole en mots par seconde, nombre de silences par mot, déviation moyenne du chunk de parole sans disfluences, distance moyenne entre les syllabes accentuées en secondes, variations de la durée des voyelles ;
- les variables basées sur le vocabulaire : nombre de collocations de noms par proposition, ratios type/token, fréquence normalisée des mots de basse fréquence, fréquence moyenne des types de mots ;
- les variables grammaticales : nombre moyen de mots par phrase, nombre de phrases longues, nombre normalisé de phrases.

Les résultats d'une régression linéaire multiple ont montré qu'une combinaison de douze variables a atteint une corrélation multiple de 0,76 avec les perceptions humaines de la difficulté des textes parlés. La variable la plus efficace est la fréquence normalisée des mots peu fréquents qui mesure la difficulté lexicale des phrases. En deuxième position, se trouve le nombre normalisé de phrases, qui mesure la difficulté syntaxique, et, en troisième position, le débit de parole de l'oral. Néanmoins, la recherche de Yoon *et al.* (2016) présente un inconvénient : la méthode de calcul de chaque variable ne paraît pas clairement définie. Cependant, ces résultats semblent indiquer qu'il s'avère nécessaire d'envisager surtout les caractéristiques linguistiques du texte (par rapport au signal oral) afin d'évaluer efficacement la complexité des textes parlés.

2.2.5.4. Ruggia (2019)

La recherche de Ruggia (2019) essaye d'extraire et de décrire les caractéristiques linguistiques qui marquent un changement de niveau du CECR concernant les textes en français, à l'aide du Text Deconvolution Saliency (TDS). Elle vise également à analyser et expliquer ces résultats en les comparant aux inventaires des référentiels pour le français. Le TDS constitue un système d'apprentissage profond conçu grâce à l'élaboration d'un algorithme : le Query-

By-Dropout-Committee (QBDC) qui « sélectionne itérativement les échantillons les plus pertinents pour être ajoutés à l'ensemble d'entraînement afin que le modèle soit amélioré de façon optimale » (Ducoffe *et al.*, 2016: 158).

Le corpus utilisé comprend des textes oraux en français de niveau A1 et B2 extraits de divers manuels de français langue étrangère (FLE). Ces textes sont distingués en deux catégories de séquences discursives à l'oral, à savoir les interactions et les monologues, conformément au classement du CECR. Comme résultats de l'analyse, un dialogue est prédit A1 avec un score de 100 %, et trois exemples de monologues de B2 sont prédits B2 avec 97,37 %. L'hypothèse suivante a été donc bien testée : le TDS paraît capable d'extraire les caractéristiques de textes en français et plus précisément il est capable d'extraire les saillances qui marquent un changement de niveau selon le CECR.

2.2.5.5. Alghamdi *et al.* (2022)

L'étude de Alghamdi *et al.* (2022) s'est penchée sur les informations visuelles. Ainsi, elle porte sur certaines dimensions de la complexité linguistique afin de comprendre la manière dont celles-ci contribuent à la difficulté globale des documents audiovisuels. Les auteurs construisent un corpus de 640 vidéos dans un corpus nommé Second Language Videotext Complexity (SLVC). Pour cette étude, seulement 320 documents audiovisuels de ce corpus sont utilisés. Les documents audiovisuels déploient des conceptions pédagogiques différentes, selon les typologies de conception pédagogique des vidéoconférences fournies par Crook & Schofield (2017). Les vidéoconférences sélectionnées ont été prononcées par des locuteurs natifs de l'anglais dans les domaines disciplinaires des sciences humaines, des études sociales, de l'éducation, de l'informatique, des mathématiques, du commerce et de la gestion, ainsi que des sciences de la vie et de la médecine. Les participants comptent 322 étudiants en première année de collège, inscrits à un cours d'anglais intensif. Leur niveau d'anglais s'élève à B1. Afin d'évaluer la difficulté des vidéoconférences, les auteurs adoptent et modifient une échelle initialement développée de manière à évaluer la difficulté de la compréhension orale en L2 (Yoon *et al.*, 2016). Sur la base de ces évaluations par

des apprenants, les auteurs construisent les modèles de régression permettant de prédire les estimations subjectives de la difficulté des cours vidéo.

Les résultats ont démontré qu'une proportion significative de la variance de la difficulté des vidéoconférences ($R^2 = 0,52$) était expliquée par les caractéristiques acoustiques, lexicales et syntaxiques et que la cohésion textuelle ne contribuait pas de manière significative. Cela indique l'utilisation de caractéristiques de complexité linguistique permettant de prédire la difficulté globale d'un document audiovisuel, et soulèvent la possibilité de développer des systèmes automatisés afin de mesurer la difficulté des vidéos, semblables à ceux déjà disponibles pour estimer la lisibilité des documents écrits.

Au cours de cette période, un certain nombre d'analyses ont été effectuées en utilisant l'apprentissage automatique ainsi que le traitement statistique. Les données sur lesquelles les études sont basées s'avèrent également diverses et anticipent les développements futurs de la recherche sur la facilité d'écoute. Dès lors, une attention particulière doit être accordée à Ruggia (2019, 2020, 2021). Ruggia a déjà mis en œuvre un outil appelé DeepFLE sur l'internet[27]. Le site est accompagné de cette explication :

> *DeepFLE permet d'évaluer le niveau de textes oraux en français selon les échelles du Cadre Européen Commun de Référence pour les langues (CECRL). DeepFLE s'intègre dans le cadre du projet IDEX DeepText, qui propose la classification de textes via deep learning ainsi qu'une description des saillances apprises par l'Intelligence Artificielle (I.A.) qui marquent un changement de niveau (A1, A2, B1, B2, C1, C2).*

Comme expliqué ci-dessus, il ne s'agit pas d'une évaluation du niveau du CECRL basée sur la parole seule, mais sur la transcription. S'il existe un grand nombre de recherches sur l'anglais en matière de la facilité d'écoute, la recherche a également progressé en ce qui concerne le français – comme DeepFLE – au point que des outils deviennent accessibles aux apprenants et enseignants généralistes. Cependant, il reste des questions à régler sur la manière d'évaluer les aspects phonétiques.

---

[27] http://deeptext.unice.fr/FLE/.

2.2.6. Variables concernant la facilité d'écoute

En résumant cette section, en ce qui concerne les variables, il semble clair qu'outre les caractéristiques du lexique et des phrases, les caractéristiques phonétiques et les caractéristiques de l'apprenant influencent également la facilité d'écoute. Parmi les variables abordées, les caractéristiques des apprenants, le score de test d'anglais, ont été valides comme variables indépendantes dans toutes ces études (Ueda *et al.*, 2013, 2014 ; Kotani *et al.*, 2014 ; Kotani & Yoshimi, 2017, 2019 ; Yoshimi & Kotani, 2020). Toutefois, la prise en compte ou non de ces caractéristiques de l'apprenant dépend, bien entendu, de l'application du modèle créé. Lors de la création de modèles pour la compétence des apprenants individuels, il paraît probable qu'il s'agisse d'une variable valide, comme le montrent les recherches précédentes. Toutefois, il s'agit d'une variable inutile lors de la création de modèles qui ne s'adressent pas à des particuliers. Dès lors, il s'avère nécessaire de prendre en compte le but de la création d'un tel modèle et d'écarter les variables à considérer. En ce qui concerne les autres variables, il n'existe pas de variables spécifiques qui disposent d'un résultat unifié dans l'influence de la facilité d'écoute. Il convient donc d'examiner de près les variables classées comme caractéristiques linguistiques et phonétiques.

Afin de terminer cette section, les variables prises en compte jusqu'à présent dans la construction du modèle de facilité d'écoute sont résumées dans un tableau (tableau 1). Les variables que nous traitons dans cette thèse sont décrites en détail dans un chapitre qui suit.

Tableau 1 : Liste des variables prises en compte dans les études précédentes[28]

Variables indépendantes	Article
• Longueur de phrase	
Nombre moyen de mots dans une phrase	Chall & Dial (1948), **Allen (1952)**, Cartier (1955), Harwood (1955a), O'Keefe (1971), Denbow (1975), Kiyokawa (1990), **Ueda *et al.* (2013,**

[28] Les articles en gras sont ceux qui concluent que la variable s'avère liée à la facilité d'écoute.

	2014), Kotani *et al.* (2014), Yoon *et al.* (2016), **Kotani & Yoshimi** (2017, **2019**), Yoshimi & Kotani (2020)
Nombre de collocations de noms par clause	Yoon *et al.* (2016)
Nombre de phrases longues	Yoon *et al.* (2016)
Nombre normalisé de phrases	**Yoon *et al.* (2016)**
Nombre de phrases prépositionnelles pour 100 mots	Allen (1952)
• Longueur de mot	
Nombre de syllabes par 100 mots	Chall & Dial (1948), Allen (1952), Cartier (1955), Harwood (1955a), O'Keefe (1971)
Nombre moyen de syllabes dans un mot (nombre de syllabes/nombre de mots dans une phrase)	Ueda *et al.* (2013), Ueda *et al.* (2014), **Kotani *et al.* (2014)**, **Kotani & Yoshimi** (2017, **2019**), Yoshimi & Kotani (2020)
Nombre de syllabes au-dessus d'une par mot	Fang (1967)
Nombre de mots polysyllabiques dans une phrase (nombre de mots ayant plus de deux syllabes dans une phrase)	**Ueda *et al.* (2013)**, **Kotani *et al.* (2014)**, Kotani & Yoshimi (2017), Yoshimi & Kotani (2020)
Nombre total de syllabes -1 pour chaque mot d'une phrase	**Ueda *et al.* (2014)**
• Type de mots	
Pourcentage des « mots personnels »	Cartier (1955), Harwood (1955a)
Pourcentage des « phrases personnelles »	Cartier (1955), Harwood (1955a)
Ratio type/token	Yoon *et al.* (2016)
Fréquence moyenne des types de mots	Yoon *et al.* (2016)
• Mots difficiles	
Score de Dale	Chall & Dial (1948), Allen (1952), Denbow (1975)

Nombre moyen de mots dans un échantillonnage de 100 mots qui n'apparaissent pas sur la longue liste de Dale	Rogers (1962)
Pourcentage de mots ne figurant pas dans la liste des mots proposée par Kiyokawa (1987)	Kiyokawa (1990)
Taux de mots difficiles dans une phrase (nombre des mots ne figurant pas sur la liste de mots de Kiyokawa (1990)/nombre total de mots dans une phrase)	Ueda *et al.* (2013, 2014), **Kotani *et al.* (2014)**, **Kotani & Yoshimi** (2017, **2019**), Yoshimi & Kotani (2020)
Fréquence normalisée des mots de basse fréquence	**Yoon *et al.* (2016)**
• Autres variables textuelles	
Longueur moyenne de l'unité d'idée (nombre de mots de la transcription/nombre de clauses indépendantes)	Rogers (1962)
Complexité linguistique (lexicale) (fréquence lexicale, densité lexicale, diversité lexicale, concrétude des mots du contenu)	**Révész & Brunfaut (2013)**
Complexité linguistique (syntactique) (complexité structurelle, fréquence des expressions négatives)	Révész & Brunfaut (2013)
Complexité linguistique (discursive)	**Révész & Brunfaut (2013)**
Explicitation du texte	Révész & Brunfaut (2013)
Vocabulaire (vocabulaire académique, non familiarité des mots, concrétude)	Loukina *et al.* (2016)
Syntaxe (complexité syntaxique)	Loukina *et al.* (2016)
Cohésion (cohésion lexicale, niveau d'argumentation)	Loukina *et al.* (2016)

Discours (degré de narrativité, style interactif/conversationnel)	Loukina *et al.* (2016)
• Phonétique	
Débit de parole (nombre de mots par minute ou par seconde)	Harwood (1955b), **Kotani *et al.* (2014)**, **Yoon *et al.* (2016)**, **Kotani & Yoshimi** (**2017**, 2019), Yoshimi & Kotani (2020)
Taux d'élision dans une phrase (nombre d'élision/nombre total de mots dans une phrase)	Kotani *et al.* (2014), Kotani & Yoshimi (2017)
Taux de réduction dans une phrase (nombre de réductions/nombre total de mots dans une phrase)	Kotani *et al.* (2014), Kotani & Yoshimi (2017)
Taux de contraction dans une phrase (nombre de contractions/nombre total de mots dans une phrase)	**Kotani *et al.* (2014)**, Kotani & Yoshimi (2017)
Taux de liaison dans une phrase (nombre de liaisons/nombre total de mots dans une phrase)	**Kotani *et al.* (2014)**, Kotani & Yoshimi (2017)
Taux de déduction dans une phrase (nombre de déductions/nombre total de mots dans une phrase)	**Kotani *et al.* (2014)**, Kotani & Yoshimi (2017)
Complexité linguistique (phonologique) (probabilité phonotactique, fréquence des élisions, rythme, hauteur)	Révész & Brunfaut (2013)
Nombre de silences par mot	Yoon *et al.* (2016)
Déviation moyenne du chunk de parole	Yoon *et al.* (2016)
Variations de la durée des voyelles	Yoon *et al.* (2016)
• Apprenants	
Score de compréhension orale du TOEIC	**Ueda *et al.* (2013, 2014)**, **Kotani *et al.* (2014)**, **Kotani & Yoshimi (2017)**
Score total du TOEIC	**Kotani & Yoshimi (2019)**, **Yoshimi & Kotani (2020)**

Nombre de mois pendant lesquels les apprenants ont étudié l'anglais	Kotani & Yoshimi (2017)
Nombre de mois passés par les apprenants dans des pays anglophones	**Kotani & Yoshimi (2017)**
Score sur une échelle de Likert en cinq points pour la fréquence de l'utilisation de l'anglais (1, rarement ; 2, assez rarement ; 3, modérément ; 4, assez fréquemment ; et 5, fréquemment)	Kotani & Yoshimi (2017)

Il n'existe que quelques recherches antérieures sur la facilité d'écoute du français ; nous avons essentiellement décrit des études sur la facilité d'écoute de l'anglais. Comme cette thèse porte sur le français, nous ne savons pas si les variables indiquées ci-dessus sont pertinentes concernant la facilité d'écoute en français. Cependant, il serait intéressant de lancer une étude sur la facilité d'écoute en français en s'inspirant de ces variables.

2.3. Conclusion

Ce chapitre a défini le terme clé de facilité d'écoute et a parcouru l'histoire de sa recherche dans l'ordre chronologique. La recherche sur la facilité d'écoute renvoie largement à la recherche sur la lisibilité menée avant elle, de sorte que nous avons commencé par une vue d'ensemble de la recherche sur la lisibilité. Ensuite, grâce au développement de la technologie informatique actuelle, nous avons présenté comment, à partir d'ensembles de données vastes et diversifiés, des analyses multidimensionnelles ont été effectuées. Dans ces analyses, ont été prises en compte non seulement les variables qui ne peuvent être obtenues qu'à partir des transcriptions, mais aussi diverses autres variables, telles que les variables phonétiques.

Cependant, en ce qui concerne l'étude de la facilité d'écoute du français, beaucoup de recherches restent à faire : certaines études, comme celle de Ruggia (2019, 2020, 2021), ont été menées pour le français, mais les résultats concernant l'enseignement et l'apprentissage ne s'avèrent pas précis. De même, il existe des outils permettant l'évaluation automatique des textes oraux, tels que DeepFLE, mais ils sont encore limités à la prédiction de « textes » oraux. Par conséquent, il paraît nécessaire d'examiner l'universalité des résultats de prédiction dans les études sur la facilité d'écoute en français, tout en examinant l'efficacité de l'éducation, et d'analyser plus en détail les variables acoustiques.

3. Méthode

Le chapitre précédent a passé en revue les recherches antérieures concernant la facilité d'écoute et en a résumé l'historique, afin de confirmer l'utilité du développement de la recherche sur la facilité d'écoute. Ce chapitre décrira la méthodologie de cette étude sur la facilité d'écoute : les questions de recherche seront présentées à la section 3.1. et les données utilisées à la section 3.2. Puis, l'étude étant divisée en deux phases d'analyse, la section 3.3. offrira un aperçu des méthodes et outils analytiques utilisés dans chaque phase, tandis que la section 3.4. fournira une explication des variables prises en compte dans cette étude. Enfin, nous concluerons ce chapitre à 3.5.

3.1. Questions de recherche

Dans cette thèse, des questions de recherche sont posées de manière à permettre le développement d'une recherche en langue française sur la facilité d'écoute. En plus d'envisager un modèle prédictif de la facilité d'écoute, il s'avère nécessaire d'identifier les variables liées à la facilité d'écoute du français. Il paraît également utile d'examiner les problèmes à surmonter afin de créer un modèle prédictif efficace de la facilité d'écoute et d'être en mesure de faire de telle prédiction de manière automatique. Ainsi, les quatre principales questions de recherche de cette thèse sont :

1. Quelles sont les variables qui sont associées à la facilité d'écoute pour le français langue étrangère ?
2. Les caractéristiques acoustiques contribuent-elles à l'amélioration des performances au niveau de la prédiction de la facilité d'écoute ?
3. Les performances au niveau de la prédiction de la facilité d'écoute varient-elles en fonction du style de parole ?
4. Est-il possible d'automatiser la prédiction de la facilité d'écoute ?

Afin de les clarifier, nous avons réalisé des études en deux étapes. Tout d'abord, grâce aux données transcrites et annotées manuellement, nous avons construit notre premier modèle de la facilité d'écoute. Ensuite, un second modèle a été établi à l'aide des données transcrites et annotées automatiquement. Enfin, en comparant

la performance des deux modèles, nous avons évaluer la possibilité d'automatiser complètement ce modèle, ainsi que l'impact de cette automatisation sur les performances. Avant de détailler la méthodologie de recherche à chaque étape, la section suivante décrira les données utilisées.

3.2. Données

Les données sont basées sur des enregistrements audios et leurs transcriptions fournis avec 25 manuels de FLE (voir tableau 2). Les principaux manuels couramment utilisés dans les cours de FLE ont été sélectionnés comme sources. Nous avons sélectionné des manuels de différents niveaux de compétence, utilisant pour ce faire l'échelle du CECR[29] qui spécifie six niveaux de compétences : A1, A2, B1, B2, C1 et C2. Toutefois, dans notre étude, les niveaux C1 et C2 ont été regroupés au vu du petit nombre de manuels existant.

En ce qui concerne les transcriptions, les données ont été collectées en scannant les pages de transcriptions orthographiques se retrouvant à la fin des manuels. Elles ont ensuite été soumises à la Reconnaissance Optique de Caractères (OCR) et enregistrées sous la forme de fichiers texte (.txt). S'agissant des audios, les données fournies avec le manuel sont converties en fichiers wav. Par ailleurs, le contenu d'une piste audio est essentiellement traité comme une seule donnée. Toutefois, si une piste contient, par exemple, plusieurs dialogues dans différents contextes, l'audio a été divisé manuellement. Ainsi, les données prennent toujours la forme d'une paire de fichiers, reprenant l'audio et la transcription, lesquels portent le même nom (mais ont une extension différente). Afin d'éviter que la performance du modèle à chaque niveau du CECR soit biaisée par le nombre de données, les données ont été presque équilibrées en matière de nombre de données à chaque niveau. Il existe des différences dans le nombre de données par manuel, mais cela est dû au nombre d'audios fournis avec chaque manuel. Aux niveaux A1 et A2, de courts audios sont principalement présents dans les manuels. Un nombre de données légèrement plus important a été sélectionné pour les niveaux A1 et A2 afin d'assurer un équilibre en matière de longueur des contenus avec les autres niveaux.

Tableau 2 : Liste des manuels

Manuel	Éditeur	Année	Nombre de données	Durée d'enregistrement (min)
A1				

[29] Les détails sont donnés au 3.4.1.

Alter Ego	Hachette	2006	26	Environ 17
écho	CLE International	2011	63	Environ 22
Texto	Hachette	2016	19	Environ 14
Cosmopolite	Hachette	2017	104	Environ 81
Atelier	Didier	2019	78	Environ 25
A2				
Entre Nous	Maison des langues	2016	70	Environ 23
Tendances	CLE International	2016	81	Environ 54
Texto	Hachette	2016	16	Environ 22
Cosmopolite	Hachette	2017	72	Environ 74
Défi	Maison des langues	2018	35	Environ 34
Atelier	Didier	2019	35	Environ 24
B1				
Entre Nous	Maison des langues	2016	72	Environ 28
Tendances	CLE International	2016	90	Environ 63
Cosmopolite	Hachette	2018	51	Environ 62
Défi	Maison des langues	2019	27	Environ 38
B2				
Édito	Didier	2006	31	Environ 35
LE DELF B2	Didier	2016	60	Environ 70
Entre Nous	Maison des langues	2017	20	Environ 14
Tendances	CLE International	2017	50	Environ 71
Cosmopolite	Hachette	2019	80	Environ 100
C1/C2				

Réussir le DALF	Didier	2007	31	Environ 106
abc DALF	CLE International	2014	57	Environ 291
Le DALF 100 % réussite	Didier	2017	75	Environ 118
Édito	Didier	2018	70	Environ 133
Tendances	CLE International	2019	10	Environ 64

Les enregistrements audios non retenus dans nos données regroupent plusieurs éléments :

1. Ceux qui ne contiennent que des mots, des chiffres ou des phrases nominales courtes ;
2. Les exercices de grammaire qui n'ont aucun sens en tant que phrases, tels que les exercices de conjugaison ;
3. Les données figurant dans les premières et dernières unités dans chaque manuel.

Les données correspondant aux deux premiers critères ont été exclues parce qu'elles ne s'avèrent pas appropriées aux objectifs de cette étude. Par ailleurs, en général, la première unité d'un manuel constitue souvent une révision d'un niveau inférieur tandis que la dernière unité est souvent déjà plus typique du niveau suivant. Ainsi, afin de maintenir une certaine homogénéité des données considérées à chaque niveau, les données correspondant au troisième critère ont été exclues. En outre, les consignes sont exclues dans nos données puisqu'elles ne sont pas liées au niveau du CECR.

Nous signalons également que toutes ces données ont été classées manuellement selon deux styles de parole : dialogue et monologue. Les données classées dans le style dialogue correspondent à celles pour lesquelles un ou plusieurs interlocuteurs sont présents dans le discours, par exemple la conversation quotidienne, l'interview dans le média. Les données classées dans le style monologue, quant à elles, comportent celles pour lesquelles il n'existe pas d'interlocuteur dans le

discours, par exemple les documents explicatifs, les annonces au public, les messages des répondeurs téléphoniques, etc.

Par ailleurs, les données dont le contenu informatif s'avère extrêmement élevé ou faible dans une certaine catégorie de niveau et de style de parole risquent de réduire l'homogénéité lors de l'examen des caractéristiques de ce niveau et de style de parole. Pour cette raison, en utilisant les syllabes par minute sans les pauses et les mots par minute sans les pauses[30], nous avons décidé d'omettre les données si l'une ou l'autre de ces variables comportait des valeurs aberrantes extrêmes. Ces variables font partie des variables indépendantes prises en compte lors de l'élaboration de notre modèle afin de prédire la facilité d'écoute. Comme cette thèse traite des documents audios et que les variables se trouvent liées à la prosodie et au volume de la parole pure, qui exclut les pauses, les syllabes par minute sans les pauses et les mots par minute sans les pauses ont été utilisées comme méthode de mesure du contenu informatif dans cette thèse. Les cinq données suivantes ont été supprimés pour leurs valeurs aberrantes extrêmes.

Tableau 3 : Données qui ont des valeurs aberrantes extrêmes

	Nom de données	Style de parole	Syllabes par minute sans les pauses	Mots par minute sans les pauses
1	B2_Entre_Nous_28_3	Monologue	341,81	488,3
2	C1_Edito_5_2	Monologue	433,78	367,05
3	C2_Reussir_le_Dalf_16_2	Monologue	144,09	78,36
4	C2_Reussir_le_Dalf_17	Dialogue	114,29	77,13
5	C2_Reussir_le_Dalf_18	Dialogue	160,8	97,2

Dans le Tableau 3, les valeurs soulignées représentent des valeurs aberrantes extrêmes. Des statistiques descriptives ont été réalisées à l'aide du logiciel Statistical Package for the Social Sciences (SPSS) pour chaque niveau et style de parole et définies selon la manière SPSS d'exprimer les valeurs extrêmes. Les valeurs extrêmes élevées correspondent à celles qui sont supérieures à Q3 + 3 * (Q3 – Q1) et les valeurs extrêmes basses, à celles qui sont inférieures à

[30] Les détails sont donnés à la section 3.4.3.

Q1 – 3 * (Q3 – Q1) [31] . Les moyennes et les critères concernant les valeurs extrêmes en syllabes par minute sans les pauses et en mots par minute sans les pauses pour chaque niveau et style de parole sont présentés dans le tableau 4 et le tableau 5.

Tableau 4 : Critères pour les valeurs extrêmes pour le dialogue

	Syllabes par minute sans les pauses			Mots par minute sans les pauses		
Niveau	Moyenne	Critères de valeurs extrêmes élevées	Critères de valeurs extrêmes faibles	Moyenne	Critères de valeurs extrêmes élevées	Critères de valeurs extrêmes faibles
A1	258,81	477,03	41,67	192,67	397,64	-12,43
A2	289,21	492,91	87,09	210,77	397,37	20,99
B1	312,69	463,66	168,62	234,76	410,94	59,78
B2	320,51	471,47	170,33	239,1	425,98	46,24
C1/C2	313,58	435,86	193,76	221,41	356,83	84,31

[31] Lorsqu'on classe les données par ordre croissant, le quart des données les plus basses correspond au premier quartile (Q1), tandis que les trois quarts des données correspondent au troisième quartile (Q3).

Tableau 5 : Critères pour les valeurs extrêmes pour le monologue

	Syllabes par minute sans les pauses			Mots par minute sans les pauses		
Niveau	Moyenne	Critères de valeurs extrêmes élevées	Critères de valeurs extrêmes faibles	Moyenne	Critères de valeurs extrêmes élevées	Critères de valeurs extrêmes faibles
A1	274,84	546,78	10,14	186,29	365,56	9,61
A2	316,94	608,11	28,54	232,02	486,96	-33,49
B1	330,85	567,9	94,34	239,01	485,62	-9,61
B2	318,7	471,83	165,82	225,47	359,96	75,25
C1/C2	326,9	461,39	190,22	216,98	348,76	80,57

En examinant le tableau 3 par rapport au tableau 4 et au tableau 5, nous constatons que les données 1 et 2 du tableau 3 constituent des données présentant des valeurs extrêmes élevées, tandis que les autres données 3, 4 et 5 correspondent à des données présentant des valeurs extrêmes faibles. Lorsque les données sources principales ont été vérifiées, les données 1 et 2 constituaient des données très courtes et peu informatives, chacune consistant en deux phrases seulement. Les données 3, 4 et 5, quant à elles, correspondaient à des données pour lesquelles de la musique était insérée dans le discours, chacune avec un temps d'enregistrement inutilement long. Ainsi, en examinant les données en fonction des deux indicateurs, les syllabes par minute sans les pauses et les mots par minute sans les pauses, il a été possible de mettre en évidence des données qui ne devaient pas être prises en compte dans l'analyse.

De plus, comme nous le détaillerons à la section 3.4.2., cette étude utilise l'outil FABRA afin d'analyser les transcriptions. Malheureusement, le code de FABRA a rencontré des difficultés à paramétrer sept transcriptions de niveau C. Ce problème n'est pas examiné en détail dans cette étude, car il s'agit d'un problème technique avec FABRA. Cependant, comme cette étude traite l'audio et les transcriptions en tant que paire de données, cela créé un déséquilibre s'il existe des données dont les caractéristiques des transcriptions ne peuvent pas être extraites. Par conséquent, ces sept données pour lesquelles FABRA ne fonctionne pas correctement ont également été exclues de l'étude. Enfin, le nombre total de

nos paires de données (audio et transcription) s'élève à 1 323 (voir les tableaux 6, 7 et 8)[32].

Tableau 6 : Description du corpus

Niveau	Nombre de données	Nombre de mots		Durée d'enregistrement (min.)	
	Total	Total	Moyenne	Total	Durée moyenne par audio
A1	290	18 049	62,24	Environ 160	0,55
A2	309	32 191	104,18	Environ 232	0,75
B1	240	31 201	130	Environ 191	0,8
B2	241	49 777	206,54	Environ 290	1,2
C1/C2	243	123 114	506,64	Environ 711	2,93
Total	1 323	254 332	192,24	Environ 1 584	1,2

Tableau 7 : Description des données (dialogue)

Niveau	Nombre de données	Nombre de mots		Durée d'enregistrement (min.)	
	Total	Total	Moyenne	Total	Durée moyenne par audio
A1	205	13 862	67,62	Environ 120	0,59
A2	180	24 468	135,93	Environ 180	1
B1	146	24 105	165,1	Environ 147	1
B2	131	29 497	225,17	Environ 167	1,27
C1/C2	121	92 906	767,82	Environ 520	4,3
Total	783	184 838	236,06	Environ 1 134	1,45

[32] Le tableau 2 indique également la répartition des 1 323 données qui ont finalement été utilisées.

Tableau 8 : Description des données (monologue)

Niveau	Nombre de données	Nombre de mots		Durée d'enregistrement (min.)	
	Total	Total	Moyenne	Total	Durée moyenne par audio
A1	85	4 187	49,26	Environ 40	0,47
A2	129	7 723	59,87	Environ 52	0,4
B1	94	7 096	75,49	Environ 44	0,47
B2	110	20 280	184,36	Environ 123	1,12
C1/C2	122	30 208	247,61	Environ 191	1,57
Total	540	69 494	128,69	Environ 450	0,83

3.3. Méthodologie

Afin de clarifier les questions de recherche présentées à la section 3.1., une analyse est effectuée à l'aide des données présentées à la section 3.2. L'analyse est menée en deux étapes : un premier modèle de la facilité d'écoute est développé et évalué ; ensuite, un second modèle est mis au point comme point d'étape vers l'automatisation de notre modèle de facilité d'écoute. Cette section précisera la méthodologie de recherche pour chaque étape.

3.3.1. Premier modèle de la facilité d'écoute

L'objectif de la première étape d'analyse consiste à construire un premier modèle de la facilité d'écoute. Dans cette étape, nous avons utilisé les données transcrites et annotées manuellement. Les données de transcription, collectées comme indiqué à la section 3.2., ont été utilisées après la correction manuelle permettant de corriger les fautes d'orthographe, etc., conformément aux transcriptions originales des manuels. Les données audios également collectées comme indiqué à la section 3.2. ont été annotées et alignées à l'aide de SPPAS afin de réaliser l'analyse.

SPPAS constitue un outil permettant de produire des annotations automatiques comprenant des segmentations d'énoncés, de mots, de syllabes et de phonèmes à partir d'un enregistrement audio et de sa transcription (Bigi, 2012: 1748)[33]. Sur la base des transcriptions, qui ont été modifiés manuellement, du fichier texte, et de l'audio du fichier wav, un alignement est d'abord effectué. Les résultats de l'alignement sont restitués dans le format TextGrid disponible dans le logiciel Praat, ce qui permet de les visualiser[34]. Tout d'abord, l'alignement est réalisé en le segmentant au niveau des pauses, et les résultats se trouvent affichés dans Praat. En observant l'alignement par rapport au son et à l'alignement, nous avons vérifié manuellement si l'alignement et l'insertion de transcriptions dans cette section par SPPAS est bien correct ou pas. La précision de cet alignement a ensuite été assurée en effectuant des segmentations syllabiques et phonémiques après cette

[33] http://www.sppas.org.
[34] https://www.praat.org.

vérification. Ensuite, les données sont à nouveau exportées au format TextGrid, afin d'être utilisées pour l'analyse.

En ce qui concerne la performance de SPPAS, Bigi & Meunier (2018) ont utilisé des discours lus et spontanés en français afin d'évaluer la performance de l'outil dans la conversion graphème-phonème, ainsi que pour la détection des pauses remplies, des rires et des bruits. Les résultats ont montré une précision de 96,09 % pour le discours lu et de 96,48 % pour le discours spontané. Cela signifie qu'une performance élevée a été obtenue, quel que soit le type de discours. Dès lors, l'utilisation d'un outil d'annotation automatique aussi performant paraît justifiée et s'avère importante pour cette thèse.

En vue de définir une pause, SPPAS définit un silence de plus de 200 ms comme une pause. Tout d'abord, cette durée semble aisément mesurable grâce au logiciel Praat (Sauzedde, 2016: 299). Ensuite, ce seuil de 200 ms correspond au seuil minimum de cadences aisément perceptibles et comptables par les humains, ainsi qu'à celui du tempo moteur spontané, c'est-à-dire le tempo minimal auquel les humains peuvent reconnaître un rythme et y répondre naturellement, constaté expérimentalement chez les humains (Candea, 2000: 22).

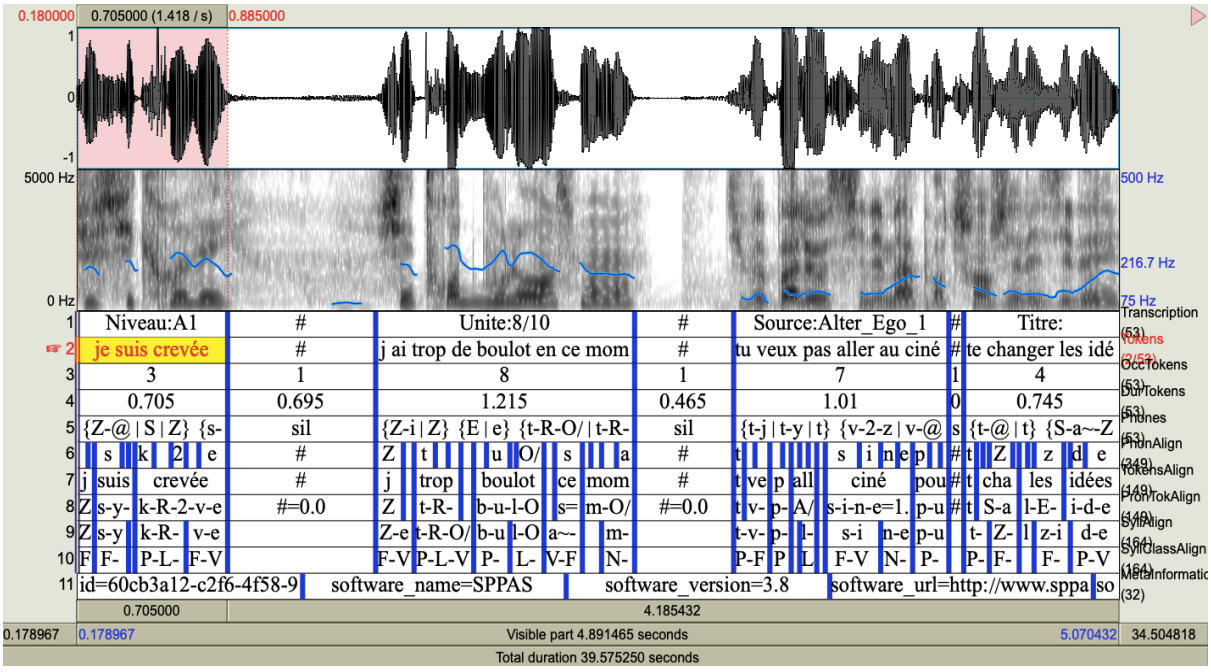

Figure 7 : Exemple de sortie dans Praat

Dans le cadre du développement de ce premier modèle, qui repose sur les variables de lisibilité de FABRA et les variables de fluidité basées sur l'annotation

décrite précédemment, et dont nous discuterons plus tard les détails, nous avons envisagé un modèle capable de prédire le niveau de facilité d'écoute d'un document en fonction de l'échelle du CECR. Pour ce faire, nous avons fait appel à des machines à vecteur de support (SVM). Les SVM produisent un algorithme statistique capable de pondérer des variables afin de séparer les données en niveaux de manière linéaire et non linéaire. Cela signifie que chaque variable peut être analysée et que les niveaux peuvent être identifiés, ce qui convient à l'objectif de cette recherche.

Afin d'approfondir leur évaluation, les modèles sont créés non seulement à l'aide de SVM, mais aussi de deux modèles de langues : wav2vec 2.0 et CamemBERT, lesquels sont ensuite comparés aux modèles créés à l'aide de SVM. Wav2vec 2.0 (désormais wav2vec) est utilisé pour l'analyse spécialisée dans le domaine de la prosodie, tandis que CamemBERT l'est pour les prédictions à partir des transcriptions. Wav2vec représente un modèle de reconnaissance vocale développé par Baevski *et al.* (2020). L'emploi de ce modèle en facilité d'écoute représente une innovation par rapport au domaine, et vise à capturer au mieux la richesse du signal vocal. En outre, ce modèle traite la forme d'onde brute du signal vocal avec un réseau neuronal convolutionnel multicouche[35] afin d'obtenir des représentations audios latentes (*Z* dans la figure 8). Ces représentations sont ensuite introduites dans un quantificateur et un transformeur. Le quantificateur choisit une unité vocale pour la représentation audio latente à partir d'un inventaire d'unités apprises (*Q* dans la figure 8). Environ la moitié des représentations audio sont masquées avant d'être introduites dans le transformeur. Le transformeur, quant à lui, ajoute des informations provenant de l'ensemble de la séquence audio. Enfin, la sortie du transformeur (*C* dans la figure 8) est utilisée pour résoudre une tâche contrastive (*L* dans la figure 8). Cette tâche exige que le modèle identifie les unités de parole quantifiées correctes pour les positions masquées. Ainsi, wav2vec possède la caractéristique de pouvoir quantifier les données audios afin de les rendre plus faciles à traiter par un ordinateur. En appliquant cette technologie de quantification, nous avons créé un modèle de prédiction permettant d'évaluer la facilité d'écoute. Des étiquettes correspondant

[35] Il s'agit d'un algorithme principalement utilisé dans la reconnaissance d'images et de la parole. Il apprend les modèles de manière hiérarchique à travers plusieurs couches et capture les caractéristiques.

aux niveaux du CECR sont attribuées aux valeurs numériques, et ce modèle apprend ces relations pour prédire les niveaux du CECR.

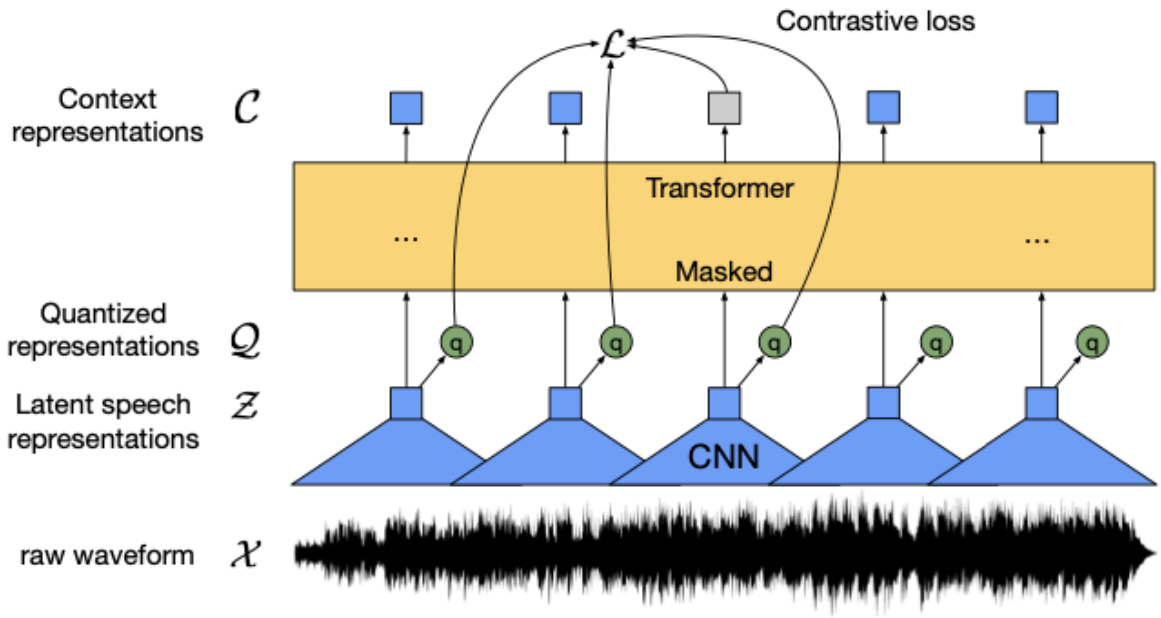

Figure 8 : Structure de wav2vec (Baevski *et al.*, 2020: 2)

Pour la version française, nous avons utilisé le modèle *facebook/wav2vec2-large-xlsr-53-french*[36], qui constitue un modèle wav2vec entraîné sur des données audios non annotées de douze langues provenant du benchmark Common Voice[37].

En revanche, CamemBERT correspond à un modèle de langue pour le français basé sur l'architecture RoBERTa pré-entraîné sur le sous-corpus français du corpus multilingue OSCAR (Martin *et al.*, 2020). RoBERTa constitue une version améliorée du modèle de langue BERT, qui génère, au niveau des phrases, des représentations intégrées, représentant le sens des phrases en fonction des phrases d'entrée, sous la forme de vecteurs numériques (Devlin *et al.*, 2018). BERT utilise un objectif de pré-entraînement de « modèle de langage masqué » (MLM). Ce dernier sélectionne aléatoirement certains des tokens de l'entrée afin de les cacher, et l'objectif consiste à prédire l'identifiant de vocabulaire original du mot caché en se basant uniquement sur son contexte. En outre, l'objectif MLM permet à la représentation de fusionner les contextes gauche et droit, ce qui permet de pré-entraîner un transformeur bidirectionnel profond. BERT utilise également la tâche

[36] https://huggingface.co/facebook/wav2vec2-large-xlsr-53-french.
[37] https://commonvoice.mozilla.org/fr/about.

de « prédiction de la phrase suivante » qui permet de pré-entraîner conjointement les représentations des paires de textes. RoBERTa s'appuie sur la stratégie de masquage linguistique de BERT, dans laquelle le système apprend à prédire des sections de texte intentionnellement cachées dans des exemples de langage par ailleurs non annotés (Liu *et al.*, 2019). Avec CamemBERT, tout comme avec wav2vec expliqué dans la section précédente, il s'avère possible de convertir une phrase en valeurs numériques, cette fois, en s'appuyant sur le contenu textuel. En utilisant ces valeurs, il devient possible construire un modèle de prédiction des niveaux du CECR, en appliquant le même principe que celui démontré avec wav2vec. Enfin, l'utilisation de CamemBERT pour les données de transcription utilisées dans cette étude permet la création de modèles de langage au niveau sémantique de la phrase.

3.3.2. Second modèle : vers l'automatisation

L'objectif de la deuxième étape d'analyse consiste à construire un second modèle qui tend à une automatisation complète. Cela signifie que, dans cette étape, nous avons utilisé des données transcrites et annotées automatiquement. Ce second modèle correspond à une situation pédagogique réelle, par exemple dans laquelle un professeur de FLE trouve un document audio sans transcription associée et doit en évaluer le niveau de facilité d'écoute. Dès lors, la transcription est obtenue à l'aide d'un processus automatique et sans correction manuelle postérieure. Pour la transcription automatique, nous avons utilisé Speech-to-Text de Google.

Speech-to-Text constitue une interface de programmation d'applications de transcription à partir de l'audio[38]. Xu *et al.* (2021) ont évalué les quatre API[39] Speech-to-Text les plus utilisées sur le discours français provenant de quatre Computing Cloud : Amazon Transcribe, Google Cloud, IBM Waston et Microsoft Azure. Cinq niveaux de bruits d'environnement différents se trouvent mélangés à 6 690 discours propres (pour un total de 17 heures). Ainsi, 100 heures de tests de discours par API Speech-to-Text ont donné 400 heures de transcription de discours. D'après Xu *et al.* (2021), notamment lorsque le travail est soumis à des contraintes de temps, Google Cloud reste le premier choix avec un temps de réponse plus

[38] https://cloud.google.com/speech-to-text.
[39] Alphabet Phonétique International.

rapide et un taux d'erreur de mot raisonnable. Dès lors, cela indique que Google Cloud offre un niveau de performance élevé, et qu'il semble approprié d'utiliser Speech-to-Text de Google afin de convertir automatiquement l'audio français en texte dans cette thèse.

Nous avons utilisé également SPPAS afin d'annoter l'audio sur la base de cette transcription, mais nous n'avons, cette fois, pas corrigé l'alignement. Nous développons sur cette base un modèle capable de prédire le niveau du CECR à l'aide du SVM. Wav2vec et CamemBERT sont également utilisés comme modèles alternatifs. Par la suite, les résultats obtenus aux deux étapes ont été comparées.

Concernant cette comparaison, nous répondons à ces deux sous-questions :

1. Existe-t-il une différence de performance entre le premier et le second modèle ?
2. Quels sont les inconvénients du traitement automatique des données ?

Sur la base de ces sous-questions de recherche, nous avons évalué un modèle de facilité d'écoute vers l'automatisation.

3.4. Variables

Dans cette thèse, le niveau de facilité d'écoute évalué en fonction de l'échelle du CECR constitue la variable dépendante, tandis que le modèle permettant de prédire cette facilité d'écoute s'appuie sur deux familles de variables indépendantes principales : les variables de lisibilité et les variables de fluidité. Ces variables se trouvent prises en compte dans les deux étapes d'analyse susmentionnées afin de construire les modèles SVM. Cette section sera consacrée à l'explication plus détaillée de ces variables.

3.4.1. CECR

Comme mentionné précédemment, cette étude utilise le niveau du CECR comme variable dépendante dans la création d'un modèle visant à prédire la facilité d'écoute. Autrement dit, nous proposons un modèle qui exprime la facilité d'écoute en matière de niveaux du CECR et permet de prédire le niveau du CECR du discours. Dans le chapitre 2, de nombreuses recherches antérieures sur la facilité d'écoute ont été abordées, dans lesquelles la variable dépendante modélisant la facilité d'écoute correspondait aux niveaux scolaires ou à la difficulté subjective de compréhension orale des phrases. Dans cette étude, les niveaux de l'échelle du CECR, largement utilisés pour l'apprentissage et l'enseignement du français, ont été choisis comme variable dépendante afin de développer à l'avenir un outil automatique de prédiction de la facilité d'écoute, accessible aussi bien aux apprenants qu'aux enseignants de français. Exprimer la facilité d'écoute à travers les niveaux du CECR présente l'avantage d'être facilement applicable dans diverses situations, telles que l'identification d'un discours adapté aux compétences des apprenants ou aux objectifs pédagogiques des cours.

Le CECR – Cadre européen commun de référence pour les langues –, « offre une base commune pour l'élaboration de programmes de langues vivantes, de référentiels, d'examens, de manuels, etc. en Europe. Il définit les niveaux de compétence qui permettent de mesurer le progrès de l'apprenant à chaque étape de l'apprentissage et à tout moment de la vie » (Conseil de l'Europe, 2001: 9). Dans la description des compétences en compréhension à l'oral dans le CECR, une

échelle est proposée afin d'illustrer la compréhension générale de l'oral et les sous-échelles suivantes sont définies :

– comprendre une interaction entre locuteurs natifs ;
– comprendre en tant qu'auditeur ;
– comprendre des annonces et instructions orales ;
– comprendre des émissions de radio et des enregistrements.

Les descriptions, niveau par niveau, de chacune de ces échelles sont les suivantes (Conseil de l'Europe, 2001: 54-56)[40] :

Compréhension générale de l'oral	
C2	Peut comprendre toute langue orale qu'elle soit en direct ou à la radio et **quel qu'en soit le débit**.
C1	Peut suivre une intervention d'une certaine longueur sur des **sujets abstraits** ou complexes même hors de son domaine mais peut avoir besoin de faire confirmer quelques détails, notamment si l'accent n'est pas familier. Peut reconnaître **une gamme étendue d'expressions idiomatiques et de tournures courantes** en relevant les changements de registre. Peut suivre une intervention **d'une certaine longueur** même si **elle n'est pas clairement structurée** et même si les relations entre les idées sont seulement implicites et non explicitement indiquées.
B2	Peut comprendre **une langue orale standard** en direct ou à la radio sur des **sujets familiers et non familiers** se rencontrant normalement dans la vie personnelle, sociale, universitaire ou professionnelle. Seul **un très fort bruit de fond**, une structure inadaptée du discours ou **l'utilisation d'expressions idiomatiques** peuvent influencer la capacité à comprendre.
	Peut comprendre les idées principales d'interventions complexes du point de vue du fond et de la forme, sur **un sujet concret ou abstrait** et **dans une langue standard**, y compris des discussions techniques dans son domaine de spécialisation.

[40] Tous les textes en gras dans les descriptions sont de l'auteur.

	Peut suivre une intervention **d'une certaine longueur** et une argumentation complexe à condition que **le sujet soit assez familier** et que le plan général de l'exposé soit indiqué par des marqueurs explicites.
B1	Peut comprendre une information factuelle directe sur **des sujets de la vie quotidienne ou relatifs au travail** en reconnaissant les messages généraux et les points de détail, à condition que **l'articulation soit claire** et l'accent courant.
	Peut comprendre les points principaux d'une intervention sur **des sujets familiers** rencontrés régulièrement au travail, à l'école, pendant les loisirs, y compris **des récits courts**.
A2	Peut comprendre assez pour pouvoir répondre à des besoins concrets à condition que **la diction soit claire** et **le débit lent**.
	Peut comprendre **des expressions et des mots porteurs de sens relatifs à des domaines de priorité immédiate** (par exemple, information personnelle et familiale de base, achats, géographie locale, emploi).
A1	Peut comprendre une intervention si elle est **lente** et **soigneusement articulée** et comprend de **longues pauses** qui permettent d'en assimiler le sens.

Ensuite, nous avons vérifié la description des quatre sous-échelles.

Comprendre une interaction entre locuteurs natifs	
C2	Comme C1
C1	Peut suivre facilement des échanges complexes entre des partenaires extérieurs dans une discussion de groupe et un débat, même sur **des sujets abstraits**, complexes et **non familiers**.
B2	Peut réellement suivre une conversation animée entre locuteurs natifs.
	Peut saisir, avec un certain effort, une grande partie de ce qui se dit en sa présence, mais pourra avoir des difficultés à effectivement participer à une discussion avec plusieurs locuteurs natifs qui ne modifient en rien leur discours.
B1	Peut également suivre les points principaux d'une longue discussion se déroulant en sa présence, à condition que **la langue soit standard** et **clairement articulée**.

A2	Peut généralement identifier le sujet d'une discussion se déroulant en sa présence si l'échange est mené **lentement** et si **l'on articule clairement**.
A1	Pas de descripteur disponible.

Comprendre en tant qu'auditeur	
C2	Peut suivre une conférence ou un exposé spécialisé employant de nombreuses formes relâchées, des régionalismes ou **une terminologie non familière**.
C1	Peut suivre la plupart des conférences, discussions et débats avec assez d'aisance.
B2	Peut suivre l'essentiel d'une conférence, d'un discours, d'un rapport et d'autres genres d'exposés éducationnels/professionnels, qui sont complexes du point de vue du fond et de la forme.
B1	Peut suivre une conférence ou un exposé dans son propre domaine à condition que **le sujet soit familier** et la présentation directe, simple et **clairement structurée**.
	Peut suivre le plan général d'**exposés courts** sur **des sujets familiers** à condition que **la langue en soit standard** et **clairement articulée**.
A2	Pas de descripteur disponible.
A1	Pas de descripteur disponible.

Comprendre des annonces et instructions orales	
C2	Comme C1
C1	Peut extraire des détails précis d'une annonce publique émise **dans de mauvaises conditions et déformée par la sonorisation** (par exemple, des annonces publiques dans une gare, un stade). Peut comprendre des informations techniques complexes, telles que des modes d'emploi, des spécifications techniques pour un produit ou un service qui lui sont familiers.
B2	Peut comprendre des annonces et des messages courants sur **des sujets concrets et abstraits**, s'ils sont en **langue standard** et émis à **un débit normal**.
B1	Peut comprendre des informations techniques simples, tels que des modes d'emploi pour un équipement d'usage courant.

	Peut suivre des directives détaillées.
A2	Peut saisir le point essentiel **d'une annonce ou d'un message brefs**, simples et **clairs**. Peut comprendre des indications simples relatives à la façon d'aller d'un point à un autre, à pied ou avec les transports en commun.
A1	Peut comprendre des instructions qui lui sont adressées **lentement** et avec soin et suivre des **directives courtes** et simples.

Comprendre des émissions de radio et des enregistrements	
C2	Comme C1
C1	Peut comprendre une gamme étendue de matériel enregistré ou radiodiffusé, y compris en **langue non standard** et identifier des détails fins incluant l'implicite des attitudes et des relations des interlocuteurs.
B2	Peut comprendre les enregistrements en **langue standard** que l'on peut rencontrer dans la vie sociale, professionnelle ou universitaire et reconnaître le point de vue et l'attitude du locuteur ainsi que le contenu informatif.
	Peut comprendre la plupart des documentaires radiodiffusés en **langue standard** et peut identifier correctement l'humeur, le ton, etc., du locuteur.
B1	Peut comprendre l'information contenue dans la plupart des documents enregistrés ou radiodiffusés, dont **le sujet est d'intérêt personnel** et **la langue standard clairement articulée**.
	Peut comprendre les points principaux des bulletins d'information radiophoniques et de documents enregistrés simples, sur **un sujet familier**, si **le débit est assez lent** et **la langue relativement articulée**.
A2	Peut comprendre et extraire l'information essentielle de **courts passages** enregistrés ayant trait à **un sujet courant prévisible**, si **le débit est lent** et **la langue clairement articulée**.
A1	Pas de descripteur disponible.

En résumé des descriptions en gras ci-dessus, plus le niveau du CECR s'avère élevé, plus les apprenants peuvent comprendre le français à l'oral, en particulier les discours comportant les caractéristiques suivantes :

1. un débit de parole rapide ;
2. un discours mal articulé ;
3. un discours comportant peu longues pauses ;
4. la présence de bruit ;
5. l'utilisation d'un vocabulaire spécialisé ;
6. l'utilisation d'expressions riches ;
7. une certaine longueur en matière de contenu ;
8. un discours mal structuré ;
9. un sujet abstrait et non familier.

En outre, un niveau élevé du CECR paraît requis afin d'écouter du français en direct ou à la radio.

Il est également suggéré que, en plus des variables issues des recherches précédentes sur la facilité d'écoute (voir le chapitre précédent), ces éléments se révèlent essentiels à la distinction des niveaux du CECR. Cependant, tous ne sont pas indiqués de manière spécifique et sans ambiguïté, ce qui laisse la place à une détermination plus détaillée des éléments. De plus, il ne s'agit que d'hypothèses basées sur les descriptions, et il demeure possible que certains éléments, non inclus dans les descriptions, soient également liés à la distinction des niveaux. En se référant à des recherches antérieures sur la facilité d'écoute et à la description des compétences de la compréhension orale selon des niveaux du CECR, nous avons examiné les variables de lisibilité et de fluidité, qui correspondent à deux types de variables distincts, afin de déterminer spécifiquement les variables impliquées.

3.4.2. Variables de lisibilité

Concernant les variables de lisibilité, 61 variables sont utilisées. Elles sont obtenues avec FABRA, un outil en ligne permettant de calculer un large éventail de variables prédictives de lisibilité pour le français (Wilkens *et al.*, 2022)[41]. Comme indiqué au chapitre 2, l'étude de la facilité d'écoute a commencé par une application des recherches en lisibilité à l'analyse des transcriptions. En outre,

[41] https://cental.uclouvain.be/fabra.

l'existence signalée de variables valides, permettant de mesurer la facilité d'écoute, ne peut être ignorée. Cette thèse se focalise dès lors sur les variables de lisibilité qui ont été considérées 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 s'avèrent moins pertinentes dans le contexte de la facilité d'écoute et ne font pas partie de notre analyse. En outre, cette thèse se concentre sur les variables liées au mot de contenu, nécessaires afin de comprendre le sens du discours, plutôt que sur les mots fonctionnels. De plus, s'il est attendu que certaines variables soient apparentées, seule l'une d'entre elles sera prise en considération (par exemple, les mots de la ressource FLELex[42] pour les niveaux A1 à C2 du CECR, dont seule la proportion de mots de la ressource FLELex pour le niveau A1 est prise en considération).

Ainsi, les variables de lisibilité étudiées dans cette thèse, sélectionnées à l'aide de FABRA, sont les suivantes :

1. 2 variables basées sur la longueur ;
2. 17 variables lexicales ;
3. 41 variables syntaxiques ;
4. un score de la formule de lisibilité de Kandel et Moles (tableau 9).

Tableau 9 : Variables de lisibilité utilisées dans cette thèse[43][44]

Variables basées 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	Proportion de lemmes partagés entre toutes les phrases du texte.

[42] FLELex consiste en un lexique pour le FLE qui rapporte les fréquences normalisées des mots (lemmes) à chaque niveau du CECR (François *et al.*, 2014, http://cental.uclouvain.be/flelex).
[43] https://cental.uclouvain.be/fabra/docs.html.
[44] Wilkens *et al.* (2022)

Diversité lexicale	LEXdvrFLC	CTTR[45] 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 moyenne 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 moyenne 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 moyenne 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 (niveau du CECR dans lequel un mot est rencontré pour la première fois).
	LEXgrdFMLA1	Proportion de lemmes de niveau A1 selon la méthode d'apprentissage automatique de Pintard & 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)

45 $\frac{\text{Nombre de types}}{\sqrt{2 \times \text{Nombre de tokens}}}$

Voisins orthographiques	LEXnghPHO	PLD20 (Yarkoni *et al.*, 2008[46]) moyen pour les mots du texte, calculés à partir de Lexique3. Il s'agit de la distance de Levenshtein orthographique et phonologique moyenne de chaque mot par rapport aux 20 mots les plus proches trouvés dans Lexique3.
Normes lexicales	LEXnrmCNCR	Niveau moyen de concrétude des mots obtenus à la base de listes (Bonin *et al.*, 2018[47] ; Bonin *et al.*, 2011[48] ; Desrochers & Thompson, 2009[49] ; Bonin *et al.*, 2003[50] ; Desrochers & Bergeron, 2000[51]).
	LEXnrmFAM	Niveau moyen de familiarité obtenu sur la base de listes (Desrochers & Thompson, 2009 ; Ferrand *et al.*, 2008[52] ; Bonin *et al.*, 2003 ; Desrochers & Bergeron, 2000).

[46] Yarkoni, T., Balota, D. A., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15, pp.971-979.

[47] Bonin, P., Méot, A., & Bugaiska, A. (2018). Concreteness norms for 1,659 french words: Relationships with other psycholinguistic variables and word recognition times. *Behavior research methods*, 50 (6), pp.2366-2387.

[48] Bonin, P., Méot, A., Ferrand, L., & Roux, S. (2011). L'imageabilité : normes et relations avec d'autres variables psycholinguistiques. *L'année Psychologique*, 111 (2), pp.327-357.

[49] Desrochers, A., & Thompson, G. L. (2009). Subjective frequency and imageability ratings for 3,600 french nouns. *Behavior research methods*, 41 (2), pp.546-557.

[50] Bonin, P., Méot, A., Aubert, L.-F., Malardier, N., Niedenthal, P., & Capelle-Toczek, M.-C. (2003). Normes de concrétude, de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots. *L'année Psychologique*, 103 (4), pp.655-694.

[51] Desrochers, A., & Bergeron, M. (2000). Valeurs de fréquence subjective et d'imagerie pour un échantillon de 1,916 substantifs de la langue française. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 54 (4), p.274.

[52] Ferrand, L., Bonin, P., Méot, A., Augustinova, M., New, B., Pallier, C., & Brysbaert, M. (2008). Age-of-acquisition and subjective frequency estimates for all generally known monosyllabic french words and their relation

	LEXnrmIMG	Valeurs moyennes d'imageabilité obtenue à partir de listes (Bonin *et al.*, 2011 ; Desrochers & Thompson, 2009 ; Bonin *et al.*, 2003 ; Desrochers & Bergeron, 2000).
Sophistication lexicale	LEXsopFK1	Proportion de lemmes dans les premières bandes de fréquences de 1 000 mots de FLELex.
	LEXsopGK1	Proportion de mots dans les premières bandes de fréquence de 1 000 mots de la liste de vocabulaire de Gougenheim[53].
	LEXsopLLK1	Proportion de lemmes dans les premières bandes de fréquence de 1 000 mots de Lexique3.
Caractéristiques graduées	LEXgrdBLA1	Proportion d'expressions lexicales de niveau A1 dans le chapitre 5 de Beacco[54].
Variables syntaxiques		
Développement du langage	SYNdevNPHRS	Nombre moyen de constituants/phrases.
	SYNdevTUL	Longueur de l'unité T.
	SYNdevVG1	Proportion de verbes du 1er groupe français dans le texte.
Caractéristiques de la proposition	SYNclsLEN	Longueur moyenne des propositions en nombre de mots.
Caractéristiques des temps verbaux	SYNtnsfINDP	Proportion de verbes à l'indicatif présent par rapport au nombre de verbes.
	SYNtnsfINDI	Proportion de verbes à l'indicatif imparfait par rapport au nombre de verbes.
	SYNtnsfINDPS	Proportion de verbes à l'indicatif passé simple par rapport au nombre de verbes.

with other psycholinguistic variables. *Behavior research methods*, 40 (4), pp.1049-1054.

[53] Gougenheim, G., Michea, R., Rivenc, P., & Sauvageot, A. (1964). L'elaboration du français fondamental (1er degré). *Étude sur l'établissement d'un vocabulaire et d'une grammaire de bas*, Paris: Didier.

[54] Beacco, J.-C., & Porquier, R. (2007). *Niveau A1 pour le français : Un référentiel*. Paris: Didier.

	SYNtnsfINDPC	Proportion de verbes à l'indicatif passé composé par rapport au nombre de verbes.
	SYNtnsfINDPQP	Proportion de verbes à l'indicatif plus-que-parfait par rapport au nombre de verbes.
	SYNtnsfINDPA	Proportion de verbes à l'indicatif passé antérieur par rapport au nombre de verbes.
	SYNtnsfINDFS	Proportion de verbes à l'indicatif futur simple par rapport au nombre de verbes.
	SYNtnsfINDFA	Proportion de verbes à l'indicatif futur antérieur par rapport au nombre de verbes.
	SYNtnsfCNDP	Proportion de verbes au conditionnel présent par rapport au nombre de verbes.
	SYNtnsfCNDPS	Proportion de verbes au conditionnel passé par rapport au nombre de verbes.
	SYNtnsfIMPP	Proportion de verbes à l'impératif présent par rapport au nombre de verbes.
	SYNtnsfIMPPS	Proportion de verbes à l'impératif passé par rapport au nombre de verbes.
	SYNtnsfSUBP	Proportion de verbes au subjonctif présent par rapport au nombre de verbes.
	SYNtnsfSUBPS	Proportion de verbes au subjonctif passé par rapport au nombre de verbes.
	SYNtnsfSUBI	Proportion de verbes au subjonctif imparfait par rapport au nombre de verbes.
	SYNtnsfSUBPQP	Proportion de verbes au subjonctif plus-que-parfait par rapport au nombre de verbes.
	SYNtnsfINF	Proportion de verbes à l'infinitif par rapport au nombre de verbes.
	SYNtnsfINFC	Proportion de verbes à l'infinitif composé par rapport au nombre de verbes.
	SYNtnsfPP	Proportion de verbes au participe présent par rapport au nombre de verbes.

	SYNtnsfPPS	Proportion de verbes au participe passé par rapport au nombre de verbes.
	SYNtnsfGERP	Proportion de verbes au gérondif par rapport au nombre de verbes.
	SYNtnsfGERPS	Proportion de verbes au gérondif passé par rapport au nombre de verbes.
	SYNtnsfTAP	Proportion de verbes au temps passif par rapport au nombre de verbes.
POS Tag	SYNposADJ	Proportion de mots identifiés comme ADJ (adjectif), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.
	SYNposADP	Proportion de mots identifiés comme ADP (adposition), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.
	SYNposADV	Proportion de mots identifiés comme ADV (adverbe), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.
	SYNposAUX	Proportion de mots identifiés comme AUX (auxiliaire), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.
	SYNposCCONJ	Proportion de mots identifiés comme CCONJ (conjonction de coordination), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.
	SYNposINTJ	Proportion de mots identifiés comme INTJ (interjection), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.

	SYNposNOUN	Proportion de mots identifiés comme NOUN (nom), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.
	SYNposNUM	Proportion de mots identifiés comme NUM (numéral), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.
	SYNposPART	Proportion de mots identifiés comme PART (particule), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.
	SYNposPRON	Proportion de mots identifiés comme PRON (pronom), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.
	SYNposPROPN	Proportion de mots identifiés comme PROPN (nom propre), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.
	SYNposSCONJ	Proportion de mots identifiés comme SCONJ (conjonction de subordination), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.
	SYNposVERB	Proportion de mots identifiés comme VERB (verbe), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.
	SYNposX	Proportion de mots identifiés comme X (autre), suivant les étiquettes POS universelles et l'analyseur Stanza par rapport au nombre de mots du texte.

Variables de la formule de lisibilité		
Service de formule de lisibilité	REAfrmKM	Score de la formule de lisibilité pour le français (Kandel-Moles)[55].

Au niveau des agrégateurs proposés par FABRA, nous avons utilisé la valeur brute pour LEXdvrFLC et REAfrmKM, tandis que la moyenne (« avg », la valeur moyenne de toutes les observations dans le texte) a été utilisée pour les autres variables.

Comme décrit précédemment, un total de 61 variables de lisibilité pouvant être extraites de FABRA – variables extraites des transcriptions – sont analysées concernant leur corrélation avec les niveaux du CECR. Elles sont générées à partir de fichiers texte via un code de programmation et sont produites par l'intermédiaire de FABRA.

3.4.3. Variables de fluidité

Dans le chapitre précédent, nous avons mentionné un nombre limité de variables liées à la phonétique prises en compte dans les recherches sur la facilité d'écoute. Nous avons augmentons ici le nombre de ces variables en nous référant au concept de fluidité, qui s'avère souvent utilisé dans les recherches sur la production orale. La notion de 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 & Foster, 1999: 96). Étant donné que la compréhension orale, tout comme la production orale, consiste en une aptitude qui requiert une telle compétence, il paraît raisonnable d'appliquer ce concept. Les indicateurs de fluidité sont les suivants (Skehan, 2003: 8 ; Tavakoli & Skehan, 2005: 254-255) :

[55] Kandel, L. & Moles, A. (1958). Application de l'indice de Flesch à la langue française. *Cahiers Etudes de Radio-Télévision*, 19 (1958), pp.253-274.

- le débit :

– nombre de mots et de syllabes par minute (débit de parole, vitesse d'articulation, volume de la parole) ;

– nombre moyen de syllabes entre les pauses ;

- la pause :

– longueur et nombre de pauses remplies, de pauses silencieuses et d'intervalles silencieux ;

- la reformulation :

– nombre de reformulations ;

– nombre de répétitions.

Le chapitre précédent montre que, dans les études précédentes sur la facilité d'écoute, le débit de parole a été principalement examiné parmi les variables susmentionnées. Dans cette thèse, les variables liées à la reformulation n'ont pas été prises en compte, car il s'agit d'une perspective qui peut également être envisagée à partir de la transcription, et, surtout, les manuels d'accompagnement audio utilisés comme données contiennent rarement des reformulations ou des répétitions. Par conséquent, cette thèse examine les variables liées à la prosodie en enrichissant les variables liées au débit de parole et aux pauses par rapport aux études précédentes. Par ailleurs, un silence de plus de 200 ms est défini comme une pause, selon la description de la section 3.3.1. En d'autres termes, les pauses n'ont pas été distinguées entre les pauses remplies (par exemple : « euh ») et les pauses silencieuses. Ainsi, les dix variables basées sur la fluidité prises en compte dans cette étude sont les suivantes :

1. le nombre de syllabes par minute ;
2. le nombre de syllabes par minute sans les pauses ;
3. le nombre moyen de syllabes entre les pauses ;
4. la durée moyenne d'une syllabe ;
5. le nombre de mots par minute ;
6. le nombre de mots par minute sans les pauses ;
7. le nombre moyen de mots entre les pauses ;
8. la durée moyenne d'un mot ;
9. la durée moyenne des pauses ;
10. le nombre de pauses par minute.

Les variables 1, 3, 5 et 7 se rapportent au débit de parole, tandis que 2 et 6 se rapportent à la vitesse d'articulation. Parmi ces variables, seule 5 a été prise en compte dans certaines des études précédentes sur la facilité d'écoute, mais le nombre de syllabes s'avère également pris en compte dans cette thèse. Les variables 4 et 8, quant à elles, correspondent à des variables pour lesquelles les dénominateurs et les numérateurs des formules de calcul des variables 2 et 6 sont inversés, et elles représentent également le débit de parole. La longueur et le nombre de pauses sont ensuite considérés dans les variables 9 et 10. Les méthodes utilisées afin de calculer ces variables se réfèrent à celles spécifiées dans les études précédentes sur la fluidité.

Tableau 10 : Méthode de calcul des variables de fluidité[56]

Variables	Méthode de calcul	Références
(1) Nombre de syllabes par minute	Nombre total de syllabes/durée totale d'enregistrement [minutes]	Préfontaine (2010), Ginther *et al.* (2010)
(2) Nombre de syllabes par minute sans les pauses	Nombre total de syllabes/(durée totale d'enregistrement [minutes] – durée totale de pauses [minutes])	Préfontaine (2010), Peltonen (2017)
(3) Nombre moyen de syllabes entre les pauses	Nombre total de syllabes/(nombre total de pauses + 1)	Ginther *et al.* (2010), Préfontaine (2010), Tavakoli

[56] Les dix variables de fluidité sont désignées dans cette thèse respectivement par les noms suivants : *syllables_per_minute* pour (1), nombre de syllabes par minute, *syllables_per_minutes_without_pauses* pour (2), nombre de syllabes par minute sans les pauses, s*yllables_between_pauses* pour (3), nombre moyen de syllabes entre les pauses, *duration_syllable* pour (4), durée moyenne d'une syllabe, *words_per_minute* pour (5), nombre de mots par minute, *words_per_minutes_without_pauses* pour (6), nombre de mots par minute sans les pauses, *words_between_pauses* pour (7), nombre moyen de mots entre les pauses, *duration_word* pour (8), durée moyenne d'un mot, *length_pauses* pour (9), durée moyenne des pauses, *pauses_per_minute* pour (10), nombre de pauses par minute.

		(2016), Segalowitz *et al.* (2017)
(4) Durée moyenne d'une syllabe	(Durée totale d'enregistrement [minutes] – durée totale de pauses [minutes])/nombre total de syllabes	Segalowitz *et al.* (2017)
(5) Nombre de mots par minute	Nombre total de mots/durée totale d'enregistrement [minutes]	Recherches précédentes sur la facilité d'écoute (Harwood (1955b), Kotani *et al.* (2014), Yoon *et al.* (2016), Kotani & Yoshimi (2017, 2019), Yoshimi & Kotani (2020))
(6) Nombre de mots par minute sans les pauses	Nombre total de mots/(durée totale d'enregistrement [minutes] – durée totale de pauses [minutes])	Méthode de variable (2)
(7) Nombre moyen de mots entre les pauses	Nombre total de mots/(nombre total de pauses + 1)	Lennon (1990)
(8) Durée moyenne d'un mot	(Durée totale d'enregistrement [minutes] – durée totale de pauses [minutes])/nombre total de mots	Méthode de variable (6)
(9) Durée moyenne des pauses	Durée totale de pauses [minutes]/nombre total de pauses	Segalowitz *et al.* (2017), Ginther *et al.* (2010)

(10) Nombre de pauses par minute	Nombre total de pauses/durée totale d'enregistrement [minutes]	Cucchiarini *et al.* (2002)

Comme décrit ci-dessus, un total de dix variables de fluidité sont analysées pour leur corrélation avec les niveaux du CECR en tant que variables pouvant être extraites du son. Les variables de fluidité sont calculées à partir des fichiers TextGrid à l'aide d'un code de programmation.

3.5. Conclusion

Dans ce chapitre, la méthodologie de recherche de cette thèse a été décrite. Un modèle de prédiction de la facilité d'écoute de la langue française est examiné sur la base des données présentées à la section 3.2. afin d'obtenir des réponses aux questions de recherche présentées à la section 3.1. L'examen et l'analyse du modèle s'effectuent en deux étapes principales, comme indiqué à la section 3.3. La première consiste en l'analyse du modèle à partir des données avec des corrections manuelles, tandis que la seconde correspond à l'analyse du modèle sans correction manuelle mais grâce à une automatisation aussi poussée que possible. En comparant les résultats des deux étapes, les difficultés de l'automatisation du modèle de prédiction de la facilité d'écoute du français ont été identifiées.

Par ailleurs, des modèles prédictifs ont été considérés à chaque étape à partir de trois approches différentes : SVM, wav2vec et CamemBERT (voir figure 9). Concernant le modèle SVM particulièrement, les variables importantes étaient expliquées à la section 3.4. La comparaison des résultats des trois approches permet d'envisager la création d'un modèle de prédiction optimale. Sur la base de la méthodologie de recherche décrite dans ce chapitre, le premier modèle – une analyse du modèle avec des données corrigées manuellement – est examiné dans le chapitre suivant et les résultats sont présentés.

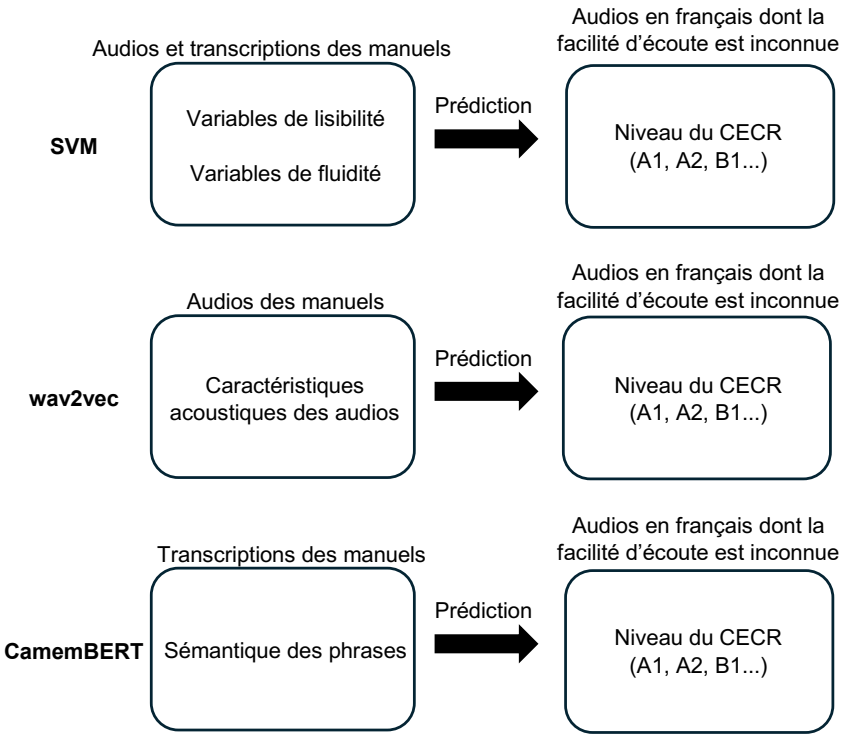

Figure 9 : Déroulement général de la création d'un modèle de la facilité d'écoute

4. Premiers modèles de la facilité d'écoute[57]

Ce chapitre présente les résultats des premiers modèles prédictifs de la facilité d'écoute utilisant des informations provenant des transcriptions et des signaux vocaux sur la base des données et des méthodes présentées dans le chapitre précédent. En ce qui concerne les premiers modèles, la performance des modèles est évaluée en utilisant des données qui ont été corrigées manuellement.

Les premiers modèles sont examinés dans le but de répondre aux cinq sous-questions de recherche suivantes :

1. Dans quelle mesure les informations extraites du signal vocal aident-elles à la prédiction de la facilité d'écoute ?
2. Dans quelle mesure les informations linguistiques provenant de la transcription aident-elles à la prédiction de la facilité d'écoute ?
3. La précision de la prédiction de la facilité d'écoute varie-t-elle en fonction du style de parole ?
4. Existe-t-il des différences dans la performance de prédiction à différents niveaux ?
5. Quel est le meilleur modèle permettant de prédire la facilité d'écoute ?

Tout d'abord, les sous-questions 1 et 2 concernent la question de recherche 1 (Quelles sont les variables qui sont associées à la facilité d'écoute pour le français langue étrangère ?) mentionnée à la section 3.1. En obtenant des réponses à ces sous-questions, il devient également possible d'examiner la question de recherche 2 (Les caractéristiques acoustiques contribuent-elles à l'amélioration des performances au niveau de la prédiction de la facilité d'écoute ?). La sous-question 3, quant à elle, correspond à la question de recherche 3 (Les performances au niveau de la prédiction de la facilité d'écoute varient-elles en fonction du style de parole ?). Enfin, les sous-questions 4 et 5 servent à confirmer la performance du modèle et à examiner l'automatisation des prédictions de facilité d'écoute. Cela correspond à la question de recherche 4 (Est-il possible d'automatiser la prédiction de la facilité d'écoute ?).

---

[57] Ce chapitre est basé sur Ozawa, M., Wilkens, R. S., Sugiyama, K., & François, T. (2024), avec des ajouts et des modifications.

Par ailleurs, trois approches, SVM, wav2vec et CamemBERT, sont utilisées afin de mettre au point ces premiers modèles. Elles sont ensuite comparées et discutées : les résultats du modèle SVM seront présentés à la section 4.1., ceux de wav2vec à la section 4.2. et CamemBERT à la section 4.3. Enfin, à la section 4.4., ces modèles seront comparés et les conclusions des premiers modèles, basés sur une approche partiellement manuelle et donc non automatisable, seront présentées.

4.1. Modèle SVM

Avant d'entraîner le modèle SVM, une analyse corrélationnelle des variables a été menée en vue de sélectionner les variables à inclure dans le modèle. Ces résultats seront d'abord présentés à la section 4.1.1. tandis que les résultats de la modélisation SVM basée sur cette analyse corrélationnelle le seront dans les sections suivantes.

4.1.1. Analyse corrélationnelle des variables

Avant tout, nous avons examiné les corrélations bivariées entre 71 variables de lisibilité et de fluidité et le niveau du CECR des audios de notre corpus. Cette étape vise à déterminer les variables indépendantes à incorporer dans le modèle SVM. Les niveaux du CECR constituant des variables ordinales, le coefficient de corrélation de Spearman semble le plus approprié afin de réaliser cette analyse corrélationnelle. Concernant chaque style de parole (corpus entier, dialogue et monologue), les variables pour lesquelles la valeur absolue du coefficient de corrélation avec le niveau du CECR s'avère supérieure à 0,2 sont listées ci-dessous. Comme des études plus détaillées paraissent nécessaires afin de déterminer les variables finales à inclure dans le SVM, le coefficient de corrélation de 0,2 (p < ,01) est fixé comme valeur seuil à ce stade de l'analyse, en laissant autant de variables que possible.

Tableau 11 : 39 variables pour le corpus entier (p < ,01)

Variables	Corrélation Spearman	Variables	Corrélation Spearman
LENsntWRD	0,63	LEXgrdFSOOUA1	-0,61

syllables_between_pauses	0,62	LEXgrdFFOA1	-0,58
words_between_pauses	0,6	LEXgrdFMLA1	-0,58
SYNdevTUL	0,59	REAfrmKM	-0,56
SYNdevNPHRS	0,59	LEXgrdBA1	-0,45
syllables_per_minute	0,57	duration_syllable	-0,42
SYNclsLEN	0,55	LEXsopFK1	-0,37
LEXdvrFLC	0,47	pauses_per_minute	-0,35
words_per_minute	0,46	length_pauses	-0,34
syllables_per_minutes_with out_pauses	0,42	SYNtnsfINDP	-0,3
SYNtnsfPPS	0,41	LEXsopLLK1	-0,26
SYNtnsfPP	0,37	duration_word	-0,25
SYNtnsfTAP	0,36	SYNposINTJ	-0,25
SYNposADP	0,36	LEXnrmIMG	-0,24
SYNposSCONJ	0,31		
SYNtnsfSUBP	0,3		
SYNposNOUN	0,29		
LENwrdSYL	0,27		
LEXnghPHO	0,27		
SYNtnsfINDI	0,26		
words_per_minutes_without _pauses	0,25		
SYNposADJ	0,25		
SYNtnsfINDPQP	0,25		
SYNtnsfINDFS	0,23		
SYNtnsfINF	0,22		

Le tableau 11 montre les variables pour lesquelles la valeur absolue du coefficient de corrélation avec le niveau du CECR se révèle supérieure à 0,2 pour le corpus entier. En ce qui concerne les caractéristiques, par exemple, il existe *LENsntWRD*, *SYNdevTUL* et *SYNdevNPHRS* qui constituent des variables liées à la longueur des phrases. Il existe également les variables liées au débit de parole telles que *syllables_between_pauses*, *words_between_pauses* et *syllables_per_minute*. En outre, les variables présentant une forte corrélation

négative avec le niveau du CECR comprennent *LEXgrdFSOOUA1*, *LEXgrdFFOA1* et *LEXgrdFMLA1*, c'est-à-dire des variables liées à la difficulté du lexique.

Les coefficients de corrélation entre le niveau du CECR et les variables indépendantes dans les données dialogiques sont, quant à eux, présentés dans le tableau 12.

Tableau 12 : 44 variables pour le dialogue (p < ,01)

Variables	Corrélation Spearman	Variables	Corrélation Spearman
words_between_pauses	0,67	LEXgrdFSOOUA1	-0,61
syllables_between_pauses	0,66	LEXgrdFFOA1	-0,58
LENsntWRD	0,65	REAfrmKM	-0,57
syllables_per_minute	0,64	LEXgrdFMLA1	-0,54
SYNclsLEN	0,59	duration_syllable	-0,49
SYNdevNPHRS	0,59	LEXgrdBA1	-0,45
SYNdevTUL	0,58	pauses_per_minute	-0,43
words_per_minute	0,57	SYNtnsfINDP	-0,38
syllables_per_minutes_without_pauses	0,5	length_pauses	-0,37
LEXdvrFLC	0,5	SYNposINTJ	-0,35
SYNposADP	0,4	LEXnrmIMG	-0,34
SYNposSCONJ	0,4	duration_word	-0,33
SYNtnsfTAP	0,4	LEXsopFK1	-0,31
SYNtnsfPP	0,39	LEXnrmCNCR	-0,24
SYNtnsfPPS	0,37		
SYNtnsfSUBP	0,36		
SYNtnsfINDI	0,35		
words_per_minutes_without_pauses	0,35		
SYNtnsfINDPQP	0,32		
SYNposNOUN	0,27		
SYNtnsfINDFS	0,26		
SYNposX	0,24		
SYNposADJ	0,23		

LENwrdSYL	0,23	
SYNtnsfCNDPS	0,23	
LEXnghPHO	0,22	
SYNtnsfINDPC	0,22	
SYNtnsfINF	0,21	
SYNtnsfINFC	0,21	
SYNtnsfINDPS	0,2	

De même que les résultats pour le corpus entier, les résultats concernant le dialogue montrent que les variables disposant de coefficients de corrélation élevés s'avèrent également typiques. Il s'agit, par exemple, de variables liées au débit de parole, comme *words_between_pauses*, *syllables_between_pauses* et *syllables_per_minute,* et à la longueur de phrase, comme *LENsntWRD*, *SYNclsLEN*, *SYNdevNPHRS* et *SYNdevTUL*. En outre, il existe également des variables liées au niveau de difficulté du lexique comme *LEXgrdFSOOUA1*, *LEXgrdFFOA1* et *LEXgrdFMLA1*.

Enfin, les résultats à propos des données monologue sont présentés dans le tableau 13.

Tableau 13 : 34 variables pour le monologue (p < ,01)

Variables	**Corrélation Spearman**	**Variables**	**Corrélation Spearman**
SYNdevTUL	0,57	LEXgrdFMLA1	-0,61
LENsntWRD	0,55	LEXgrdFSOOUA1	-0,61
SYNdevNPHRS	0,54	LEXgrdFFOA1	-0,55
syllables_between_pauses	0,54	REAfrmKM	-0,53
LEXdvrFLC	0,51	LEXgrdBA1	-0,42
words_between_pauses	0,48	LEXsopFK1	-0,41
SYNclsLEN	0,47	LEXsopLLK1	-0,35
SYNtnsfPPS	0,45	duration_syllable	-0,27
syllables_per_minute	0,38	length_pauses	-0,26
SYNtnsfPP	0,36	SYNposPRON	-0,22
SYNtnsfTAP	0,31	pauses_per_minute	-0,22
LEXnghPHO	0,3		

LENwrdSYL	0,28	
SYNposNOUN	0,27	
SYNposADV	0,26	
SYNposADP	0,26	
syllables_per_minutes_without_pauses	0,26	
words_per_minute	0,25	
SYNposSCONJ	0,24	
SYNtnsfSUBP	0,23	
SYNposADJ	0,22	
SYNtnsfINF	0,21	
SYNtnsfINDFS	0,21	

De même, en ce qui concerne le monologue, les variables présentant des coefficients de corrélation élevés s'avèrent typiques : *SYNdevTUL*, *LENsntWRD* et *SYNdevNPHRS* représentent la longueur de la phrase. De plus, *LEXgrdFMLA1*, *LEXgrdFSOOUA1* et *LEXgrdFFOA1* constituent des variables représentant le niveau de difficulté du lexique. Parmi les variables liées au débit de parole, *syllables_between_pauses* présente un coefficient de corrélation de 0,54. Cependant, à part cette exception, aucune d'entre elles ne présente de coefficient de corrélation supérieur à 0,5. Il s'agit d'une différence par rapport aux résultats obtenus pour le corpus entier et le dialogue. En d'autres termes, l'association entre les variables de fluidité et les niveaux du CECR se révèle moins élevée à propos des monologues, par rapport aux dialogues.

En comparant les résultats des trois styles de parole ci-dessus, le nombre de variables disposant d'une valeur absolue de coefficient de corrélation de plus de 0,2 avec le niveau du CECR s'élève à 44 sur la section du corpus constitué uniquement de dialogues, contre 34 dans la section constituée de monologues. Entre les deux, 39 variables ont été identifiées comme pertinentes concernant le corpus entier. Ainsi, la comparaison des dialogues et des monologues montre que le niveau du CECR offre une relation marquée avec plus de variables pour les dialogues que pour les monologues. Par ailleurs, dans les trois catégories de styles de parole, plusieurs variables fortement corrélées avec le niveau du CECR font parties de la même famille, par exemple celles qui indiquent la longueur des

phrases et le niveau de difficulté du lexique. Dès lors, il semble possible que des corrélations apparaissent entre des variables similaires, c'est-à-dire apportant plus ou moins la même information. Dans chaque style de parole, les coefficients de corrélation de Spearman sont donc à vérifier aussi entre les variables dans le but de minimiser la colinéarité lors de la construction du modèle. Par ailleurs, concernant les variables dont la valeur absolue du coefficient de corrélation avec le niveau du CECR s'avère supérieure à 0,2 ci-dessus, nous avons examiné les coefficients de corrélation entre elles.

Pour ce faire, nous avons d'abord procédé à un regroupement approximatif des variables énumérées dans les tableaux suivants, en fonction de ce que les variables mesurent et représentent. Les groupes constitués et les variables qui y appartiennent pour le corpus entier sont présentés dans le tableau 14.

Tableau 14 : Catégorisation de 39 variables pour le corpus entier

Catégorie	Nombre de variables	Variables
Temps verbal	9	SYNtnsfPPS, SYNtnsfPP, SYNtnsfTAP, SYNtnsfSUBP, SYNtnsfINDI, SYNtnsfINDPQP, SYNtnsfINDFS, SYNtnsfINF, SYNtnsfINDP
Débit de parole	8	syllables_between_pauses, words_between_pauses, syllables_per_minute, words_per_minute, syllables_per_minutes_without_pauses, words_per_minutes_without_pauses, duration_word, duration_syllable
Niveau de difficulté du lexique	6	LEXsopLLK1, LEXgrdFSOOUA1, LEXgrdFFOA1, LEXgrdFMLA1, LEXgrdBA1, LEXsopFK1
Syntaxe	5	SYNposADP, SYNposSCONJ, SYNposINTJ, SYNposNOUN, SYNposADJ
Longueur de phrase	5	LENsntWRD, SYNclsLEN, SYNdevNPHRS, SYNdevTUL, LENwrdSYL
Pause	2	pauses_per_minute, length_pauses
Type de vocabulaire	1	LEXnrmIMG
Phonologie	1	LEXnghPHO
Richesse lexicale	1	LEXdvrFLC
Formule de lisibilité	1	REAfrmKM

Ensuite, le tableau 15 montre les groupes et les variables concernant le dialogue.

Tableau 15 : Catégorisation de 44 variables pour le dialogue

Catégorie	Nombre de variables	Variables
Temps verbal	13	SYNtnsfTAP, SYNtnsfPP, SYNtnsfPPS, SYNtnsfSUBP, SYNtnsfINDI, SYNtnsfINDPQP, SYNtnsfINDFS, SYNtnsfCNDPS, SYNtnsfINDPC, SYNtnsfINF, SYNtnsfINFC, SYNtnsfINDPS, SYNtnsfINDP
Débit de parole	8	words_between_pauses, syllables_between_pauses, syllables_per_minute, words_per_minute, syllables_per_minutes_without_pauses, words_per_minutes_without_pauses, duration_word, duration_syllable
Syntaxe	6	SYNposADP, SYNposSCONJ, SYNposINTJ, SYNposNOUN, SYNposX, SYNposADJ
Niveau de difficulté du lexique	5	LEXgrdFSOOUA1, LEXgrdFFOA1, LEXgrdFMLA1, LEXgrdBA1, LEXsopFK1
Longueur de phrase	5	LENsntWRD, SYNclsLEN, SYNdevNPHRS, SYNdevTUL, LENwrdSYL
Type de vocabulaire	2	LEXnrmIMG, LEXnrmCNCR
Pause	2	pauses_per_minute, length_pauses
Phonologie	1	LEXnghPHO
Richesse lexicale	1	LEXdvrFLC
Formule de lisibilité	1	REAfrmKM

Enfin, le tableau 16 montre les groupes et les variables concernant le monologue.

Tableau 16 : Catégorisation de 34 variables pour le monologue

Catégorie	Nombre de variables	Variables
Temps verbal	6	SYNtnsfPPS, SYNtnsfPP, SYNtnsfTAP, SYNtnsfSUBP, SYNtnsfINF, SYNtnsfINDFS
Syntaxe	6	SYNposNOUN, SYNposADV, SYNposADP, SYNposSCONJ, SYNposADJ, SYNposPRON
Niveau de difficulté du lexique	6	LEXgrdFMLA1, LEXgrdFSOOUA1, LEXgrdFFOA1, LEXgrdBA1, LEXsopFK1, LEXsopLLK1
Débit de parole	6	syllables_between_pauses, words_between_pauses, syllables_per_minute, syllables_per_minutes_without_pauses, words_per_minute, duration_syllable
Longueur de phrase	5	SYNdevTUL, LENsntWRD, SYNdevNPHRS, SYNclsLEN, LENwrdSYL
Pause	2	length_pauses, pauses_per_minute
Phonologie	1	LEXnghPHO
Richesse lexicale	1	LEXdvrFLC
Formule de lisibilité	1	REAfrmKM

La catégorisation des variables révèle que, particulièrement en ce qui concerne le dialogue, il existe treize variables liées aux temps verbaux pour lesquelles la valeur absolue du coefficient de corrélation s'avère supérieure à 0,2, comparé à six variables pour le monologue. Cela suggère que les temps verbaux s'avèrent plus importants pour les dialogues que pour les monologues comme indice du niveau du CECR.

Enfin, les corrélations entre les variables ont été examinées à l'aide du coefficient de corrélation de Spearman, sur les 39 variables pour le corpus entier,

les 44 variables pour le dialogue et les 34 variables pour le monologue. En ce qui concerne les variables pour lesquelles la valeur absolue du coefficient de corrélation avec d'autres variables se révèle supérieure à 0,7, lorsque certaines de ces variables sont incluses dans une même catégorie (voir tableaux 14, 15 et 16), seule celle qui offre la corrélation la plus forte avec le niveau du CECR est conservée pour chaque catégorie, tandis que les autres variables sont exclues.

Concernant le dialogue, par exemple, le coefficient de corrélation entre *LENsntWRD* et *SYNclsLEN* pour la longueur de phrase s'élève à 0,96 ($p < ,01$) et entre *SYNdevNPHRS* et *SYNdevTUL*, à 0,94 ($p < ,01$). Ainsi, parmi ces trois variables, seule *LENsntWRD*, qui présente le coefficient de corrélation le plus élevé avec le niveau du CECR, est retenue tandis que les autres sont exclues. Concernant le monologue, le coefficient de corrélation entre *LEXgrdFSOOUA1* et *LEXgrdFMLA1*, représentant le niveau de difficulté du lexique, s'élève à 0,89 ($p < ,01$), et entre *LEXgrdFMLA1* et *LEXgrdFFOA1,* à 0,83 ($p < ,01$), indiquant une très forte corrélation. Dans ce cas, seule *LEXgrdFMLA1*, qui présente le coefficient de corrélation le plus élevé avec le niveau du CECR, est retenue tandis que les autres sont exclues. Cependant, il n'est pas évident de justifier que, par exemple, la proportion de présent de l'indicatif et la proportion de passé composé de l'indicatif mesurent des choses tout à fait similaires ; de même pour la proportion d'adjectifs et celle de prépositions. Pour cette raison, concernant les variables liées au temps verbal et à la syntaxe, toutes les variables ont été conservées pour inclusion dans le modèle si la valeur absolue du coefficient de corrélation se trouve supérieure à 0,2.

À l'issue de ce processus, les variables indépendantes finalement incorporées dans le SVM pour chaque style de parole sont répertoriées dans les tableaux suivants. Des diagrammes en boîte à moustaches sont également présentés concernant les variables les plus fortement corrélées avec le niveau du CECR concernant les dialogues et les monologues. Tout d'abord, le tableau 17 introduit les variables finales du corpus entier.

Tableau 17 : 24 variables finales pour le corpus entier (p < ,01)[58]

Variables	Catégorie	Corrélation Spearman
LENsntWRD	Longueur de phrase	0,63
syllables_between_pauses	Débit de parole	0,62
LEXdvrFLC	Richesse lexicale	0,47
SYNtnsfPPS	Temps verbal	0,41
SYNtnsfPP	Temps verbal	0,37
SYNtnsfTAP	Temps verbal	0,36
SYNposADP	Syntaxe	0,36
SYNposSCONJ	Syntaxe	0,31
SYNtnsfSUBP	Temps verbal	0,3
SYNposNOUN	Syntaxe	0,29
LEXnghPHO	Phonologie	0,27
SYNtnsfINDI	Temps verbal	0,26
SYNposADJ	Syntaxe	0,25
SYNtnsfINDPQP	Temps verbal	0,25
SYNtnsfINDFS	Temps verbal	0,23
SYNtnsfINF	Temps verbal	0,22
LEXnrmIMG	Type de vocabulaire	-0,24
SYNposINTJ	Syntaxe	-0,25
SYNtnsfINDP	Temps verbal	-0,3
length_pauses	Pause	-0,34
pauses_per_minute	Pause	-0,35
LEXgrdBA1	Difficulté du lexique	-0,45
REAfrmKM	Formule de lisibilité	-0,56
LEXgrdFSOOUA1	Difficulté du lexique	-0,61

Ensuite, le tableau 18 montre les variables finales du dialogue tandis que la figure 10 présente le diagramme en boîte à moustaches pour *words_between_pauses* (nombre moyen de mots entre les pauses ; r = 0,67).

[58] Dans les tableaux 17, 18 et 19, les variables ombrées constituent des variables de fluidité et les autres des variables de lisibilité.

Tableau 18 : 29 variables finales pour le dialogue (p < ,01)

Variables	Catégorie	Corrélation Spearman
words_between_pauses	Débit de parole	0,67
LENsntWRD	Longueur de phrase	0,65
LEXdvrFLC	Richesse lexicale	0,5
SYNposADP	Syntaxe	0,4
SYNposSCONJ	Syntaxe	0,4
SYNtnsfTAP	Temps verbal	0,4
SYNtnsfPP	Temps verbal	0,39
SYNtnsfPPS	Temps verbal	0,37
SYNtnsfSUBP	Temps verbal	0,36
SYNtnsfINDI	Temps verbal	0,35
SYNtnsfINDPQP	Temps verbal	0,32
SYNposNOUN	Syntaxe	0,27
SYNtnsfINDFS	Temps verbal	0,26
SYNposX	Syntaxe	0,24
SYNposADJ	Syntaxe	0,23
SYNtnsfCNDPS	Temps verbal	0,23
LEXnghPHO	Phonologie	0,22
SYNtnsfINDPC	Temps verbal	0,22
SYNtnsfINF	Temps verbal	0,21
SYNtnsfINFC	Temps verbal	0,21
SYNtnsfINDPS	Temps verbal	0,2
LEXnrmIMG	Type de vocabulaire	-0,34
SYNposINTJ	Syntaxe	-0,35
length_pauses	Pause	-0,37
SYNtnsfINDP	Temps verbal	-0,38
pauses_per_minute	Pause	-0,43
LEXgrdBA1	Difficulté du lexique	-0,45
REAfrmKM	Formule de lisibilité	-0,57
LEXgrdFSOOUA1	Difficulté du lexique	-0,61

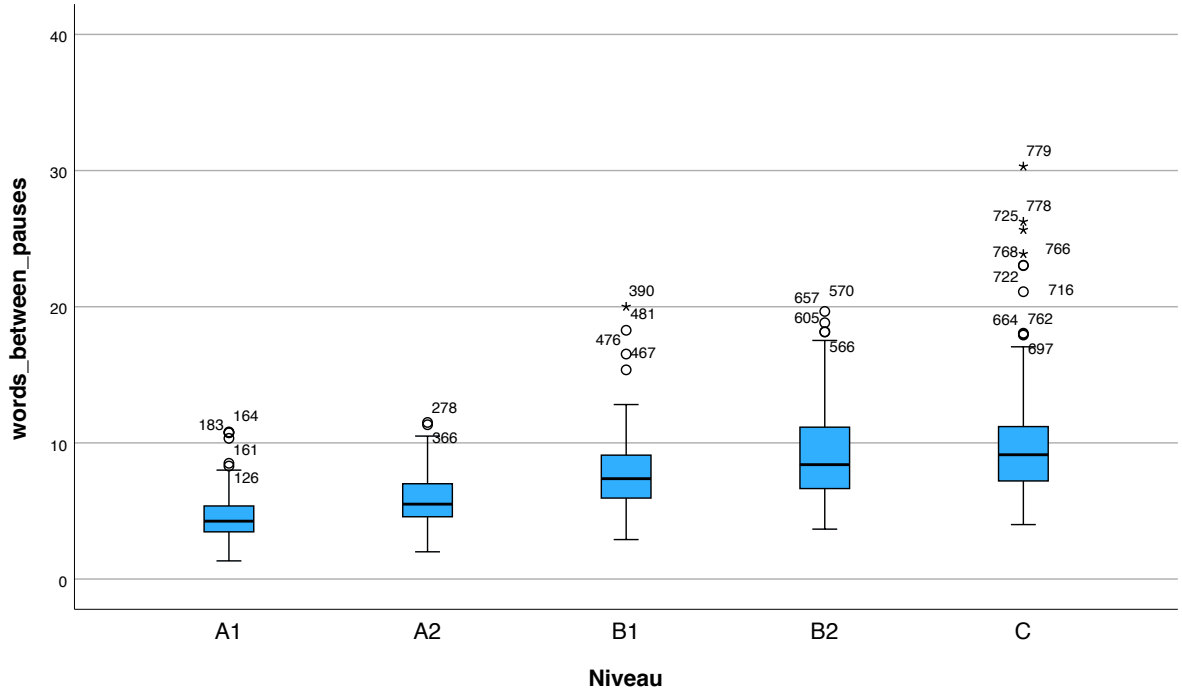

Figure 10 : Boîte à moustaches de *words_between_pauses*

Enfin, le tableau 19 montre les variables finales du monologue tandis que la figure 11 présente le diagramme en boîte à moustaches pour *LEXgrdFMLA1* (proportion de lemmes de niveau A1 ; r = -0,61).

Tableau 19 : 21 variables finales pour le monologue (p < ,01)

Variables	Catégorie	Corrélation Spearman
SYNdevTUL	Longueur de phrase	0,57
syllables_between_pauses	Débit de parole	0,54
LEXdvrFLC	Richesse lexicale	0,51
SYNtnsfPPS	Temps verbal	0,45
SYNtnsfPP	Temps verbal	0,36
SYNtnsfTAP	Temps verbal	0,31
LEXnghPHO	Phonologie	0,3
SYNposNOUN	Syntaxe	0,27
SYNposADV	Syntaxe	0,26
SYNposADP	Syntaxe	0,26
SYNposSCONJ	Syntaxe	0,24
SYNtnsfSUBP	Temps verbal	0,23
SYNposADJ	Syntaxe	0,22
SYNtnsfINF	Temps verbal	0,21
SYNtnsfINDFS	Temps verbal	0,21
pauses_per_minute	Pause	-0,22
SYNposPRON	Syntaxe	-0,22
length_pauses	Pause	-0,26
LEXgrdBA1	Difficulté du lexique	-0,42
REAfrmKM	Formule de lisibilité	-0,53
LEXgrdFMLA1	Difficulté du lexique	-0,61

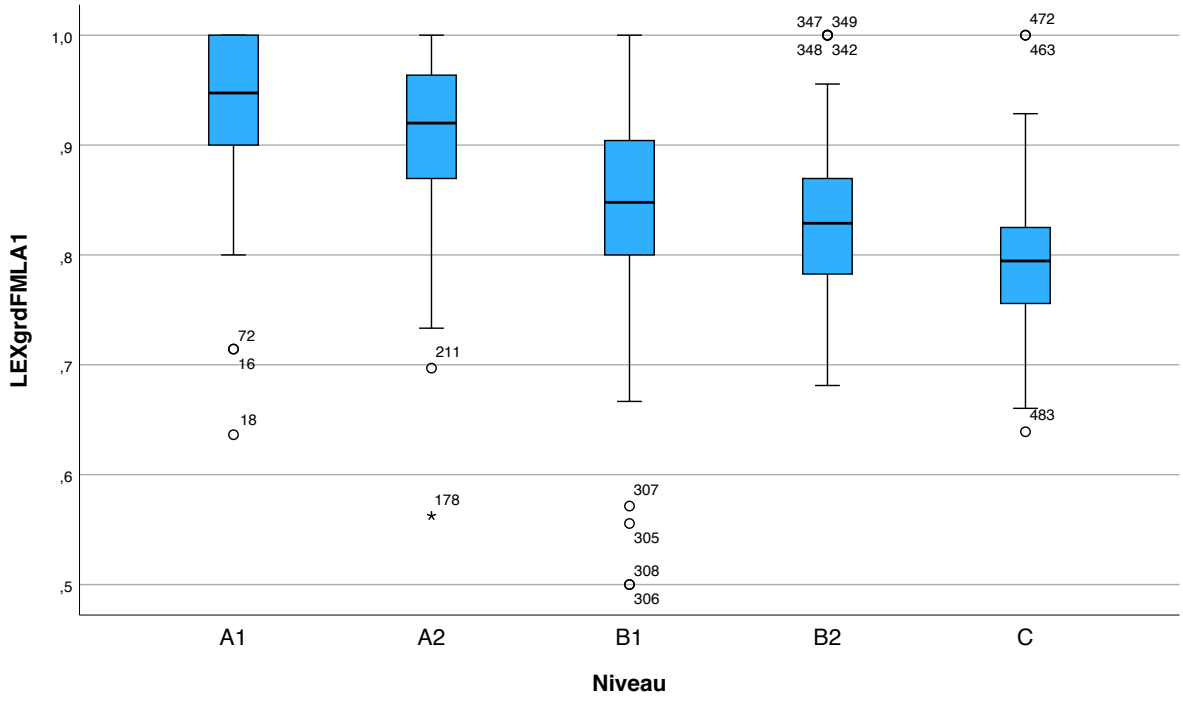

Figure 11 : Boîte à moustaches de *LEXgrdFMLA1*

Parmi les variables de fluidité, le coefficient de corrélation le plus élevé avec le niveau du CECR pour le dialogue s'élève à 0,67 pour *words_between_pauses* qui représente le débit de parole. De même, pour le monologue, la meilleure variable atteint 0,54, à savoir les *syllables_between_pauses* pour le débit de parole. Les variables relatives aux pauses ont également été incorporées dans les variables finales, avec un coefficient de corrélation de -0,43 pour *pauses_per_minute* et de -0,37 pour *length_pauses* pour le dialogue. Concernant le monologue, en revanche, le coefficient de corrélation pour *pauses_per_minute* s'élève à -0,22 et *length_pauses* à -0,26. En raison de ces différences entre le dialogue et le monologue, les variables de fluidité peuvent être plus importantes concernant le dialogue que pour le monologue.

Les résultats obtenus jusqu'à présent peuvent être résumés en accordant une attention particulière aux résultats des dialogues et des monologues. Trois variables communes paraissent fortement corrélées avec le niveau du CECR :

1. les variables concernant la longueur de phrase (exemple : *LENsntWRD* (r = 0,65) pour les dialogues, *SYNdevTUL* (r = 0,57) pour les monologues) ;

2. les variables concernant le débit de parole (exemple : *words_between_pauses* (r = 0,67) pour les dialogues, *syllables_between_pauses* (r = 0,54) pour les monologues) ;
3. les variables concernant le niveau de difficulté du lexique (exemple : *LEXgrdFSOOUA1* (r = -0,61) pour les dialogues, *LEXgrdFMLA*1 (r = -0,61) pour les monologues).

Indépendamment du style de parole, ces trois variables sont considérées comme liées à la facilité d'écoute du français. En ce qui concerne la différence entre les dialogues et les monologues, le niveau du CECR possède une relation avec plus de variables pour les dialogues que pour les monologues. En particulier, les variables liées aux temps verbaux se trouvent davantage corrélées avec le niveau du CECR concernant les dialogues. Cela signifie que les temps verbaux peuvent s'avérer particulièrement important lors de la compréhension de dialogues. En particulier, ceux qui révèlent une corrélation commune entre les dialogues et les monologue sont le participe passé, le participe présent, le temps passif, le subjonctif présent, l'infinitif et l'indicatif futur simple. En revanche, certaines corrélations n'ont été trouvées que pour les dialogues : l'indicatif imparfait, l'indicatif plus-que-parfait, le conditionnel passé, l'indicatif passé composé, l'infinitif composé, l'indicatif passé simple et l'indicatif présent. Par ailleurs, la corrélation n'apparaît négative que pour l'indicatif présent, de sorte que plus la proportion de l'indicatif présent est élevée, plus le niveau de discours du CECR s'avère bas. Concernant les dialogues, le locuteur peut parler d'événements à n'importe quel moment, comme le passé ou le futur, dans un seul contexte, le moment auquel le locuteur parle constituant l'axe principal. En revanche, concernant les monologues, de nombreuses occasions permettent de se concentrer sur un seul sujet. Bien sûr, de nombreux temps verbaux peuvent également être utilisés dans un monologue, mais en considérant le contexte comme une unité, un plus grand développement de l'histoire peut avoir lieu dans un dialogue. Par conséquent, les temps verbaux et le niveau du CECR sont considérés comme étant plus étroitement liés pour les dialogues.

Sur la base des résultats ci-dessus concernant les temps verbaux, des exemples représentatifs de transcription de dialogues et de monologues sont présentés respectivement. Nous avons évité de montrer des exemples du niveau A1, pour lequel les transcriptions tendent à être relativement courtes, ainsi que ceux du

niveau C, trop longues, et nous avons privilégié des exemples du niveau B1 afin de fournir des exemples faciles à comprendre. Le premier consiste en un dialogue. Les temps verbaux de la transcription, considérés comme des variables liées aux temps verbaux dans cette étude, sont soulignés.

> – SOS Médecins Paris, service d'urgence, bonsoir !
> – Bonsoir madame.
> Je ne vais pas très bien.
> Il est tard et ça m'inquiète.
> J'aimerais qu'un médecin vienne chez moi au plus vite.
> – D'accord.
> Est-ce que vous avez pris votre température ?
> – Oui, j'ai 40 de fièvre.
> – Vous avez pris un médicament ?
> – Un peu de paracétamol.
> Mais je n'en ai plus, et je me sens pas bien du tout.
> – Entendu.
> Je vais prendre vos coordonnées pour vous envoyer quelqu'un...
>
> (Dialogue, B1, *Cosmopolite*, Unité 2, p.6)

La transcription suivante propose un exemple de monologue.

> Votre demande concerne l'habitation.
> Vous souhaitez faire un devis pour assurer votre habitation : tapez 1, modifier votre contrat : tapez 2, résilier votre contrat : tapez 3, déclarer un sinistre : tapez 4, suivre un sinistre déjà déclaré : tapez 5.
> Pour toute autre demande : tapez 6.
>
> (Monologue, B1, *Tendances*, Unité 8, pp.168-169)

L'exemple de dialogue ci-dessus constitue l'enregistrement d'un appel téléphonique passé au service d'urgence. L'exemple de monologue, quant à lui, correspond à un enregistrement provenant d'un service de réponse vocale automatisée. Bien qu'il s'agisse dans les deux cas de sons liés au téléphone, nous

pouvons conclure que les dialogues s'avèrent plus susceptibles d'utiliser une variété de temps verbaux dans leur contenu.

En outre, les variables indiquant le débit de parole, indépendamment du style de parole, se sont avérées fortement corrélée avec le niveau du CECR. Toutefois, il a également été constaté que les variables de fluidité peuvent être plus importantes pour le dialogue que pour le monologue. En effet, concernant les monologues, le locuteur parle seul et tend donc à pouvoir maintenir un certain débit de parole et des pauses. Pour le dialogue, en revanche, outre le locuteur, le débit de parole et les pauses d'un ou de plusieurs interlocuteurs se révèlent également importants. Ainsi, il est possible que la présence d'interlocuteurs impacte les variables de fluidité concernant les dialogues.

4.1.2. Résultats du modèle SVM

En utilisant les variables des tableaux 17, 18 et 19 de la section précédente, des modèles SVM ont été construits de manière à prédire le niveau du CECR des documents audios. Afin d'examiner l'impact des informations provenant des transcriptions et des signaux vocaux – c'est-à-dire capturées via les variables de lisibilité et de fluidité – sur la construction d'un modèle de facilité d'écoute, trois modèles SVM ont été entraînés pour chaque style de parole. Le premier consiste en un modèle utilisant uniquement les variables de lisibilité, le deuxième, uniquement les variables de fluidité, et le troisième utilisant à la fois les variables de lisibilité et de fluidité.

Une fois le jeu de prédicteurs défini, chaque modèle SVM nécessite d'être doublement entraîné. En effet, il convient non seulement d'entraîner les paramètres du modèle sur le jeu d'entraînement, mais également de sélectionner les meilleurs hyperparamètres sur le jeu de développement. Les hyperparamètres constituent des paramètres pour les algorithmes d'apprentissage automatique. Dans les SVM, les hyperparamètres comportent *C* (paramètre de régularisation), *kernel* (type de kernel) et *gamma* (contrôle de la largeur de la fonction kernel). Tout d'abord, *C* ajuste le degré de tolérance des erreurs de classification et empêche le surentraînement, qui augmente la précision pour les données d'entraînement mais diminue la précision pour les données inconnues. Lorsque *C* est petit, les erreurs de classification s'avèrent plus fréquentes, mais la précision pour les données inconnues augmente. En revanche, si *C* est grand, les erreurs de

classification sont réduites, tout comme la précision pour les données inconnues. Par ailleurs, dans *kernel*, il est possible de définir si les données sont séparées linéairement ou non linéairement. À propos de *kernel*, *gamma* constitue un hyperparamètre lié à la limite de décision séparant les données. Un petit gamma sépare sur des frontières de décision simples, tandis qu'un grand gamma le réalise sur des frontières de décision complexes. Plus le gamma s'avère grand, plus le risque de surapprentissage devient important.

En ce qui concerne la recherche des hyperparamètres, nous les avons explorés à l'aide d'une grille de recherche. Cette dernière renvoie à une méthode permettant de fixer les hyperparamètres à leurs valeurs optimales. À l'aide de celle-ci, nous avons exploré les hyperparamètres, C = [0,001 ; 0,05 ; 0,1] et C = [1 ; 10 ; 100 ; 1000], qui ont été utilisés comme coefficient de régularisation. Puis, nous avons comparé des noyaux linéaires et RBF (non linéaire) et les valeurs de gamma (γ) = [1e-3 ; 1e-4] comme hyperparamètres spécifiques au noyau RBF. Après avoir essayé toutes les combinaisons de ces hyperparamètres, la meilleure a été sélectionnée. Les performances des modèles, quant à elles, ont été évaluées au moyen d'une procédure de validation croisée à dix plis et à l'aide des mesures d'évaluation d'exactitude, de précision, de rappel, et de macro F1 et de F1 pondérée.

L'exactitude correspond à la proportion de prédictions correctes par rapport à l'ensemble des prédictions (c'est-à-dire, dans le tableau 20, (VP + VN)/(VP + VN + FP + FN)). La précision, quant à elle, correspond à la proportion réellement positive les prédictions positives (VP/(VP + FP)) tandis que le rappel consiste en la proportion de prédictions effectivement positives parmi toutes celles qui auraient dû être prédites positives (VP/(VP + FN)).

Afin d'expliquer la macro F1 et la F1 pondérée, il faut d'abord expliciter la F1. La F1 correspond à la moyenne harmonique de la précision et du rappel ((2 × précision × rappel)/(précision + rappel)). La macro F1 représente la F1 moyenne calculée par classe (à savoir par niveau du CECR) tandis que la F1 pondérée consiste en la somme de « nombre de données d'une classe/données totales × Macro F1 d'une classe ». En d'autres termes, la macro F1 traite toutes les classes avec le même poids, quelles que soient les différences dans la distribution des données par classe. En revanche, la F1 pondérée pondère les classes en fonction du nombre de données. Cela signifie que les classes disposant de plus de données engendrent un impact plus important. Par ailleurs, les données

utilisées de cette étude pour cinq niveaux du CECR dans chaque style de parole ne s'avèrent pas présentes dans des proportions exactement équilibrées. Le modèle doit être correctement évalué alors qu'il n'est pas possible de former le même nombre de positifs et de négatifs. Ainsi, la macro F1 est considérée comme un premier indicateur d'évaluation.

Tableau 20 : Indicateurs d'évaluation des modèles SVM

	Positif	Négatif
Positif prédit	VP (vrai positif)	FP (faux positif)
Négatif prédit	FN (faux négatif)	VN (vrai négatif)

Lors de la validation croisée à dix plis, 80 % des données servent à la phase d'entraînement, 10 % pour le développement et 10 % pour le test. Les moyennes obtenues sur les dix plis de tests pour chaque score sont rapportées. En outre, la macro F1 pour chacun des cinq niveaux du CECR est également indiquée.

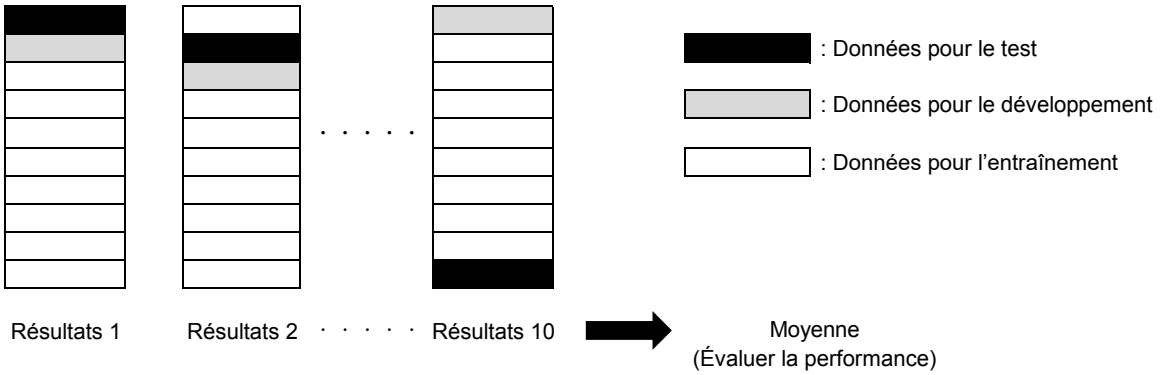

Figure 12 : Déroulement de la validation croisée

Afin de sélectionner un C susceptible d'améliorer les performances du modèle, les résultats utilisant C = [0,001 ; 0,05 ; 0,1] sont présentés dans le tableau 21.

Tableau 21 : Résultat pour la tâche de prédiction du niveau de facilité d'écoute (C = [0,001 ; 0,05 ; 0,1])

	Lisibilité + Fluidité		
	Corpus entier	Dialogue	Monologue
	avg (std)	avg (std)	avg (std)
Exactitude	0,48 (0,04)	0,52 (0,06)	0,45 (0,06)
Précision	0,46 (0,05)	0,53 (0,05)	0,47 (0,09)
Rappel	0,47 (0,04)	0,51 (0,06)	0,44 (0,05)
Macro F1	0,45 (0,04)	0,5 (0,05)	0,42 (0,05)
F1 pondérée	0,45 (0,04)	0,51 (0,05)	0,43 (0,04)
F1-A1	0,64 (0,06)	0,67 (0,09)	0,53 (0,11)
F1-A2	0,47 (0,06)	0,49 (0,08)	0,44 (0,07)
F1-B1	0,22 (0,11)	0,37 (0,11)	0,22 (0,13)
F1-B2	0,29 (0,07)	0,36 (0,07)	0,33 (0,14)
F1-C	0,61 (0,07)	0,63 (0,1)	0,6 (0,13)

	Lisibilité		
	Corpus entier	Dialogue	Monologue
	avg (std)	avg (std)	avg (std)
Exactitude	0,47 (0,04)	0,47 (0,06)	0,43 (0,06)
Précision	0,41 (0,08)	0,46 (0,07)	0,38 (0,11)
Rappel	0,45 (0,04)	0,45 (0,06)	0,38 (0,05)
Macro F1	0,4 (0,03)	0,43 (0,06)	0,33 (0,05)
F1 pondérée	0,42 (0,05)	0,44 (0,05)	0,36 (0,06)
F1-A1	0,6 (0,07)	0,61 (0,08)	0,08 (0,12)
F1-A2	0,47 (0,04)	0,46 (0,08)	0,47 (0,07)
F1-B1	0,01 (0,03)	0,16 (0,08)	0,06 (0,11)
F1-B2	0,33 (0,08)	0,28 (0,16)	0,41 (0,13)
F1-C	0,61 (0,06)	0,62 (0,09)	0,62 (0,13)

	Fluidité		
	Corpus entier	Dialogue	Monologue
	avg (std)	avg (std)	avg (std)
Exactitude	0,38 (0,05)	0,38 (0,05)	0,4 (0,05)
Précision	0,36 (0,08)	0,38 (0,09)	0,33 (0,11)
Rappel	0,37 (0,04)	0,34 (0,05)	0,38 (0,04)
Macro F1	0,32 (0,04)	0,32 (0,05)	0,31 (0,04)
F1 pondérée	0,34 (0,05)	0,35 (0,06)	0,32 (0,05)
F1-A1	0,6 (0,06)	0,63 (0,09)	0,54 (0,09)
F1-A2	0,39 (0,07)	0,27 (0,11)	0,4 (0,09)
F1-B1	0,13 (0,08)	0,29 (0,1)	0,08 (0,1)
F1-B2	0,07 (0,06)	0,29 (0,07)	0 (0)
F1-C	0,42 (0,05)	0,14 (0,11)	0,54 (0,07)

Ensuite, les résultats utilisant C = [1 ; 10 ; 100 ; 1000] sont présentés dans le tableau 22.

Tableau 22 : Résultat du modèle SVM pour la tâche de prédiction du niveau de facilité d'écoute (C = [1 ; 10 ; 100 ; 1000])

	Lisibilité + Fluidité		
	Corpus entier	Dialogue	Monologue
	avg (std)	avg (std)	avg (std)
Exactitude	0,52 (0,05)	0,58 (0,06)	0,49 (0,09)
Précision	0,52 (0,05)	0,59 (0,07)	0,48 (0,1)
Rappel	0,51 (0,05)	0,56 (0,06)	0,49 (0,09)
Macro F1	0,51 (0,05)	0,56 (0,06)	0,47 (0,09)
F1 pondérée	0,52 (0,04)	0,58 (0,05)	0,47 (0,03)
F1-A1	0,71 (0,07)	0,75 (0,06)	0,65 (0,17)
F1-A2	0,52 (0,06)	0,58 (0,08)	0,49 (0,11)
F1-B1	0,34 (0,08)	0,38 (0,1)	0,32 (0,16)
F1-B2	0,4 (0,11)	0,5 (0,12)	0,3 (0,19)
F1-C	0,59 (0,06)	0,61 (0,1)	0,6 (0,14)

	Lisibilité		
	Corpus entier	Dialogue	Monologue
	avg (std)	avg (std)	avg (std)
Exactitude	0,49 (0,02)	0,52 (0,05)	0,49 (0,08)
Précision	0,48 (0,02)	0,53 (0,06)	0,48 (0,08)
Rappel	0,48 (0,02)	0,51 (0,05)	0,48 (0,08)
Macro F1	0,47 (0,02)	0,5 (0,06)	0,47 (0,08)
F1 pondérée	0,48 (0,04)	0,52 (0,05)	0,47 (0,03)
F1-A1	0,69 (0,05)	0,7 (0,08)	0,56 (0,16)
F1-A2	0,5 (0,04)	0,51 (0,1)	0,51 (0,11)
F1-B1	0,24 (0,09)	0,25 (0,1)	0,27 (0,15)
F1-B2	0,36 (0,11)	0,41 (0,15)	0,4 (0,19)
F1-C	0,58 (0,04)	0,65 (0,09)	0,59 (0,12)

	Fluidité		
	Corpus entier	Dialogue	Monologue
	avg (std)	avg (std)	avg (std)
Exactitude	0,39 (0,05)	0,42 (0,06)	0,39 (0,03)
Précision	0,36 (0,05)	0,42 (0,07)	0,31 (0,07)
Rappel	0,38 (0,04)	0,4 (0,06)	0,37 (0,03)
Macro F1	0,33 (0,03)	0,4 (0,06)	0,3 (0,03)
F1 pondérée	0,35 (0,05)	0,42 (0,05)	0,31 (0,05)
F1-A1	0,58 (0,06)	0,65 (0,08)	0,49 (0,15)
F1-A2	0,4 (0,07)	0,36 (0,13)	0,39 (0,08)
F1-B1	0,12 (0,08)	0,29 (0,09)	0,1 (0,1)
F1-B2	0,12 (0,06)	0,35 (0,1)	0 (0)
F1-C	0,45 (0,07)	0,35 (0,15)	0,55 (0,07)

Tout d'abord, nous avons comparé le tableau 21 des résultats lorsque C = [0,001 ; 0,05 ; 0,1] et le tableau 22 des résultats lorsque C = [1 ; 10 ; 100 ; 1000]. Or, la précision pour le corpus entier utilisant uniquement les variables de fluidité demeure la même pour les deux résultats. De plus, le résultat pour C = [0,001 ; 0,05 ; 0,1] s'avère légèrement plus élevé pour le monologue, qui utilise également uniquement les variables de fluidité. De même, les résultats pour l'exactitude, la précision, le rappel, la macro F1 et la F1 pondérée se révèlent plus élevés lorsque C = [1 ; 10 ; 100 ; 1000]. Cela signifie que les modèles utilisant C = [1 ; 10 ; 100 ; 1000] sont principalement plus performants. Par conséquent, à partir de maintenant, nous nous sommes concentrés sur les résultats du tableau 22 avec C = [1 ; 10 ; 100 ; 1000], et nous avons analysé et évalué les premiers modèles, tandis que les modèles SVM ont été évalués en matière de styles de parole et de niveau du CECR.

4.1.3. Analyse des modèles SVM

L'analyse des modèles SVM se fait en se basant sur le tableau 22. Concernant l'exactitude, la précision, le rappel, la macro F1 et la F1 pondérée, ces valeurs s'avèrent plus élevées pour les dialogues. Cela signifie que le niveau du CECR des documents oraux peut être prédit de manière légèrement plus précise pour les

dialogues que sur le corpus entier ou sur les monologues. Cela peut impliquer les résultats de l'analyse de corrélation à la section 4.1.1. En particulier, lorsque les résultats des dialogues et des monologues sont comparés, nous constatons que les variables fortement corrélées s'avèrent plus nombreuses pour les dialogues que pour les monologues (tableaux 18 et 19). L'inclusion de ces variables dans le modèle SVM peut avoir conduit à une augmentation de la performance du modèle concernant les dialogues.

Ensuite, nous avons examiné les trois modèles SVM créés – lisibilité + fluidité, lisibilité uniquement, fluidité uniquement – pour chaque style de parole. Sur le corpus entier, les scores les plus élevés pour nos mesures d'évaluation sont obtenus pour *le modèle lisibilité + fluidité*, suivi par *le modèle lisibilité* et enfin *le modèle fluidité*. La même tendance peut être observée pour les dialogues.

Concernant ces deux corpus (entier et dialogue), dans *le modèle lisibilité*, une certaine précision de prédiction a été obtenue avec la macro F1 de 0,47 pour le corpus entier et 0,5 pour le dialogue. *Le modèle fluidité*, avec la macro F1 de 0,33 pour le corpus entier et 0,4 pour le dialogue, ne s'avère pas aussi précis que *le modèle lisibilité*. Cependant, en considérant *le modèle lisibilité + fluidité*, qui combine les deux, la précision augmente avec la macro F1 de 0,51 pour le corpus entier et de 0,56 pour le dialogue. Ainsi, les variables de fluidité contribuent également à l'amélioration des performances en matière de prédiction.

En revanche, concernant le monologue, les mêmes valeurs ont été généralement observées pour *le modèle lisibilité + fluidité* et *le modèle lisibilité* en ce qui concerne presque tous les indicateurs. Les valeurs les plus faibles, pour chaque indicateur, ont été trouvées dans *le modèle fluidité*, comme dans les deux autres styles de parole. En examinant spécifiquement les valeurs, la macro F1 s'élève à 0,47 dans *le modèle lisibilité* et à 0,3 dans *le modèle fluidité*. Cependant, la macro F1 dans *le modèle lisibilité + fluidité* s'élève à 0,47, ce qui paraît identique à la macro F1 dans *le modèle lisibilité*. Ainsi, les variables de fluidité se révèlent moins importantes concernant le monologue tandis que les variables de lisibilité seules s'avèrent suffisantes afin de prédire le niveau du CECR.

Enfin, le modèle est analysé en examinant la macro F1 par niveau du CECR. Indépendamment du style de parole et des trois modèles différents, la précision prédictive du niveau A1 semble globalement élevée ; la précision prédictive des niveaux A2 et C, modérée à légèrement élevée, tandis qu'elle a tendance à baisser aux niveaux B1 et B2. Cela suggère qu'il existe des différences dans la

performance du modèle à différents niveaux du CECR. En d'autres termes, Nous constatons que les prédictions de niveau B ont tendance à être particulièrement difficiles à réaliser.

4.2. Modèle wav2vec

Dans cette section, nous décrivons le modèle wav2vec. Comme indiqué à la section 3.3.1., wav2vec est utilisé pour l'analyse spécialisée dans le domaine de la prosodie. En d'autres termes, il ne prend en compte aucune des informations provenant des transcriptions, mais se base uniquement sur les signaux vocaux. Les résultats de ce modèle, qui prédit le niveau du CECR des documents audios à partir des signaux vocaux, seront présentés à la section 4.2.1. et son analyse à la section 4.2.2.

4.2.1. Résultats du modèle wav2vec

Le modèle wav2vec extrait automatiquement des caractéristiques acoustiques des audios. Il est entraîné et évalué afin de prédire le niveau du CECR à partir de ces caractéristiques. Le tableau 23 décrit les résultats du modèle wav2vec. Comme pour les modèles SVM, une validation croisée à dix plis est effectuée, avec 80 % des données utilisées pour l'entraînement, 10 % pour le développement et 10 % pour le test. Trois modèles ont été entraînés, en fonction du style de parole.

Tableau 23 : Résultat de premier modèle wav2vec pour la tâche de prédiction du niveau de facilité d'écoute

	Corpus entier	Dialogue	Monologue
	avg (std)	avg (std)	avg (std)
Exactitude	0,31 (0,04)	0,29 (0,04)	0,26 (0,06)
Précision	0,28 (0,04)	0,24 (0,04)	0,27 (0,06)
Rappel	0,24 (0,09)	0,2 (0,11)	0,16 (0,08)
Macro F1	0,21 (0,05)	0,18 (0,07)	0,17 (0,05)
F1 pondérée	0,4 (0,04)	0,38 (0,04)	0,35 (0,08)
F1-A1	0,48 (0,16)	0,43 (0,08)	0,41 (0,18)
F1-A2	0,3 (0,09)	0,22 (0,12)	0,31 (0,1)
F1-B1	0,13 (0,12)	0,1 (0,11)	0 (0)
F1-B2	0,09 (0,1)	0,1 (0,16)	0,04 (0,09)
F1-C	0,02 (0,05)	0,03 (0,06)	0,08 (0,12)

4.2.2. Analyse du modèle wav2vec

Dans l'ensemble, les macro F1 se situent autour de 0,2, ce qui s'avère assez faible. Même en comparant les trois styles de parole, il n'apparaît pas de différence notable dans les valeurs. Cela signifie qu'il semble assez difficile de prédire le niveau du CECR à partir des seuls signaux vocaux.

En considérant la macro F1 pour chaque niveau, des valeurs modérées au niveau A1 (de 0,41 à 0,48) sont obtenus ; les résultats au niveau A2 (de 0,22 à 0,31) s'avèrent également supérieurs à la macro F1 pour chaque modèle. En revanche, pour les autres niveaux supérieurs à B1, les macros F1 se révèlent très faibles. Par conséquent, si les performances de prédiction semblent quelque peu garanties au niveau A, il s'est avéré plus difficile de prédire les niveaux supérieurs à B1 en utilisant les signaux vocaux. En d'autres termes, il se peut qu'aux niveaux B et supérieurs, il n'existe pas de caractéristiques du signal vocal qui soient suffisantes afin de discriminer les niveaux.

4.3. Modèle CamemBERT

Dans cette section, le modèle CamemBERT est examiné. En utilisant les transcriptions des audios comme données d'entrée, il est basé sur les caractéristiques des phrases françaises. Cela signifie que, alors que le modèle wav2vec s'appuie uniquement sur des signaux vocaux, le modèle CamemBERT prend uniquement en compte des informations provenant des transcriptions. La section 4.3.1. présentera les résultats du modèle CamemBERT, qui prédit le niveau du CECR des documents audios à partir des transcriptions. L'analyse de ces résultats sera ensuite présentée à la section 4.3.2.

4.3.1. Résultats du modèle CamemBERT

Les résultats du modèle CamemBERT sont repris ci-dessous. Comme pour les modèles SVM et wav2vec, une validation croisée à dix plis a été effectuée, 80 % des données servant à la phase d'entraînement, 10 % pour le développement et 10 % pour le test. De nouveau, trois modèles ont été créés, en fonction du style de parole.

Tableau 24 : Résultat du premier modèle CamemBERT pour la tâche de prédiction du niveau de facilité d'écoute[59]

	Corpus entier	Dialogue	Monologue
	avg (std)	avg (std)	avg (std)
Exactitude	0,63 (0,05)	0,6 (0,05)	0,58 (0,05)
Précision	0,64 (0,06)	0,57 (0,11)	0,62 (0,06)
Rappel	0,61 (0,05)	0,54 (0,06)	0,58 (0,05)
Macro F1	0,61 (0,05)	0,52 (0,08)	0,57 (0,05)
F1 pondérée	0,62 (0,05)	0,57 (0,06)	0,57 (0,05)
F1-A1	0,8 (0,05)	0,83 (0,04)	0,64 (0,14)
F1-A2	0,63 (0,11)	0,6 (0,08)	0,57 (0,05)
F1-B1	0,49 (0,12)	0,41 (0,2)	0,4 (0,09)
F1-B2	0,52 (0,09)	0,43 (0,17)	0,54 (0,14)
F1-C	0,6 (0,11)	0,34 (0,19)	0,68 (0,11)

4.3.2. Analyse du modèle CamemBERT

Dans cette section, comme dans les sections précédentes, les modèles sont examinés en utilisant la macro F1 comme indicateur d'évaluation. La macro F1 s'avère supérieure à 0,5 pour tous les modèles, ce qui indique une certaine précision de prédiction, même à partir des seules informations provenant des transcriptions. En comparant les styles de parole, elle semble légèrement supérieure pour les monologues (0,57) par rapport aux dialogues (0,52). Cela signifie que notre modèle basé sur CamemBERT paraît plus performant lorsqu'il s'agit de prédire le niveau du CECR des monologues à partir des transcriptions.

Ensuite, les macro F1 par niveau obtiennent également des valeurs moyennes dans l'ensemble. Les valeurs au niveau B1 se situent dans la plage de 0,4 pour tous les styles de parole, ce qui indique une tendance à des performances de prédiction comparativement faibles. En comparant le dialogue et le monologue du point de vue des niveaux du CECR, une différence notable réside dans les valeurs de la macro F1 au niveau C. En effet, la macro F1 pour le dialogue s'élève à 0,34, ce qui en fait la valeur la plus basse parmi les macros F1 par niveau du CECR pour

[59] Les résultats présentés ici diffèrent légèrement de ceux d'Ozawa *et al.* (2024), car nous avons depuis réussi à résoudre le problème de variance extrême du modèle CamemBERT, ce qui a entraîné quelques différences de performance.

le dialogue. En revanche, pour le monologue, elle atteint 0,68 au niveau C, représentant la valeur la plus élevée des macros F1 par niveau du CECR pour le monologue. Cependant, en ce qui concerne les performances de prédiction au niveau A1, nous observons que la macro F1 de dialogue (0,83) s'avère supérieure à celle de monologue (0,64). Enfin, nous constatons que le modèle spécialisé sur les dialogues excelle particulièrement dans la prédiction au niveau A1, tandis que le modèle dédié aux monologues semble capable d'effectuer des prédictions de manière équilibrée pour tous les niveaux sauf le niveau B1, avec des performances particulièrement élevées pour le niveau C comparées à celles de dialogue.

4.4. Comparaison des modèles

Dans cette section, les résultats des modèles décrits jusqu'à présent par SVM, wav2vec et CamemBERT sont comparés, et le meilleur modèle permettant de prédire la facilité d'écoute du français est examiné. Tout d'abord, nous avons examiné l'axe des styles de parole. Dans les tableaux 25, 26 et 27 ci-après, les tableaux 22, 23 et 24 sont récapitulés pour chaque style de parole. Dans ces tableaux, seules l'exactitude et la macro F1 et celle par niveau sont listées pour chaque modèle. Le tableau 25 présente les résultats de tous les modèles pour le corpus entier[60].

Tableau 25 : Résultat (SVM, wav2vec, CamemBERT) pour la tâche de prédiction du niveau de facilité d'écoute sur l'ensemble des données

	SVM-LF	SVM-L	SVM-F	wav2vec	CamemBERT
Exactitude (std)	0,52 (0,05)	0,49 (0,02)	0,39 (0,05)	0,31 (0,04)	0,63 (0,05)
Macro F1 (std)	0,51 (0,05)	0,47 (0,02)	0,33 (0,03)	0,21 (0,05)	0,61 (0,05)
F1-A1 (std)	0,71 (0,07)	0,69 (0,05)	0,58 (0,06)	0,48 (0,16)	0,8 (0,05)
F1-A2 (std)	0,52 (0,06)	0,5 (0,04)	0,4 (0,07)	0,3 (0,09)	0,63 (0,11)
F1-B1 (std)	0,34 (0,08)	0,24 (0,09)	0,12 (0,08)	0,13 (0,12)	0,49 (0,12)
F1-B2 (std)	0,4 (0,11)	0,36 (0,11)	0,12 (0,06)	0,09 (0,1)	0,52 (0,09)
F1-C (std)	0,59 (0,06)	0,58 (0,04)	0,45 (0,07)	0,02 (0,05)	0,6 (0,11)

Ensuite, le tableau 26 présente les résultats de tous les modèles pour le dialogue.

[60] Dans les tableaux 25, 26 et 27, SVM-LF représente le modèle SVM basé sur les variables de lisibilité et de fluidité, SVM-L représente le modèle utilisant uniquement les variables de lisibilité et SVM-F représente le modèle utilisant uniquement les variables de fluidité.

Tableau 26 : Résultat (SVM, wav2vec, CamemBERT) pour la tâche de prédiction du niveau de facilité d'écoute sur les données de style dialogue

	SVM-LF	SVM-L	SVM-F	wav2vec	CamemBERT
Exactitude (std)	0,58 (0,06)	0,52 (0,05)	0,42 (0,06)	0,29 (0,04)	0,6 (0,05)
Macro F1 (std)	0,56 (0,06)	0,5 (0,06)	0,4 (0,06)	0,18 (0,07)	0,52 (0,08)
F1-A1 (std)	0,75 (0,06)	0,7 (0,08)	0,65 (0,08)	0,43 (0,08)	0,83 (0,04)
F1-A2 (std)	0,58 (0,08)	0,51 (0,1)	0,36 (0,13)	0,22 (0,12)	0,6 (0,08)
F1-B1 (std)	0,38 (0,1)	0,25 (0,1)	0,29 (0,09)	0,1 (0,11)	0,41 (0,2)
F1-B2 (std)	0,5 (0,12)	0,41 (0,15)	0,35 (0,1)	0,1 (0,16)	0,43 (0,17)
F1-C (std)	0,61 (0,1)	0,65 (0,09)	0,35 (0,15)	0,03 (0,06)	0,34 (0,19)

Enfin, le tableau 27 présente les résultats pour le monologue.

Tableau 27 : Résultat (SVM, wav2vec, CamemBERT) pour la tâche de prédiction du niveau de facilité d'écoute sur les données de style monologue

	SVM-LF	SVM-L	SVM-F	wav2vec	CamemBERT
Exactitude (std)	0,49 (0,09)	0,49 (0,08)	0,39 (0,03)	0,26 (0,06)	0,58 (0,05)
Macro F1 (std)	0,47 (0,09)	0,47 (0,08)	0,3 (0,03)	0,17 (0,05)	0,57 (0,05)
F1-A1 (std)	0,65 (0,17)	0,56 (0,16)	0,49 (0,15)	0,41 (0,18)	0,64 (0,14)
F1-A2 (std)	0,49 (0,11)	0,51 (0,11)	0,39 (0,08)	0,31 (0,1)	0,57 (0,05)
F1-B1 (std)	0,32 (0,16)	0,27 (0,15)	0,1 (0,1)	0 (0)	0,4 (0,09)
F1-B2 (std)	0,3 (0,19)	0,4 (0,19)	0 (0)	0,04 (0,09)	0,54 (0,14)
F1-C (std)	0,6 (0,14)	0,59 (0,12)	0,55 (0,07)	0,08 (0,12)	0,68 (0,11)

En considérant uniquement les macro F1 pour chaque modèle, l'ordre de grandeur pour le corpus entier est CamemBERT > SVM-LF > SVM-L > SVM-F > wav2vec ; pour le dialogue, SVM-LF > CamemBERT > SVM-L > SVM-F > wav2vec ; pour le monologue, CamemBERT > SVM-LF = SVM-L > SVM-F > wav2vec. En d'autres termes, dans le modèle CamemBERT, le modèle SVM créé avec des variables de lisibilité et de fluidité et celui créé sur la base des variables de lisibilité uniquement, en extrayant des caractéristiques des informations provenant des transcriptions, la performance de prédiction s'avère relativement élevée. En revanche, la performance de prédiction paraît faible pour le modèle SVM créé sur la base des variables de fluidité uniquement, ainsi que pour le modèle wav2vec créé en extrayant des caractéristiques des signaux vocaux. Ainsi, cela signifie qu'il semble difficile de prédire le niveau du CECR de discours français en utilisant uniquement des informations provenant des audios. Inversement, les résultats montrent également que le niveau du CECR de discours français peut être prédit, dans une certaine mesure, à partir d'informations provenant uniquement des transcriptions.

Dès lors, il convient de se demander si l'information issue de l'audio s'avère inutile pour prédire le niveau du CECR de discours français, et s'il suffit de consulter la transcription afin de prédire la facilité d'écoute du français. Comme mentionné à la section 4.1.2., il convient de faire une comparaison entre les trois modèles SVM à ce sujet. Tout d'abord, concernant le dialogue, les variables de fluidité, c'est-à-dire l'information provenant du son, contribuent également à une meilleure performance de prédiction, alors que pour le monologue, les variables de fluidité s'avèrent moins importantes. Cela est confirmé par les résultats de CamemBERT puisque sa macro F1 concernant le monologue s'élève à 0,57, et est ainsi plus élevée que celle concernant le dialogue. Par conséquent, cela suggère que les informations provenant des transcriptions sont nettement plus essentielles que les informations liées à la fluence afin de prédire le niveau du CECR concernant les monologues. En outre, en prenant en considération l'ordre de grandeur de la macro F1 mentionné précédemment, la macro F1 semble légèrement plus élevée pour SVM-LF que pour CamemBERT uniquement pour le dialogue, ce qui suggère que les variables de fluidité contribuent à une meilleure performance de prédiction concernant les dialogues.

Ensuite, nous avons comparé les modèles par SVM, wav2vec et CamemBERT autour de l'axe des niveaux. Nous avons constaté que la macro F1 de B1 et de B2 s'avère fondamentalement faible. En effet, la faible valeur de la macro F1 pour la plupart de B1 et B2 peut être liée aux caractéristiques de ces niveaux, qui, en tant que niveaux intermédiaires du CECR, peuvent être un mélange de données proches du niveau A et du niveau C. Dès lors, la prédiction de niveau peut s'avérer plus complexe en raison de ces caractéristiques.

4.5. Conclusion

Dans ce chapitre, les modèles de prédiction de la facilité d'écoute de la langue française ont été examinés à partir de trois approches : SVM, wav2vec et CamemBERT. L'analyse des premiers modèles se termine par les réponses à nos cinq sous-questions de recherche. Concernant la première, « *Dans quelle mesure les informations extraites du signal vocal aident-elles à la prédiction de la facilité d'écoute ? »*, il s'est avéré qu'elles sont peu utiles à la prédiction. Cependant, en comparant les résultats du modèle SVM basé sur les variables de lisibilité et de fluidité et ceux du modèle SVM basé sur les variables de fluidité uniquement, il a été montré que les variables de fluidité ont également contribué à améliorer la performance de prédiction concernant le corpus entier et le dialogue. Ainsi, les informations provenant du son ne paraissent pas totalement dénuées d'utilité concernant la prédiction.

Concernant la deuxième question*, « Dans quelle mesure les informations linguistiques provenant de la transcription aident-elles à la prédiction de la facilité d'écoute ? »*, en examinant les résultats du modèle SVM basé sur les variables de lisibilité et de fluidité, du modèle SVM basé uniquement sur la variable de lisibilité et du modèle CamemBERT, les informations provenant des transcriptions peuvent s'avérer très utiles pour la prédiction. Elles se sont également révélées très importantes, en particulier dans le cas de monologues.

En ce qui concerne la troisième question, *« La précision de la prédiction de la facilité d'écoute varie-t-elle en fonction du style de parole ? »*, les résultats montrent qu'effectivement la précision des modèles diffère. En général, la précision de prédiction semble plus élevée pour le dialogue que pour le monologue. Cependant, la précision de prédiction de CamemBERT pour le monologue est plus élevée pour le dialogue.

D'ailleurs, pour la quatrième question*, « Existe-t-il des différences dans la performance de prédiction à différents niveaux ? »*, il apparaît qu'aux niveaux B1 et B2, la précision de prédiction se révèle souvent faible. Ainsi, il existe des différences entre les niveaux. En outre, la faible performance de prédiction pour les niveaux B1 et B2 pourrait être influencée par le fait que ces niveaux se situent à un niveau intermédiaire entre A et C, qui sont adjacents.

Enfin, concernant la cinquième question*, « Quel est le meilleur modèle permettant de prédire la facilité d'écoute ? »*, dans le modèle CamemBERT qui

contient des informations sur la transcription et le modèle SVM basé sur les variables de lisibilité et de fluidité, le modèle SVM basé uniquement sur les variables de lisibilité, la performance de prédiction s'est avérée élevée. Par conséquent, les performances des modèles créés à l'aide de la transcription semblent satisfaisantes.

5. Seconds modèles de la facilité d'écoute

Dans le chapitre précédent, le modèle de la facilité d'écoute a été examiné au travers d'un pré-traitement des données réalisé manuellement. Cette façon de faire, si elle paraît idéale dans un contexte scientifique permettant d'explorer des questions de recherche, n'est malheureusement pas applicable dans le contexte d'un outil de facilité d'écoute en ligne, devant rapidement traiter une requête. Ainsi, dans le présent chapitre, des modèles de la facilité d'écoute reposant sur une automatisation sont examinés. La méthodologie générale a déjà été décrite à la section 3.3.2., mais une description détaillée des données et de la méthodologie pour le second modèle sera donnée à la section 5.1. Comme pour les premiers modèles, trois approches différentes ont également été utilisées : SVM, wav2vec et CamemBERT. Les résultats et l'analyse du modèle SVM seront présentés à la section 5.2., ceux du modèle wav2vec à la section 5.3. et ceux du modèle CamemBERT à la section 5.4. Les résultats de ces trois modèles seront comparés à la section 5.5., et les conclusions concernant le second modèle seront menées à la section 5.6. Afin de pouvoir comparer avec précision les premiers et les seconds modèles dans le chapitre suivant, l'analyse a été menée dans la même perspective que les premiers modèles. Dès lors, les questions de recherche de ce chapitre restent les mêmes que celles des chapitres précédents.

1. Dans quelle mesure les informations extraites du signal vocal aident-elles à la prédiction de la facilité d'écoute ?
2. Dans quelle mesure les informations linguistiques provenant de la transcription aident-elles à la prédiction de la facilité d'écoute ?
3. La précision de la prédiction de la facilité d'écoute varie-t-elle en fonction du style de parole ?
4. Existe-t-il des différences dans la performance de prédiction à différents niveaux ?
5. Quel est le meilleur modèle permettant de prédire la facilité d'écoute ?

Les réponses à ces questions dans le second modèle seront présentées à la fin de ce chapitre, puis comparées à celles du premier modèle dans le chapitre suivant.

5.1. Données et méthodologie

Les données utilisées dans ce chapitre sont principalement les mêmes que celles décrites à la section 3.2. Toutefois, certaines données ont été exclues à cause de problème technique. Nous décrivons les données pour le second modèle dans cette section. Afin de préparer les données, la transcription automatique a d'abord été effectuée à l'aide de Speech-to-Text de Google, système de reconnaissance vocale, sur la base des données audio décrites à la section 3.2. La ponctuation s'avérant nécessaire lors de l'analyse des transcriptions générées, la transcription a été réalisée en utilisant la fonction de ponctuation automatique proposée par Speech-to-Text. Il a été confirmé que la transcription automatique s'effectuait bien pour toutes les données audios.

Ensuite, un alignement automatique a été effectué à l'aide de SPPAS sur la base des transcriptions générées automatiquement et des données audios. Aucune correction manuelle de l'alignement ni aucune correction manuelle de l'insertion de transcriptions n'ont été effectuées, comme cela avait été le cas lors de la création du premier modèle. À ce stade, il est apparu clairement que l'alignement automatique n'avait pas été réalisé avec succès pour 19 données. La répartition de ces données est présentée dans le tableau 28.

Tableau 28 : Données qui n'ont pas pu être alignées automatiquement[61]

	Nom des données	Style de parole	Nombre de mots	Durée d'enregistrement (min)
1	A1_Cosmopolite_22	Dialogue	50	0,57
2	B1_Cosmopolite_29	Dialogue	331	2,49
3	B2_Cosmopolite_19	Dialogue	447	2,31
4	B2_Le_DELF_B2_31	Dialogue	387	2,28
5	C1_abcDALF_13	Dialogue	1 652	9,1
6	C1_abcDALF_22	Monologue	1 189	6,87
7	C1_C2_Tendances_1	Dialogue	933	4,4
8	C1_C2_Tendances_7	Dialogue	1 143	6,46

[61] Le nombre de mots pour chaque donnée des tableaux 28 et 29 sont calculés à partir de transcriptions générées automatiquement.

9	C1_C2_Tendances_8	Dialogue	1 234	6,3
10	C1_Edito_19	Dialogue	1 078	5,58
11	C1_Edito_26	Dialogue	758	4,16
12	C1_Edito_34	Dialogue	1 470	8,78
13	C1_Le_DALF_25	Dialogue	354	2
14	C1_Le_DALF_64	Dialogue	1 216	6,58
15	C1_Reussir_le_Dalf_12	Dialogue	1 079	5,7
16	C2_abcDALF_3	Dialogue	2 484	13,64
17	C2_Le_DALF_56	Dialogue	868	5,1
18	C2_Le_DALF_63	Dialogue	2 612	14,1
19	C2_Reussir_le_Dalf_24	Dialogue	1 800	9,89

Le tableau 28 montre que plus de la moitié des données qui n'ont pas pu être alignées automatiquement représentent des données de niveau C et que toutes les données sauf une correspondent à des données de dialogue. En comparant le tableau 28 avec les tableaux 7 et 8, qui présentent la description des données, nous déduisons que l'alignement automatique n'a pas fonctionné correctement, principalement concernant les données pour lesquelles le nombre de mots et la durée d'enregistrement s'avèrent supérieurs à la moyenne du niveau. Cependant, comme nous ne considérons pas ces questions techniques dans cette étude, nous avons simplement exclu ces données dans l'analyse des seconds modèles.

Ensuite, les variables de lisibilité et de fluidité présentées aux sections 3.4.2. et 3.4.3. ont été extraites à l'aide d'un code de programmation basé sur les fichiers texte de transcription générés automatiquement et les fichiers TextGrid alignés automatiquement. Cependant, 17 fichiers texte ont été détectés du fait que l'extraction des variables de lisibilité à l'aide de FABRA n'avait pas donné de résultats avec succès. La répartition de ces données est présentée dans le tableau 29.

Tableau 29 : Données pour lesquelles aucune variable de lisibilité n'a été extraite

	Nom de données	Style de parole	Nombre de mots	Durée d'enregistrement (min)

1	C1_abcDALF_37	Monologue	597	9,52
2	C1_C2_Tendances_3	Dialogue	351	6,31
3	C1_C2_Tendances_4	Dialogue	869	11,66
4	C1_Edito_12	Dialogue	491	9,01
5	C1_Edito_3	Dialogue	333	6,67
6	C1_Edito_37	Dialogue	260	6,46
7	C1_Edito_39	Dialogue	504	8,17
8	C2_abcDALF_1	Dialogue	781	12,29
9	C2_abcDALF_10	Dialogue	860	14,92
10	C2_abcDALF_12	Dialogue	832	13,5
11	C2_abcDALF_2	Dialogue	711	14,37
12	C2_abcDALF_5	Monologue	1 216	14,79
13	C2_abcDALF_6	Dialogue	1 103	14,46
14	C2_abcDALF_7	Monologue	582	14,42
15	C2_abcDALF_8	Dialogue	749	14,1
16	C2_Reussir_le_Dalf_20	Monologue	829	10,55
17	C2_Reussir_le_Dalf_22	Dialogue	704	12,15

Le tableau 29 montre que toutes les données pour lesquelles les variables de lisibilité n'ont pas pu être extraites par FABRA correspondent à des données de niveau C. En outre, plus de la moitié d'entre elles constituent des données de dialogue. Comme indiqué à la section 3.2., des problèmes similaires se sont également posés lors de l'élaboration des données originales, et les données problématiques n'ont pas été incluses dans les données originales. Lors de la création des seconds modèles, les 17 données mentionnées précédemment ont également été exclues. En d'autres termes, 36 données au total se trouvent exclues du second modèle, dont 19 données qui ont posé des problèmes lors de la phase d'alignement automatique et 17 données qui ont posé des problèmes lors de l'extraction des variables de lisibilité. Comme mentionné ci-dessus, cette thèse n'aborde pas les questions techniques liées à l'alignement automatique ou à l'extraction automatique des variables de lisibilité. Compte tenu des caractéristiques des données à exclure, décrites dans cette section (tableaux 28 et 29), il paraît probable que les problèmes soient plus susceptibles de se produire

avec des données volumineuses, avec un grand nombre de mots et des durées d'enregistrement relativement longues.

Les données à utiliser dans les seconds modèles sont résumées ci-après. Tout d'abord, une liste de manuels est présentée dans le tableau 30. Lorsque les données diffèrent des données originales (les données utilisées dans les premiers modèles ; tableau 2), elles sont soulignées dans la colonne « nombre de données ».

Tableau 30 : Liste des manuels

Manuel	Éditeur	Année	Nombre de données	Durée d'enregistrement (min)
A1				
Alter Ego	Hachette	2006	26	Environ 17
écho	CLE International	2011	63	Environ 22
Texto	Hachette	2016	19	Environ 14
Cosmopolite	Hachette	2017	<u>103</u>	Environ 81
Atelier	Didier	2019	78	Environ 25
A2				
Entre Nous	Maison des langues	2016	70	Environ 23
Tendances	CLE International	2016	81	Environ 54
Texto	Hachette	2016	16	Environ 22
Cosmopolite	Hachette	2017	72	Environ 74
Défi	Maison des langues	2018	35	Environ 34
Atelier	Didier	2019	35	Environ 24
B1				
Entre Nous	Maison des langues	2016	72	Environ 28
Tendances	CLE International	2016	90	Environ 63
Cosmopolite	Hachette	2018	<u>50</u>	Environ 60

Défi	Maison des langues	2019	27	Environ 38
B2				
Édito	Didier	2006	31	Environ 35
LE DELF B2	Didier	2016	59	Environ 68
Entre Nous	Maison des langues	2017	20	Environ 14
Tendances	CLE International	2017	50	Environ 71
Cosmopolite	Hachette	2019	79	Environ 98
C1/C2				
Réussir le DALF	Didier	2007	27	Environ 67
abc DALF	CLE International	2014	45	Environ 136
Le DALF 100 % réussite	Didier	2017	71	Environ 90
Édito	Didier	2018	63	Environ 84
Tendances	CLE International	2019	5	Environ 128

Ensuite, nous présentons la description des données pour chaque style de parole. Tout d'abord, le tableau 31 montre la description des données pour le corpus entier.

Tableau 31 : Description du corpus pour les seconds modèles

Niveau	Nombre de données	Nombre de mots		Durée d'enregistrement (min)	
	Total	Total	Moyenne	Total	Moyenne
A1	289	17 238	59,44	Environ 159	0,55
A2	309	30 738	99,48	Environ 232	0,75
B1	239	30 119	125,5	Environ 188	0,79
B2	239	48 126	199,69	Environ 285	1,19
C1/C2	211	118 713	488,53	Environ 404	1,91
Total	1 287	244 934	185,14	Environ 1 268	0,99

Ensuite, le tableau 32 présente la description des données pour le dialogue.

Tableau 32 : Description des données (dialogue) pour les seconds modèles

Niveau	Nombre de données	Nombre de mots		Durée d'enregistrement (min)	
	Total	Total	Moyenne	Total	Moyenne
A1	204	13 168	64,23	Environ 119	0,58
A2	180	23 168	128,71	Environ 180	1
B1	145	23 167	158,68	Environ 144	1
B2	129	28 483	217,43	Environ 162	1,26
C1/C2	94	89 065	736,07	Environ 270	2,87
Total	752	177 051	226,12	Environ 875	1,16

Enfin, le tableau 33 présente la description des données pour le monologue.

Tableau 33 : Description des données (monologue) pour les seconds modèles

Niveau	Nombre de données	Nombre de mots		Durée d'enregistrement (min)	
	Total	Total	Moyenne	Total	Moyenne
A1	85	4 070	47,88	Environ 40	0,47
A2	129	7 570	58,68	Environ 52	0,4
B1	94	6 952	73,96	Environ 44	0,47
B2	110	19 643	178,57	Environ 123	1,12
C1/C2	117	29 648	243,02	Environ 134	1,15
Total	535	67 883	125,71	Environ 393	0,73

Sur la base de l'ensemble de ces données, un second modèle est envisagé.

5.2. Modèle SVM

Sur la base des données présentées à la section 5.1., nous avons créé et analysé un modèle SVM. Comme pour le premier modèle, une analyse corrélationnelle des variables sera d'abord effectuée à la section 5.2.1., ensuite les résultats du modèle SVM créé sur la base de cette analyse seront présentés à la section 5.2.2. Enfin, l'analyse de ce modèle sera menée à la section 5.2.3.

5.2.1. Analyse corrélationnelle des variables

Dans cette section, nous analysons la corrélation entre les 71 variables décrites aux sections 3.4.2. et 3.4.3. et le niveau du CECR, et nous envisageons les variables indépendantes à incorporer dans le second modèle de SVM. Le coefficient de corrélation de Spearman est utilisé afin d'examiner la relation corrélationnelle dans chaque style de parole (corpus entier, dialogue et monologue). Tout d'abord, les variables pour lesquelles la valeur absolue du coefficient de corrélation avec le niveau du CECR s'avère supérieure à 0,2 ($p < ,01$) sont présentées ci-dessous. Le tableau 34 énumère ces variables pour le corpus entier.

Tableau 34 : 32 variables pour le corpus entier (p < ,01)

Variables	Corrélation Spearman	Variables	Corrélation Spearman
LENsntWRD	0,5	LEXgrdFSOOUA1	-0,59
SYNdevNPHRS	0,5	LEXgrdFMLA1	-0,56
SYNdevTUL	0,48	LEXgrdFFOA1	-0,55
LEXdvrFLC	0,43	REAfrmKM	-0,44
SYNclsLEN	0,36	LEXgrdBA1	-0,41
SYNtnsfPPS	0,34	LEXsopFK1	-0,35
SYNtnsfPP	0,33	SYNtnsfINDP	-0,3
SYNposADP	0,32	LEXsopLLK1	-0,26
SYNtnsfTAP	0,32	SYNposINTJ	-0,24
syllables_between_pauses	0,31	LEXnrmIMG	-0,21
words_between_pauses	0,31	LEXsopGK1	-0,2
LEXnghPHO	0,27		
SYNposSCONJ	0,27		
SYNtnsfSUBP	0,25		
SYNposNOUN	0,25		
SYNtnsfINF	0,24		
LENwrdSYL	0,24		
SYNtnsfINDFS	0,22		
SYNposADJ	0,21		
SYNtnsfINDI	0,21		
SYNtnsfINDPQP	0,21		

La classification catégorielle des variables ci-dessus concernant le corpus entier est présentée dans le tableau 35.

Tableau 35 : Catégorisation de 32 variables pour le corpus entier

Catégorie	Nombre de variables	Variables
Temps verbal	9	SYNtnsfPPS, SYNtnsfPP, SYNtnsfTAP, SYNtnsfSUBP, SYNtnsfINF, SYNtnsfINDFS, SYNtnsfINDI, SYNtnsfINDPQP, SYNtnsfINDP
Niveau de difficulté du lexique	7	LEXgrdFSOOUA1, LEXgrdFMLA1, LEXgrdFFOA1, LEXgrdBA1, LEXsopFK1, LEXsopLLK1, LEXsopGK1
Longueur de phrase	5	LENsntWRD, SYNdevNPHRS, SYNdevTUL, SYNclsLEN, LENwrdSYL
Syntaxe	5	SYNposADP, SYNposSCONJ, SYNposNOUN, SYNposADJ, SYNposINTJ
Débit de parole	2	syllables_between_pauses, words_between_pauses
Formule de lisibilité	1	REAfrmKM
Phonologie	1	LEXnghPHO
Richesse lexicale	1	LEXdvrFLC
Type de vocabulaire	1	LEXnrmIMG

Les tableaux 34 et 35 montrent que, concernant le corpus entier, les niveaux du CECR se trouvent fortement corrélés avec les variables relatives à la longueur de phrase et au niveau de difficulté du lexique en particulier. Par catégorie, nous constatons que, en plus de ces deux variables, les niveaux du CECR ont également tendance à être corrélés avec les variables relatives aux temps verbaux et à la syntaxe.

Le tableau 36, quant à lui, énumère les variables pour lesquelles la valeur absolue du coefficient de corrélation avec le niveau du CECR se révèle supérieure à 0,2 ($p < ,01$) concernant le dialogue.

Tableau 36 : 28 variables pour le dialogue (p < ,01)

Variables	Corrélation Spearman	Variables	Corrélation Spearman
LENsntWRD	0,45	LEXgrdFSOOUA1	-0,58
SYNdevNPHRS	0,45	LEXgrdFFOA1	-0,54
SYNdevTUL	0,44	LEXgrdFMLA1	-0,51
LEXdvrFLC	0,44	LEXgrdBA1	-0,4
SYNposSCONJ	0,36	REAfrmKM	-0,39
SYNtnsfPP	0,34	SYNtnsfINDP	-0,35
words_between_pauses	0,33	LEXnrmIMG	-0,3
syllables_between_pauses	0,33	SYNposINTJ	-0,29
SYNtnsfTAP	0,33	LEXsopFK1	-0,28
SYNposADP	0,32		
SYNtnsfPPS	0,32		
SYNtnsfINDI	0,31		
SYNtnsfSUBP	0,29		
SYNtnsfINDPQP	0,27		
SYNclsLEN	0,27		
SYNtnsfINF	0,24		
SYNtnsfINDFS	0,24		
SYNtnsfINDPC	0,24		
LEXnghPHO	0,2		

La classification catégorielle des variables ci-dessus concernant le dialogue est présentée dans le tableau 37.

Tableau 37 : Catégorisation de 28 variables pour le dialogue

Catégorie	Nombre de variables	Variables
Temps verbal	10	SYNtnsfPP, SYNtnsfTAP, SYNtnsfPPS, SYNtnsfINDI, SYNtnsfSUBP, SYNtnsfINDPQP, SYNtnsfINF, SYNtnsfINDFS, SYNtnsfINDPC, SYNtnsfINDP
Niveau de difficulté du lexique	5	LEXgrdFSOOUA1, LEXgrdFFOA1, LEXgrdFMLA1, LEXgrdBA1, LEXsopFK1
Longueur de phrase	4	LENsntWRD, SYNdevNPHRS, SYNdevTUL, SYNclsLEN
Syntaxe	3	SYNposSCONJ, SYNposADP, SYNposINTJ
Débit de parole	2	words_between_pauses, syllables_between_pauses
Formule de lisibilité	1	REAfrmKM
Phonologie	1	LEXnghPHO
Richesse lexicale	1	LEXdvrFLC
Type de vocabulaire	1	LEXnrmIMG

Les tableaux 36 et 37 montrent que, concernant le dialogue, les variables relatives au niveau de difficulté du lexique et à la longueur de phrase se trouvent fortement corrélées avec le niveau du CECR. En examinant les catégories, nous constatons que, en plus de ces deux variables, les variables liées aux temps verbaux, qui sont corrélées avec le niveau du CECR, s'avèrent particulièrement nombreuses. Ces tendances sont similaires aux résultats du corpus entier.

Enfin, dans le tableau 38, la liste des variables pour lesquelles la valeur absolue du coefficient de corrélation avec le niveau du CECR s'avère supérieure à 0,2 ($p < ,01$) pour le monologue est indiquée.

Tableau 38 : 28 variables pour le monologue (p < ,01)

Variables	Corrélation Spearman	Variables	Corrélation Spearman
LENsntWRD	0,57	LEXgrdFMLA1	-0,58
SYNdevNPHRS	0,57	LEXgrdFSOOUA1	-0,57
SYNdevTUL	0,53	LEXgrdFFOA1	-0,53
LEXdvrFLC	0,5	REAfrmKM	-0,5
SYNclsLEN	0,46	LEXsopFK1	-0,38
SYNtnsfPPS	0,35	LEXgrdBA1	-0,37
SYNtnsfTAP	0,31	LEXsopLLK1	-0,34
LEXnghPHO	0,31	SYNposPRON	-0,22
SYNtnsfPP	0,3		
SYNposNOUN	0,28		
LENwrdSYL	0,27		
SYNposADP	0,27		
SYNposADV	0,26		
words_between_pauses	0,25		
syllables_between_pauses	0,25		
SYNposADJ	0,24		
SYNposSCONJ	0,23		
SYNtnsfINDFS	0,23		
SYNtnsfSUBP	0,21		
SYNtnsfINF	0,21		

La classification catégorielle des variables ci-dessus concernant le monologue est présentée dans le tableau 39.

Tableau 39 : Catégorisation de 28 variables pour le monologue

Catégorie	Nombre de variables	Variables
Niveau de difficulté du lexique	6	LEXgrdFMLA1, LEXgrdFSOOUA1, LEXgrdFFOA1, LEXsopFK1, LEXgrdBA1, LEXsopLLK1
Syntaxe	6	SYNposNOUN, SYNposADP, SYNposADV, SYNposADJ, SYNposSCONJ, SYNposPRON
Temps verbal	6	SYNtnsfPPS, SYNtnsfTAP, SYNtnsfPP, SYNtnsfINDFS, SYNtnsfSUBP, SYNtnsfINF
Longueur de phrase	5	LENsntWRD, SYNdevNPHRS, SYNdevTUL, SYNclsLEN, LENwrdSYL
Débit de parole	2	words_between_pauses, syllables_between_pauses
Formule de lisibilité	1	REAfrmKM
Phonologie	1	LEXnghPHO
Richesse lexicale	1	LEXdvrFLC

Les tableaux 38 et 39 montrent que pour le monologue aussi, il existe notamment une forte corrélation entre le niveau du CECR et les variables liées à la longueur de phrase et au niveau de difficulté du lexique. En matière de catégories, nous constatons qu'en plus de ces deux variables, il existe également de nombreuses variables liées à la syntaxe et aux temps verbaux qui se trouvent corrélées avec le niveau du CECR. Ces grandes tendances s'avèrent similaires aux résultats pour le corpus entier et le dialogue.

Cependant, afin de déterminer les variables finales à utiliser pour créer le modèle SVM, il semble nécessaire de réduire la colinéarité autant que possible. Par conséquent, comme dans le processus de création du premier modèle, concernant chaque style de parole, les corrélations entre les variables ci-dessus sont examinées à l'aide du coefficient de corrélation de Spearman. Ainsi, par exemple, une corrélation de 0,98 ($p < ,01$) a été trouvée entre *words_between_pauses* et *syllables_between_pauses*, indiquant le débit de parole,

à la fois pour le dialogue et le monologue. De plus, une corrélation de 1 ($p < ,01$) a été observée entre *LENsntWRD* et *SYNdevNPHRS* qui indiquent la longueur de phrase, également pour le dialogue et le monologue. En ce qui concerne les variables pour lesquelles la valeur absolue du coefficient de corrélation avec d'autres variables s'avère supérieure à 0,7, lorsque plusieurs variables sont incluses dans une même catégorie (tableaux 35, 37 et 39), seule celle qui offre la corrélation la plus forte avec le niveau du CECR ont été conservées pour chaque catégorie. Dès lors, les autres variables ont été exclues. Cependant, pour les variables liées aux temps verbaux et à la syntaxe, elles sont laissées comme variables à inclure dans le modèle si la valeur absolue du coefficient de corrélation s'avère supérieure à 0,2.

Après ce processus, les variables indépendantes à inclure dans le modèle SVM ont été déterminées pour chaque style de parole. Des diagrammes en boîte à moustaches sont également présentés concernant les variables les plus fortement corrélées avec le niveau du CECR concernant les dialogues et les monologues. Le tableau 40 présente les variables finales pour le corpus entier.

Tableau 40 : 23 variables finales pour le corpus entier (p < ,01)[62]

Variables	Catégorie	Corrélation Spearman
LENsntWRD	Longueur de phrase	0,5
LEXdvrFLC	Richesse lexicale	0,43
SYNtnsfPPS	Temps verbal	0,34
SYNtnsfPP	Temps verbal	0,33
SYNposADP	Syntaxe	0,32
SYNtnsfTAP	Temps verbal	0,32
syllables_between_pauses	Débit de parole	0,31
LEXnghPHO	Phonologie	0,27
SYNposSCONJ	Syntaxe	0,27
SYNtnsfSUBP	Temps verbal	0,25
SYNposNOUN	Syntaxe	0,25
SYNtnsfINF	Temps verbal	0,24
SYNtnsfINDFS	Temps verbal	0,22
SYNposADJ	Syntaxe	0,21
SYNtnsfINDI	Temps verbal	0,21
SYNtnsfINDPQP	Temps verbal	0,21
LEXsopGK1	Difficulté du lexique	-0,2
LEXnrmIMG	Type de vocabulaire	-0,21
SYNposINTJ	Syntaxe	-0,24
SYNtnsfINDP	Temps verbal	-0,3
LEXgrdBA1	Difficulté du lexique	-0,41
REAfrmKM	Formule de lisibilité	-0,44
LEXgrdFSOOUA1	Difficulté du lexique	-0,59

Les variables finales du dialogue sont ensuite présentées dans le tableau 41 tandis que la figure 13 propose le diagramme en boîte à moustaches pour *LEXgrdFSOOUA1* (proportion de lemmes de niveau A1 ; r = -0,58).

[62] Dans les tableaux 40, 41 et 42, les variables ombrées constituent des variables de fluidité et les autres des variables de lisibilité.

Tableau 41 : 21 variables finales pour le dialogue (p < ,01)

Variables	Catégorie	Corrélation Spearman
LENsntWRD	Longueur de phrase	0,45
LEXdvrFLC	Richesse lexicale	0,44
SYNposSCONJ	Syntaxe	0,36
SYNtnsfPP	Temps verbal	0,34
words_between_pauses	Débit de parole	0,33
SYNtnsfTAP	Temps verbal	0,33
SYNposADP	Syntaxe	0,32
SYNtnsfPPS	Temps verbal	0,32
SYNtnsfINDI	Temps verbal	0,31
SYNtnsfSUBP	Temps verbal	0,29
SYNtnsfINDPQP	Temps verbal	0,27
SYNtnsfINF	Temps verbal	0,24
SYNtnsfINDFS	Temps verbal	0,24
SYNtnsfINDPC	Temps verbal	0,24
LEXnghPHO	Phonologie	0,2
SYNposINTJ	Syntaxe	-0,29
LEXnrmIMG	Type de vocabulaire	-0,3
SYNtnsfINDP	Temps verbal	-0,35
REAfrmKM	Formule de lisibilité	-0,39
LEXgrdBA1	Difficulté du lexique	-0,4
LEXgrdFSOOUA1	Difficulté du lexique	-0,58

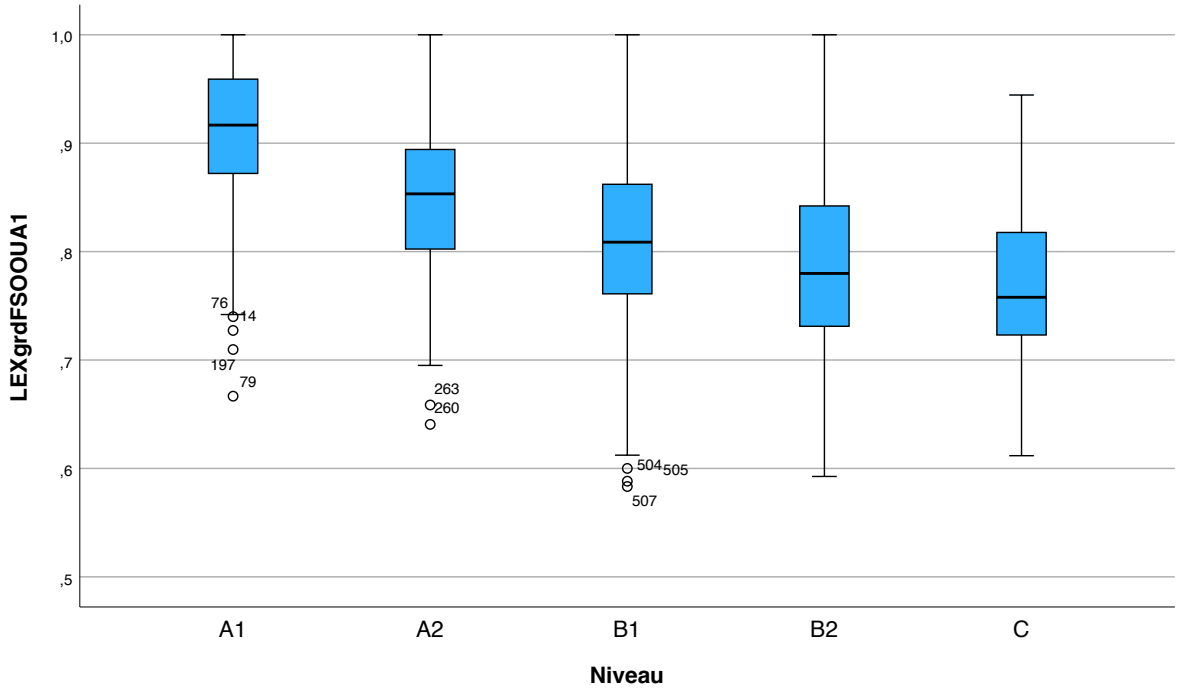

Figure 13 : Boîte à moustaches de *LEXgrdFSOOUA1*

Enfin, le tableau 42 présente les variables finales concernant le monologue tandis que la figure 14 propose le diagramme en boîte à moustaches pour *LEXgrdFMLA1* (proportion de lemmes de niveau A1 ; r = -0,58).

Tableau 42 : 19 variables finales pour le monologue (p < ,01)

Variables	Catégorie	Corrélation Spearman
LENsntWRD	Longueur de phrase	0,57
LEXdvrFLC	Richesse lexicale	0,5
SYNtnsfPPS	Temps verbal	0,35
SYNtnsfTAP	Temps verbal	0,31
LEXnghPHO	Phonologie	0,31
SYNtnsfPP	Temps verbal	0,3
SYNposNOUN	Syntaxe	0,28
SYNposADP	Syntaxe	0,27
SYNposADV	Syntaxe	0,26
words_between_pauses	Débit de parole	0,25
SYNposADJ	Syntaxe	0,24
SYNposSCONJ	Syntaxe	0,23
SYNtnsfINDFS	Temps verbal	0,23
SYNtnsfSUBP	Temps verbal	0,21
SYNtnsfINF	Temps verbal	0,21
SYNposPRON	Syntaxe	-0,22
LEXgrdBA1	Difficulté du lexique	-0,37
REAfrmKM	Formule de lisibilité	-0,5
LEXgrdFMLA1	Difficulté du lexique	-0,58

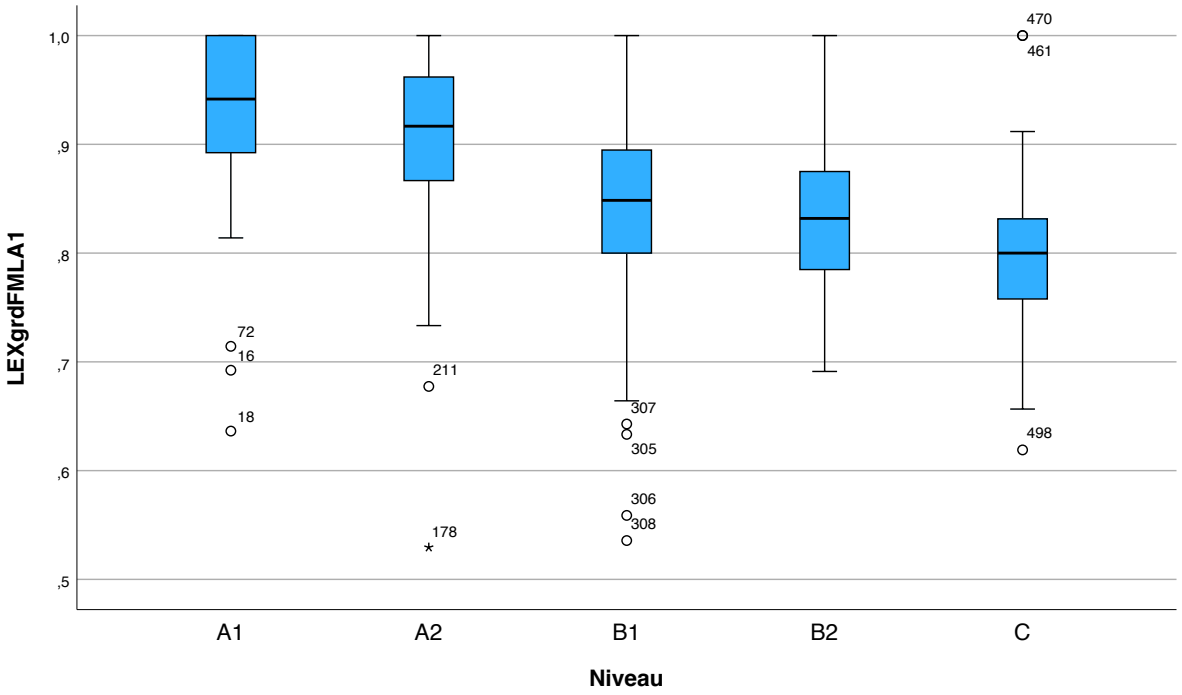

Figure 14 : Boîte à moustaches à *LEXgrdFMLA1*

En résumé des résultats de l'analyse de corrélation des variables dans chacun des styles de parole susmentionnés, les variables présentant la corrélation la plus élevée avec le niveau du CECR comportent *LEXgrdFSOOUA1* (-0,59) pour le corpus entier, *LEXgrdFSOOUA1* (-0,58) pour les dialogues et *LEXgrdFMLA1* (-0,58) pour les monologues, à savoir toutes les variables liées au niveau de vocabulaire. La deuxième variable la plus corrélée, dans tous les styles de parole, est *LENsntWRD*, une variable liée à la longueur des phrases, avec un coefficient de corrélation de 0,5 pour le corpus entier, 0,45 pour le dialogue et 0,57 pour le monologue. Ainsi, les résultats montrent que, pour chaque style de parole, la tendance des variables fortement corrélées avec le niveau du CECR s'avère similaire. En revanche, les différences, surtout entre le dialogue et le monologue, se situent au niveau du nombre de variables liées aux temps verbaux et à la syntaxe, qui sont corrélées avec le niveau du CECR. En effet, le nombre de variables liées aux temps verbaux paraît plus élevé pour le dialogue, tandis que celles liées à la syntaxe se révèlent plus élevées pour le monologue.

Jusqu'à présent, nous n'avons mentionné que les variables de lisibilité, mais lorsque nous examinons celles de fluidité, une seule d'entre elles a été incluse dans le jeu de variables final concernant chaque style de parole. Pour le corpus entier, il s'agit de *syllables_between_pauses* (r = 0,31) ; *words_between_pauses*

pour le dialogue (r = 0,33) et pour le monologue également (r = 0,25), qui sont toutes liées au débit de parole. L'une des caractéristiques du monologue est que la force de la corrélation de cette variable de fluidité avec les niveaux du CECR ne s'avère pas très élevée par rapport à d'autres variables du monologue.

5.2.2. Résultats du modèle SVM

En utilisant les variables des tableaux 40, 41 et 42, un modèle SVM a été construit afin de prédire la facilité d'écoute des documents audios selon l'échelle du CECR. Pour chaque style de parole, trois modèles ont été créés : un modèle utilisant uniquement les variables de lisibilité, un autre uniquement basé sur les variables de fluidité et un troisième utilisant à la fois les variables de lisibilité et de fluidité. Comme indiqué dans la section précédente, il n'existe qu'une seule variable de fluidité pour chaque style de parole. Cependant, afin de faciliter la comparaison des résultats globaux avec ceux de premier modèle, trois modèles ont été créés pour chaque style de parole.

En ce qui concerne la recherche des hyperparamètres, nous les avons explorés à l'aide d'une grille de recherche. Dans le SVM, les valeurs de C = [1 ; 10 ; 100 ; 1000], adoptées dans le premier modèle, sont utilisés comme coefficient de régularisation, les kernels explorés s'avèrent linéaire et RBF comme types de noyau et (γ) = [1e-3 ; 1e-4] comme hyperparamètres du noyau RBF. Les performances des modèles ont été évaluées au moyen d'une procédure de validation croisée à dix plis et à l'aide des mesures d'évaluation d'exactitude, de précision, de rappel, de macro F1 et de F1 pondérée. En outre, la macro F1 pour chacun des cinq niveaux du CECR s'avère également indiquée.

Comme dans le premier modèle, 80 % des données servent à la phase d'entraînement, 10 % pour le développement et 10 % pour le test. Les moyennes des dix plis obtenus pour chaque score sont rapportées[63]. Les résultats du modèle SVM sont présentés dans le tableau 43.

[63] Pour plus d'informations sur la méthode, voir la section 4.1.2.

Tableau 43 : Résultat pour la tâche de prédiction du niveau de facilité d'écoute (C = [1 ; 10 ; 100 ; 1000])

	Lisibilité + Fluidité		
	Corpus entier	Dialogue	Monologue
	avg (std)	avg (std)	avg (std)
Exactitude	0,46 (0,04)	0,51 (0,05)	0,45 (0,07)
Précision	0,45 (0,04)	0,5 (0,03)	0,49 (0,07)
Rappel	0,44 (0,04)	0,49 (0,04)	0,47 (0,07)
Macro F1	0,44 (0,04)	0,47 (0,04)	0,45 (0,07)
F1 pondérée	0,44 (0,04)	0,49 (0,06)	0,45 (0,02)
F1-A1	0,61 (0,09)	0,7 (0,12)	0,58 (0,13)
F1-A2	0,47 (0,06)	0,52 (0,07)	0,41 (0,09)
F1-B1	0,26 (0,09)	0,28 (0,09)	0,32 (0,17)
F1-B2	0,38 (0,07)	0,4 (0,1)	0,44 (0,14)
F1-C	0,46 (0,08)	0,46 (0,07)	0,51 (0,1)

	Lisibilité		
	Corpus entier	Dialogue	Monologue
	avg (std)	avg (std)	avg (std)
Exactitude	0,46 (0,04)	0,49 (0,04)	0,45 (0,05)
Précision	0,46 (0,07)	0,48 (0,05)	0,49 (0,05)
Rappel	0,5 (0,04)	0,47 (0,03)	0,48 (0,06)
Macro F1	0,43 (0,04)	0,46 (0,04)	0,46 (0,05)
F1 pondérée	0,45 (0,04)	0,47 (0,05)	0,45 (0,02)
F1-A1	0,64 (0,09)	0,67 (0,1)	0,62 (0,13)
F1-A2	0,48 (0,06)	0,47 (0,07)	0,42 (0,08)
F1-B1	0,2 (0,08)	0,29 (0,15)	0,31 (0,19)
F1-B2	0,4 (0,07)	0,35 (0,17)	0,43 (0,14)
F1-C	0,46 (0,11)	0,5 (0,09)	0,5 (0,09)

	Fluidité		
	Corpus entier	Dialogue	Monologue
	avg (std)	avg (std)	avg (std)
Exactitude	0,29 (0,03)	0,28 (0,06)	0,2 (0,05)
Précision	0,24 (0,07)	0,24 (0,13)	0,14 (0,07)
Rappel	0,27 (0,04)	0,25 (0,05)	0,21 (0,05)
Macro F1	0,23 (0,04)	0,2 (0,06)	0,12 (0,04)
F1 pondérée	0,24 (0,03)	0,22 (0,04)	0,14 (0,03)
F1-A1	0,25 (0,08)	0,33 (0,15)	0,01 (0,04)
F1-A2	0,41 (0,05)	0,37 (0,09)	0,28 (0,05)
F1-B1	0,03 (0,06)	0,07 (0,07)	0 (0)
F1-B2	0,21 (0,09)	0,05 (0,08)	0,17 (0,16)
F1-C	0,24 (0,15)	0,17 (0,13)	0,16 (0,14)

Sur la base des résultats ci-dessus, le modèle SVM a été évalué en matière de différences entre les styles de parole et les niveaux du CECR.

5.2.3. Analyse du modèle SVM

Nous avons consulté le tableau 43 afin d'analyser les résultats du modèle SVM. En comparant les résultats pour chaque style de parole, les résultats pour *le modèle Lisibilité + Fluidité* et *le modèle Lisibilité*, en matière d'exactitude, la précision, le rappel, la macro F1 et la F1 pondérée se révèlent tous similaires. Par ailleurs, les résultats du *modèle Fluidité* se trouvent nettement inférieurs à ceux du *modèle Lisibilité + Fluidité*, en particulier pour le monologue. Dès lors, il paraît difficile d'affirmer que, pour le second modèle de SVM, les variables de fluidité contribuent de manière importante à la performance du modèle pour tous les styles de parole, et cela s'avère remarquable pour le monologue.

Ensuite, nous avons analysé chaque modèle à partir de la macro F1 de chaque niveau du CECR. Concernant *le modèle Lisibilité + Fluidité* et *le modèle Lisibilité*, la macro F1 pour A1 détient la valeur la plus élevée et la macro F1 pour B1 la valeur la plus basse, pour tous les styles de parole. Par ailleurs, dans *le modèle Fluidité*, la macro F1 semble souvent très faible, surtout aux niveaux B1 et B2. Ainsi, la performance prédictive de la facilité d'écoute diffère selon le niveau du

CECR. Finalement, l'ensemble de ces résultats montre qu'il est particulièrement difficile de prédire la facilité d'écoute au niveau B1.

5.3. Modèle wav2vec

Dans cette section, le modèle wav2vec est considéré. Dans la section 5.3.1., les résultats du modèle wav2vec créé en utilisant les données audios présentées dans la section 5.1. seront indiqués, tandis que l'analyse sera présentée dans la section 5.3.2.

5.3.1. Résultats du modèle wav2vec

Le tableau 44 montre les résultats du second modèle de wav2vec pour chaque style de parole. Ce modèle a été entraîné et évalué avec les données qui ont été divisées en 80 % de données pour l'entraînement, 10 % pour le développement et 10 % pour le test pour la validation croisée à dix plis.

Tableau 44 : Résultat du second modèle wav2vec pour la tâche de prédiction du niveau de facilité d'écoute

	Corpus entier	Dialogue	Monologue
	avg (std)	avg (std)	avg (std)
Exactitude	0,31 (0,06)	0,31 (0,06)	0,24 (0,07)
Précision	0,26 (0,04)	0,25 (0,03)	0,26 (0,05)
Rappel	0,24 (0,15)	0,24 (0,12)	0,16 (0,12)
Macro F1	0,19 (0,06)	0,17 (0,05)	0,16 (0,06)
F1 pondérée	0,19 (0,06)	0,17 (0,06)	0,13 (0,06)
F1-A1	0,5 (0,08)	0,45 (0,13)	0,44 (0,13)
F1-A2	0,23 (0,09)	0,13 (0,13)	0,25 (0,11)
F1-B1	0,06 (0,09)	0,07 (0,1)	0 (0)
F1-B2	0,1 (0,08)	0,2 (0,13)	0,07 (0,12)
F1-C	0,02 (0,04)	0,02 (0,06)	0,03 (0,06)

5.3.2. Analyse du modèle wav2vec

Le tableau 44 montre que les performances des modèles s'avèrent extrêmement faibles : la macro F1 ne s'élève qu'à 0,19 pour le corpus entier, à 0,17 pour le dialogue et à 0,16 pour le monologue. Par conséquent, il s'avère très difficile

d’utiliser wav2vec afin de prédire le niveau du CECR à partir des données audios lors de la construction du second modèle.

En examinant la macro F1 pour chaque niveau, la performance prédictive la plus élevée est observée au niveau A1 pour tous les styles de parole. Au niveau A2, la performance de prédiction chute à la moitié ou moins de la valeur de la macro F1 au niveau A1 pour tous les styles de parole. Aux niveaux B1 et supérieurs, tous montrent des macro F1 inférieures à 0,2, ce qui indique une très faible performance de prédiction. Cela signifie que la prédiction ne paraît possible que pour le niveau A1.

5.4. Modèle CamemBERT

Dans cette section, en utilisant les transcriptions automatiquement générées à partir des audios, le modèle CamemBERT est examiné. Les résultats seront présentés dans la section 5.4.1. et l'analyse dans la section 5.4.2.

5.4.1. Résultats du modèle CamemBERT

Les résultats du modèle CamemBERT se trouvent dans le tableau 45. Comme pour tous les modèles, la validation croisée à dix plis a été effectuée, tandis que le modèle a été entraîné avec 80 % des données utilisées pour la phase d'entraînement, 10 % pour le développement et 10 % pour le test. Trois modèles ont été créés en fonction du style de parole.

Tableau 45 : Résultat du second modèle CamemBERT pour la tâche de prédiction du niveau de facilité d'écoute

	Corpus entier	Dialogue	Monologue
	avg (std)	avg (std)	avg (std)
Exactitude	0,57 (0,05)	0,6 (0,06)	0,55 (0,08)
Précision	0,59 (0,07)	0,56 (0,06)	0,58 (0,09)
Rappel	0,55 (0,05)	0,54 (0,04)	0,55 (0,08)
Macro F1	0,54 (0,05)	0,52 (0,05)	0,55 (0,08)
F1 pondérée	0,56 (0,06)	0,58 (0,05)	0,56 (0,08)
F1-A1	0,78 (0,05)	0,83 (0,08)	0,73 (0,12)
F1-A2	0,59 (0,09)	0,63 (0,06)	0,5 (0,15)
F1-B1	0,47 (0,1)	0,4 (0,14)	0,38 (0,17)
F1-B2	0,43 (0,12)	0,44 (0,14)	0,53 (0,12)
F1-C	0,44 (0,16)	0,33 (0,24)	0,59 (0,17)

5.4.2. Analyse du modèle CamemBERT

D'après les résultats ci-dessus, les valeurs de la macro F1 se situent autour de 0,5 pour tous les styles de parole, ce qui indique que le modèle CamemBERT possède une certaine capacité de prédiction. Lorsque la performance de prédiction

a été examinée pour chaque niveau du CECR, il a été montré que les valeurs de macro F1 s'avéraient les plus élevées au niveau A1, pour tous les styles de parole.

Concernant le monologue, la valeur de la macro F1 s'avère inférieure à 0,5 uniquement pour le niveau B1. Par conséquent, nous constatons qu'il présente une performance de prédiction relativement élevée, et ce même pour les niveaux B2 et C, qui ont tendance à montrer une baisse de performance prédictive au sein du corpus entier et du dialogue. En particulier, pour le monologue, la valeur de la macro F1 au niveau C apparaît plus élevée par rapport aux deux autres styles de parole. Malgré la performance similaire du modèle global pour les trois styles de parole, en particulier le dialogue et le monologue, il semble clair que la réalité interne du modèle diffère lorsqu'il est vérifié à chaque niveau.

5.5. Comparaison des modèles

Dans cette section, nous établissons la comparaison parmi les résultats des seconds modèles par SVM, wav2vec et CamemBERT. Pour ce faire, nous avons examiné le meilleur second modèle permettant de prédire la facilité d'écoute du français. Dans les tableaux 46, 47 et 48, les tableaux 43, 44 et 45 se trouvent récapitulés pour chaque style de parole. Dans ces tableaux, seules l'exactitude et la macro F1 et celle par niveau sont listées pour chaque modèle. En se basant sur ces tableaux, nous avons examiné l'axe des styles de parole. Tout d'abord, le tableau 46 présente les résultats de tous les modèles pour le corpus entier.

Tableau 46 : Résultat (SVM, wav2vec, CamemBERT) pour la tâche de prédiction du niveau de facilité d'écoute sur l'ensemble des données

	SVM-LF	SVM-L	SVM-F	wav2vec	CamemBERT
Exactitude (std)	0,46 (0,04)	0,46 (0,04)	0,29 (0,03)	0,31 (0,06)	0,57 (0,05)
Macro F1 (std)	0,44 (0,04)	0,43 (0,04)	0,23 (0,04)	0,19 (0,06)	0,54 (0,05)
F1-A1 (std)	0,61 (0,09)	0,64 (0,09)	0,25 (0,08)	0,5 (0,08)	0,78 (0,05)
F1-A2 (std)	0,47 (0,06)	0,48 (0,06)	0,41 (0,05)	0,23 (0,09)	0,59 (0,09)
F1-B1 (std)	0,26 (0,09)	0,2 (0,08)	0,03 (0,06)	0,06 (0,09)	0,47 (0,1)
F1-B2 (std)	0,38 (0,07)	0,4 (0,07)	0,21 (0,09)	0,1 (0,08)	0,43 (0,12)
F1-C (std)	0,46 (0,08)	0,46 (0,11)	0,24 (0,15)	0,02 (0,04)	0,44 (0,16)

Puis, le tableau 47 présente les résultats de tous les modèles pour le dialogue.

Tableau 47 : Résultat (SVM, wav2vec, CamemBERT) pour la tâche de prédiction du niveau de facilité d'écoute sur les données de style dialogue

	SVM-LF	SVM-L	SVM-F	wav2vec	CamemBERT
Exactitude (std)	0,51 (0,05)	0,49 (0,04)	0,28 (0,06)	0,31 (0,06)	0,6 (0,06)
Macro F1 (std)	0,47 (0,04)	0,46 (0,04)	0,2 (0,06)	0,17 (0,05)	0,52 (0,05)
F1-A1 (std)	0,7 (0,12)	0,67 (0,1)	0,33 (0,15)	0,45 (0,13)	0,83 (0,08)
F1-A2 (std)	0,52 (0,07)	0,47 (0,07)	0,37 (0,09)	0,13 (0,13)	0,63 (0,06)
F1-B1 (std)	0,28 (0,09)	0,29 (0,15)	0,07 (0,07)	0,07 (0,1)	0,4 (0,14)
F1-B2 (std)	0,4 (0,1)	0,35 (0,17)	0,05 (0,08)	0,2 (0,13)	0,44 (0,14)
F1-C (std)	0,46 (0,07)	0,5 (0,09)	0,17 (0,13)	0,02 (0,06)	0,33 (0,24)

Enfin, le tableau 48 présente les résultats pour le monologue.

Tableau 48 : Résultat (SVM, wav2vec, CamemBERT) pour la tâche de prédiction du niveau de facilité d'écoute sur les données de style monologue

	SVM-LF	SVM-L	SVM-F	wav2vec	CamemBERT
Exactitude (std)	0,45 (0,07)	0,45 (0,05)	0,2 (0,05)	0,24 (0,07)	0,55 (0,08)
Macro F1 (std)	0,45 (0,07)	0,46 (0,05)	0,12 (0,04)	0,16 (0,06)	0,55 (0,08)
F1-A1 (std)	0,58 (0,13)	0,62 (0,13)	0,01 (0,04)	0,44 (0,13)	0,73 (0,12)
F1-A2 (std)	0,41 (0,09)	0,42 (0,08)	0,28 (0,05)	0,25 (0,11)	0,5 (0,15)
F1-B1 (std)	0,32 (0,17)	0,31 (0,19)	0 (0)	0 (0)	0,38 (0,17)
F1-B2 (std)	0,44 (0,14)	0,43 (0,14)	0,17 (0,16)	0,07 (0,12)	0,53 (0,12)
F1-C (std)	0,51 (0,1)	0,5 (0,09)	0,16 (0,14)	0,03 (0,06)	0,59 (0,17)

En observant uniquement la F1 pour chaque modèle, l'ordre de grandeur de la macro F1 est CamemBERT > SVM-LF > SVM-L > SVM-F > wav2vec pour le corpus entier et le dialogue et CamemBERT > SVM-L > SVM-LF > wav2vec > SVM-F pour le monologue. Concernant tous les styles de parole, les macro F1 dans SVM-F et wav2vec se trouvent inférieures à la moitié des valeurs de macro F1 dans CamemBERT, SVM-LF et SVM-L, ce qui montre un écart important. Ainsi, les modèles basés sur la transcription peuvent prédire le niveau du CECR du français dans une certaine mesure dans le second modèle. En revanche, il paraît difficile de le réaliser avec des caractéristiques acoustiques.

D'une part, en considérant l'importance de la prédiction du niveau du CECR à partir des caractéristiques acoustiques dans les résultats du SVM-F et de wav2vec, ces deux modèles proposent une performance de prédiction très faible. D'autre part, à propos des résultats du SVM, concernant le corpus entier et le dialogue, les valeurs de macro F1 se révèlent plus grandes pour le SVM-LF, qui prend également en compte les variables de fluidité, que pour le SVM-L, qui ne considère que les variables de lisibilité. Cependant, dans les deux cas, la valeur de macro F1 n'a

augmenté que de 0,01, il paraît ainsi difficile d'affirmer que les caractéristiques acoustiques contribuent de manière importante à la performance de prédiction du niveau du CECR. En revanche, concernant le monologue, les résultats du SVM-LF montrent une macro F1 plus faible que les résultats du SVM-L. Toutefois, la différence ne s'élève qu'à 0,01. En tout état de cause, il n'a pas été possible de confirmer que les variables de fluidité ou les caractéristiques acoustiques exercent une influence importante sur les performances de prédiction pour tous les styles de parole.

Ainsi, que le SVM ou le CamemBERT soit plus utile afin de prédire la facilité d'écoute du français à partir de transcriptions, il paraît clair, d'après la valeurs de macro F1, que le CamemBERT constitue un meilleur modèle de prédiction pour les deux styles de parole. Cependant, dans l'ensemble, les résultats des modèles SVM-LF et SVM-L ne s'avèrent pas tellement inférieurs, ce qui rend leurs performances dignes d'être évaluées.

Ensuite, en examinant les résultats de chaque modèle par niveau du CECR, nous constatons que, pour le corpus entier et le dialogue, la performance prédictive a tendance à être fondamentalement plus élevée au niveau A, en particulier A1, et plus faible au niveau B, en particulier B1. En ce qui concerne les monologues, les modèles basés sur la transcription tendent à présenter une performance prédictive pour le niveau C supérieure à celle observée pour le corpus entier et les dialogues. Par conséquent, nous pouvons constater qu'il existe des différences dans les niveaux du CECR qui se révèlent plus faciles à prédire en fonction du style de parole et du modèle.

5.6. Conclusion

Dans ce chapitre, un second modèle de prédiction de la facilité d'écoute du français est examiné en utilisant des données à partir desquelles des transcriptions sont automatiquement générées. En outre, l'alignement de la parole s'avère également effectué automatiquement sur la base des transcriptions. L'analyse des seconds modèles de la facilité d'écoute se termine par des réponses à cinq questions de recherche.

Concernant la première question, *« Dans quelle mesure les informations extraites du signal vocal aident-elles à la prédiction de la facilité d'écoute ? »*, la performance de prédiction du SVM-F et de wav2vec ne s'avérait pas très élevée, quel que soit le style de parole. Cela suggère la difficulté de prédire les niveaux du CECR de la facilité d'écoute uniquement à partir de l'audio.

Pour la deuxième question, *« Dans quelle mesure les informations linguistiques provenant de la transcription aident-elles à la prédiction de la facilité d'écoute ? »*, pour tous les styles de parole, la performance de prédiction de CamemBERT et de SVM-L semblait élevée, ce qui indique que ces modèles peuvent prédire le niveau du CECR des documents audios français dans une certaine mesure, en utilisant uniquement les informations de transcription. En particulier, les performances de CamemBERT s'avéraient relativement élevées, tandis que celles du modèle SVM n'étaient pas particulièrement inférieures en comparaison.

En ce qui concerne la troisième question, *« La précision de la prédiction de la facilité d'écoute varie-t-elle en fonction du style de parole ? »*, dans chaque modèle, globalement, il n'existe pas d'exemples de grandes différences concernant les performances de prédiction dues à des différences dans le style de parole.

Pour la quatrième question, « *Existe-t-il des différences dans la performance de prédiction à différents niveaux ? »*, comme mentionné ci-dessus, même si la performance de prédiction ne diffère pas beaucoup parmi les modèles pour chaque style de parole, la performance varie en fonction du niveau du CECR. Dans l'ensemble, la performance de prédiction s'avérait élevée au niveau A1 et faible au niveau B1. En outre, pour le monologue, la performance de prédiction s'est révélée relativement élevée même au niveau C.

Enfin, concernant la cinquième question, *« Quel est le meilleur modèle permettant de prédire la facilité d'écoute ? »*, dans le second modèle, le modèle

CamemBERT construit à partir de transcriptions a obtenu les performances les plus élevées pour tous les styles de parole. Ainsi, nous concluons que le modèle CamemBERT constitue aussi le meilleur modèle lorsqu'on travaille sur des données transcrites et annotées automatiquement.

Les résultats ci-dessus présentent des différences importantes par rapport aux résultats du premier modèle présenté au chapitre 4. Les aspects qui diffèrent et les raisons de ces différences seront examinés dans le chapitre suivant.

6. Comparaison du premier modèle et du second modèle

Ce chapitre compare les résultats des premier et second modèles, et présente les réponses aux deux questions de recherche introduites à la section 3.3.2. :

1. Existe-t-il une différence de performance entre le premier et le second modèle ?
2. Quels sont les inconvénients du traitement automatique des données ?

Tout d'abord, les résultats pour les deux modèles dans l'ensemble des données sont indiqués[64] dans le tableau 49.

[64] Les tableaux 49, 50 et 51 résument les tableaux 25, 26, 27, 46, 47 et 48.

Tableau 49 : Comparaison des résultats sur l'ensemble des données

		SVM-LF	SVM-L	SVM-F	wav2vec	CamemBERT
Exactitude (std)	Modèle 1	0,52 (0,05)	0,49 (0,02)	0,39 (0,05)	0,31 (0,04)	0,63 (0,05)
	Modèle 2	0,46 (0,04)	0,46 (0,04)	0,29 (0,03)	0,31 (0,06)	0,57 (0,05)
Macro F1 (std)	Modèle 1	0,51 (0,05)	0,47 (0,02)	0,33 (0,03)	0,21 (0,05)	0,61 (0,05)
	Modèle 2	0,44 (0,04)	0,43 (0,04)	0,23 (0,04)	0,19 (0,06)	0,54 (0,05)
F1-A1 (std)	Modèle 1	0,71 (0,07)	0,69 (0,05)	0,58 (0,06)	0,48 (0,16)	0,8 (0,05)
	Modèle 2	0,61 (0,09)	0,64 (0,09)	0,25 (0,08)	0,5 (0,08)	0,78 (0,05)
F1-A2 (std)	Modèle 1	0,52 (0,06)	0,5 (0,04)	0,4 (0,07)	0,3 (0,09)	0,63 (0,11)
	Modèle 2	0,47 (0,06)	0,48 (0,06)	0,41 (0,05)	0,23 (0,09)	0,59 (0,09)
F1-B1 (std)	Modèle 1	0,34 (0,08)	0,24 (0,09)	0,12 (0,08)	0,13 (0,12)	0,49 (0,12)
	Modèle 2	0,26 (0,09)	0,2 (0,08)	0,03 (0,06)	0,06 (0,09)	0,47 (0,1)
F1-B2 (std)	Modèle 1	0,4 (0,11)	0,36 (0,11)	0,12 (0,06)	0,09 (0,1)	0,52 (0,09)
	Modèle 2	0,38 (0,07)	0,4 (0,07)	0,21 (0,09)	0,1 (0,08)	0,43 (0,12)
F1-C (std)	Modèle 1	0,59 (0,06)	0,58 (0,04)	0,45 (0,07)	0,02 (0,05)	0,6 (0,11)
	Modèle 2	0,46 (0,08)	0,46 (0,11)	0,24 (0,15)	0,02 (0,04)	0,44 (0,16)

Ensuite, les résultats pour les deux modèles concernant le dialogue sont présentés dans le tableau 50.

Tableau 50 : Comparaison des résultats sur les données de style dialogue

		SVM-LF	SVM-L	SVM-F	wav2vec	CamemBERT
Exactitude (std)	Modèle 1	0,58 (0,06)	0,52 (0,05)	0,42 (0,06)	0,29 (0,04)	0,6 (0,05)
	Modèle 2	0,51 (0,05)	0,49 (0,04)	0,28 (0,06)	0,31 (0,06)	0,6 (0,06)
Macro F1 (std)	Modèle 1	0,56 (0,06)	0,5 (0,06)	0,4 (0,06)	0,18 (0,07)	0,52 (0,08)
	Modèle 2	0,47 (0,04)	0,46 (0,04)	0,2 (0,06)	0,17 (0,05)	0,52 (0,05)
F1-A1 (std)	Modèle 1	0,75 (0,06)	0,7 (0,08)	0,65 (0,08)	0,43 (0,08)	0,83 (0,04)
	Modèle 2	0,7 (0,12)	0,67 (0,1)	0,33 (0,15)	0,45 (0,13)	0,83 (0,08)
F1-A2 (std)	Modèle 1	0,58 (0,08)	0,51 (0,1)	0,36 (0,13)	0,22 (0,12)	0,6 (0,08)
	Modèle 2	0,52 (0,07)	0,47 (0,07)	0,37 (0,09)	0,13 (0,13)	0,63 (0,06)
F1-B1 (std)	Modèle 1	0,38 (0,1)	0,25 (0,1)	0,29 (0,09)	0,1 (0,11)	0,41 (0,2)
	Modèle 2	0,28 (0,09)	0,29 (0,15)	0,07 (0,07)	0,07 (0,1)	0,4 (0,14)
F1-B2 (std)	Modèle 1	0,5 (0,12)	0,41 (0,15)	0,35 (0,1)	0,1 (0,16)	0,43 (0,17)
	Modèle 2	0,4 (0,1)	0,35 (0,17)	0,05 (0,08)	0,2 (0,13)	0,44 (0,14)
F1-C (std)	Modèle 1	0,61 (0,1)	0,65 (0,09)	0,35 (0,15)	0,03 (0,06)	0,34 (0,19)
	Modèle 2	0,46 (0,07)	0,5 (0,09)	0,17 (0,13)	0,02 (0,06)	0,33 (0,24)

Enfin, les résultats des deux modèles pour le monologue sont montrés dans le tableau 51.

Tableau 51 : Comparaison des résultats sur les données de style monologue

		SVM-LF	SVM-L	SVM-F	wav2vec	CamemBERT
Exactitude (std)	Modèle 1	0,49 (0,09)	0,49 (0,08)	0,39 (0,03)	0,26 (0,06)	0,58 (0,05)
	Modèle 2	0,45 (0,07)	0,45 (0,05)	0,2 (0,05)	0,24 (0,07)	0,55 (0,08)
Macro F1 (std)	Modèle 1	0,47 (0,09)	0,47 (0,08)	0,3 (0,03)	0,17 (0,05)	0,57 (0,05)
	Modèle 2	0,45 (0,07)	0,46 (0,05)	0,12 (0,04)	0,16 (0,06)	0,55 (0,08)
F1-A1 (std)	Modèle 1	0,65 (0,17)	0,56 (0,16)	0,49 (0,15)	0,41 (0,18)	0,64 (0,14)
	Modèle 2	0,58 (0,13)	0,62 (0,13)	0,01 (0,04)	0,44 (0,13)	0,73 (0,12)
F1-A2 (std)	Modèle 1	0,49 (0,11)	0,51 (0,11)	0,39 (0,08)	0,31 (0,1)	0,57 (0,05)
	Modèle 2	0,41 (0,09)	0,42 (0,08)	0,28 (0,05)	0,25 (0,11)	0,5 (0,15)
F1-B1 (std)	Modèle 1	0,32 (0,16)	0,27 (0,15)	0,1 (0,1)	0 (0)	0,4 (0,09)
	Modèle 2	0,32 (0,17)	0,31 (0,19)	0 (0)	0 (0)	0,38 (0,17)
F1-B2 (std)	Modèle 1	0,3 (0,19)	0,4 (0,19)	0 (0)	0,04 (0,09)	0,54 (0,14)
	Modèle 2	0,44 (0,14)	0,43 (0,14)	0,17 (0,16)	0,07 (0,12)	0,53 (0,12)
F1-C (std)	Modèle 1	0,6 (0,14)	0,59 (0,12)	0,55 (0,07)	0,08 (0,12)	0,68 (0,11)
	Modèle 2	0,51 (0,1)	0,5 (0,09)	0,16 (0,14)	0,03 (0,06)	0,59 (0,17)

La première différence entre les deux modèles réside dans leur performance. En raison des différences dans les ensembles de données utilisés afin d'entraîner et

évaluer les deux modèles, il était attendu qu'ils ne produisent pas exactement les mêmes résultats. Néanmoins, il apparaît clairement que le second modèle offre une performance prédictive inférieure à celle du premier.

En ce qui concerne le modèle SVM, il paraissait probable que la performance de prédiction du second modèle, utilisant des données préparées de manière automatique, soit inférieure à celle du premier modèle, utilisant des données qui ont été vérifiées et corrigées manuellement. En effet, les données du premier modèle devraient inévitablement être plus cohérentes avec l'audio, la transcription et l'alignement de l'audio. En revanche, avec le second modèle, dans la mesure où il est automatique, leur cohérence est inévitablement inférieure à celle du premier modèle. Par conséquent, dans l'ensemble, la performance de prédiction s'avère sûrement plus faible pour le second modèle que pour le premier. En considérant uniquement la macro F1, la différence entre les modèles SVM-L du premier et du second modèle ne s'élève qu'à 0,01 ou 0,04 pour tous les styles de parole. En revanche, la différence avec le modèle SVM-F est comprise entre 0,1 et 0,2, la différence étant plus importante dans le premier et le second modèle que dans le modèle SVM-L.

Par ailleurs, les résultats du modèle SVM-L, dans lequel les variables de lisibilité ont été extraites à l'aide de FABRA sur la base des transcriptions, puis examinées à partir des variables de lisibilité corrélées avec le niveau du CECR, avec une légère différence entre le premier et le second modèle, suggèrent que la génération automatique de transcriptions a été réalisée avec un degré élevé de précision. Il convient de noter que, malgré les différences dans le nombre d'ensembles de données utilisés, il n'apparaît pas de divergence importante dans les résultats.

En revanche, avec le modèle SVM-F, dans lequel les variables de fluidité ont été extraites des données de la transcription en utilisant SPPAS, afin d'aligner la parole et en se basant sur les variables de fluidité corrélées avec le niveau du CECR, la difficulté d'effectuer l'alignement automatiquement peut être perçue. Si l'alignement dans le second modèle s'avère aussi efficace que dans le premier, la différence entre les résultats des deux modèles ne devrait être que faible, comme pour les résultats du modèle SVM-L. Cependant, la différence plus importante entre les résultats du modèle SVM-F et ceux du modèle SVM-L nous oblige à conclure que la précision de l'alignement paraît inférieure à celle de la génération automatique de transcriptions, même en tenant compte des différences entre les

ensembles de données. Cependant, la différence ne s'avère pas non plus étonnamment élevée. Ainsi, il ne fait aucun doute que l'alignement automatique par SPPAS se révèle assez performant dans certaine mesure.

Ensuite, en regardant les résultats du modèle wav2vec, la simple différence de macro F1 entre les deux modèles varie de 0,01 à 0,02 pour chaque style de parole. En outre, puisque les modèles sont entraînés et évalués uniquement à partir de fichiers audios, cette différence s'avère simplement due à la différence de distribution des données.

Enfin, nous nous concentrons sur les résultats du modèle CamemBERT. Comme mentionné concernant les résultats de SVM-L, qui ont été également réalisés en se basant uniquement sur la transcription, il n'apparaît pas de grande différence entre les deux modèles du CamemBERT. Ce résultat garantit la précision de la transcription automatique.

Sur la base de ces résultats, nous obtenons des réponses aux questions de recherche à propos de l'analyse comparative du premier et du second modèle. Concernant la première question, *« Existe-t-il une différence de performance entre le premier et le second modèle ? »*, nous pouvons affirmer qu'il existe une différence de performance entre les deux modèles. En effet, la performance de prédiction du second modèle se révèle moins efficace que celle du premier, et ce pour deux raisons :

1. La différence dans le nombre d'ensembles de données utilisés dans le premier et le second modèle.
2. La différence dans la précision des données elles-mêmes, entre les données préparées avec des corrections manuelles et les données préparées de manière automatique. En particulier, la difficulté consiste à effectuer l'alignement automatiquement.

Par ailleurs, ces deux raisons se trouvent liées à la deuxième question, *« Quels sont les inconvénients du traitement automatique des données ? »*. Lors de la prédiction de la facilité d'écoute de manière entièrement automatique, la précision des données à préparer automatiquement, et en particulier la précision de l'alignement automatique, peut être réduite, ce qui peut entraîner la prédiction moins performante. Ces résultats de l'analyse comparative des deux modèles examinés selon différentes méthodes tirent des conclusions utiles pour la

construction de modèles automatiques de la facilité d'écoute du français plus précis.

7. Conclusion

Sur la base des résultats des analyses précédentes, le dernier chapitre nous permet de tirer des conclusions pour cette thèse, de présenter des perspectives et d'ajouter des suggestions pour l'enseignement du français. Les conclusions seront présentées en répondant aux quatre questions de recherche énumérées à la section 3.1.

Concernant la première question, *« Quelles sont les variables qui sont associées à la facilité d'écoute pour le français langue étrangère ? »*, la réponse est donnée par les résultats de l'analyse de corrélation, réalisée de manière à construire le modèle SVM, entre les variables de lisibilité et de fluidité, d'une part, et les niveaux du CECR, d'autre part. Les ensembles de données utilisés dans le premier et le second modèle s'avérant différents, des analyses de corrélation ont été effectuées sur chaque ensemble. Les résultats se sont révélés quelque peu différents. En particulier, la forte corrélation entre les variables de fluidité et le niveau du CECR, observée avec le premier modèle, n'a pas été trouvée avec le second. En effet, cela peut être associé à la difficulté de l'alignement automatique décrite dans le chapitre précédent. En ce qui concerne l'analyse de corrélation, les autres résultats se sont avérés similaires pour les deux modèles. Par conséquent, les résultats de l'analyse de corrélation du premier modèle, basés sur les données avec correction manuelle des alignements, sont utilisés ici afin de dériver les réponses à la question de recherche. En outre, ces résultats montrent que, pour tous les styles de parole, le niveau du CECR s'avère fortement corrélé avec les variables liées à la longueur de phrase, au niveau de difficulté du lexique et au débit de parole. En revanche, le coefficient de corrélation pour les variables de fluidité s'avère plus élevé pour le dialogue que pour le monologue. En outre, les variables liées aux temps verbaux se sont également révélées plus fortement liées pour le dialogue.

Quant à la deuxième question, *« Les caractéristiques acoustiques contribuent-elles à l'amélioration des performances au niveau de la prédiction de la facilité d'écoute ? »*, les performances de prédiction, comparées par style de parole, des modèles SVM-F et wav2vec se trouvent légèrement plus élevées pour le dialogue que pour le monologue. En relation avec les résultats de l'analyse de corrélation mentionnée précédemment, les caractéristiques acoustiques s'avèrent liées à la facilité d'écoute, en particulier concernant le dialogue. Cependant, les modèles

SVM-F et wav2vec ont obtenu des résultats modestes, ce qui indique que les variables de fluidité et les caractéristiques acoustiques ne contribuent pas de manière importante à l'amélioration des performances de prédiction. Il convient de noter que cela ne signifie pas que les variables de fluidité ne contribuent pas du tout à la prédiction de la facilité d'écoute, en particulier dans le premier modèle, puisque le modèle SVM-LF, qui inclut également des variables de fluidité, se révèle légèrement plus performant que le modèle SVM-L, basé sur les seules variables de lisibilité.

Sur la base des conclusions discutées jusqu'à présent, concernant la troisième question, *« Les performances au niveau de la prédiction de la facilité d'écoute varient-elles en fonction du style de parole ? »*, il paraît clair qu'il existe des différences entre le dialogue et le monologue quant à l'examen des modèles de facilité d'écoute. Cela peut être expliqué par les différences de caractéristiques entre ces deux styles. Comme le montrent les tableaux 7, 8, 32 et 33, une différence majeure est que le dialogue contient plus de mots et a une durée d'enregistrement plus longue que le monologue. De plus, pour le dialogue et le monologue, en plus du nombre de personnes parlant, les temps verbaux utilisés diffèrent, comme l'indique l'analyse corrélationnelle entre les variables. Ainsi, en raison de nombreuses différences de caractéristiques, analyser ces styles de parole séparément permet d'obtenir des performances de prédiction adaptées à chaque style de parole.

Par ailleurs, la définition du dialogue dans cette étude indique que le nombre de locuteurs doit être de deux ou plus. Or, nous n'avons pas examiné si l'augmentation du nombre de locuteurs, deux, trois locuteurs ou plus, affecte la facilité d'écoute, ni si le nombre et la durée des pauses varient. Par conséquent, il serait pertinent d'analyser davantage les différences pour style de parole et de les subdiviser dans une analyse plus détaillée, ce qui rendrait plus facile le traitement de la prédiction de discours authentiques.

Enfin, en ce qui concerne la quatrième question, *« Est-il possible d'automatiser la prédiction de la facilité d'écoute ? »*, nous concluons qu'il est possible, même si des progrès restent à faire. Dans cette thèse, les modèles ont été construits et analysés comparativement à l'aide de trois méthodes, SVM, wav2vec et CamemBERT. Nous pouvons affirmer que la méthode utilisant CamemBERT paraît la plus efficace, car elle offre de meilleures performances en matière de prédiction. Dans le futur, il sera également possible de combiner CamemBERT

avec l'analyse de l'audio, par exemple en construisant un nouveau modèle croisé avec le modèle wav2vec et en étudiant ses performances. Cela permettrait de faire de nouvelles propositions concernant la prédiction automatique de la facilité d'écoute.

Bien que les variables de fluidité aient été utilisées afin de construire le modèle SVM dans cette étude, l'audio français comprend de nombreuses autres caractéristiques, par exemple, l'articulation, l'omission de sons, les liaisons, la distinction entre les pauses remplies et les pauses silencieuses, voire le nombre de personnes parlant, etc. Il a été constaté que le SVM-LF, qui inclut les variables de fluidité, n'est pas particulièrement inférieur aux performances du modèle CamemBERT. Par conséquent, lors de l'examen du modèle SVM afin d'améliorer ses performances, il serait nécessaire de créer et de prendre en compte d'autres variables liées à la parole en français, comme celles mentionnées ci-dessus, en plus des variables de fluidité. En conséquence, il s'avère possible que les performances de prédiction du modèle SVM, incluant ces variables, dépassent celles de CamemBERT, qui a pourtant obtenu les meilleures performances cette fois-ci. En outre, l'audio devrait être analysée plus finement afin d'explorer ce qui est impliqué dans la détermination et la prédiction des niveaux du CECR. Même si ces analyses ne sont pas immédiatement réalisables de manière automatique, elles peuvent être effectuées, dans un premier temps, à l'aide de méthodes manuelles. Ainsi, nous sommes convaincus que les résultats se révèleront utiles lors de la création éventuelle d'un modèle entièrement automatisé de la facilité d'écoute du français.

De plus, bien que nous en soyons venus à utiliser des ensembles de données différents pour le premier et le second modèle, nous avons également besoin de créer à nouveau le premier modèle avec le second ensemble de données et de comparer les deux. Il serait également utile d'envisager d'autres méthodes concernant l'automatisation du processus de prédiction. Ensuite, en ce qui concerne l'alignement de la parole, l'utilisation de SPPAS a été envisagée dans cette thèse. Comme il existe d'autres supports qui peuvent être utilisés afin d'aligner la parole, tels que EasyAlign, il serait également utile d'en explorer les apports en vue d'obtenir de meilleurs modèles.

En plus des défis cités précédemment, il s'avère également nécessaire de prendre en compte la facilité d'écoute de chaque manuel. En effet, certains manuels peuvent avoir tendance à être prédits à des niveaux plus élevés du CECR,

ou inversement à des niveaux plus bas. Il serait également intéressant d'examiner les modèles de la facilité d'écoute pour chaque manuel, ou pour chaque éditeur, en tenant compte des caractéristiques du manuel.

Cependant, dans le cadre du développement futur de cette recherche, il s'avère nécessaire de noter les caractéristiques des données utilisées comme base pour le modèle de la facilité d'écoute dans cette thèse. Cette recherche est basée sur les audios et les transcriptions accompagnant les manuels, et nous pouvons dire que l'audio des manuels constitue fondamentalement un texte précédé d'un script. Dès lors, la question de savoir s'il s'avère possible de prédire un discours totalement authentique sur la base d'un modèle présentant de telles caractéristiques devrait être étudiée plus avant à l'avenir, par exemple en faisant écouter aux apprenants le discours prédit par ce modèle et en l'évaluant. En outre, les données de cette thèse étaient constituées de discours non accompagnés de vidéo. Or, les discours avec vidéo et transcriptions peuvent être étudiés en utilisant des méthodes similaires à la présente étude, afin d'examiner le rôle de la vidéo pour la compréhension orale du français. Étant donné que dans la vie réelle, les informations visuelles sont utilisées afin d'obtenir des informations, une telle étude permettrait d'examiner la facilité d'écoute du français dans des situations plus réalistes.

Enfin, nous présentons les implications pédagogiques concernant l'enseignement du français, et en particulier pour l'enseignement de la compréhension orale. Bien que l'objectif de cette étude ait été de construire un modèle de la facilité d'écoute du français, nous envisageons d'affiner le mécanisme et de l'intégrer dans une interface accessible à tous. Si cela devient possible, il sera plus facile pour les enseignants de sélectionner le matériel pédagogique, pour les apprenants de promouvoir l'auto-apprentissage et d'être plus motivés pour apprendre. Même au stade actuel, dans lequel cet objectif n'a pas été atteint, les résultats de cette thèse permettent de saisir des éléments qui peuvent être utiles pour l'enseignement de la compréhension orale en français. En particulier, les résultats de l'analyse de corrélation entre le niveau du CECR et les variables de lisibilité et de fluidité menée s'avèrent utiles pour l'enseignement. Tout d'abord, il faut considérer l'existence de différences dans les variables corrélées avec le niveau du CECR pour le dialogue et le monologue. Dès lors, lorsque les apprenants écoute un audio, l'instruction peut être donnée avec différents points d'attention en fonction du style de parole. Par exemple, pour le

dialogue, les variables liées au temps verbal sont liées à la distinction du niveau du CECR. Par conséquent, l'intégration de questions concernant les temps verbaux peut améliorer la compréhension orale du dialogue par les apprenants. En outre, le niveau de vocabulaire s'avère fortement corrélé avec le niveau du CECR pour le dialogue et le monologue. Ainsi, les résultats indiquent qu'il faut réfléchir à enrichir le vocabulaire ; par exemple, en faisant apprendre à l'avance aux étudiants le vocabulaire important du contenu qu'ils doivent écouter, de manière à améliorer leur compréhension orale de l'audio. De cette manière, l'enseignement devrait d'abord être basé sur des audios spécifiques sélectionnés par l'enseignant, tels que les fichiers audios accompagnant le manuel, afin que les étudiants puissent acquérir progressivement des compétences de base en matière de la compréhension orale. Des documents authentiques devraient ensuite être incorporés dans le processus d'apprentissage. Cela préparera les apprenants au moment où ils auront besoin de communiquer avec des francophones et d'obtenir des informations en français à l'oral. À cette fin, il semble urgent de développer des modèles prédictifs de la facilité d'écoute.

Bibliographie

Akita, T. (2009). "Kiku", "kiku", "kiku" : mittsu no "kiku chikara" wo hagukumu torikumi (Entendre, écouter et demander : initiatives visant à développer les trois compétences d'écoute) (in Japanese). *Obirin today: In search of a learner-centered education*, 9, pp.97-112.

Alghamdi, E. A., Gruba, P., & Velloso, E. (2022). The Relative Contribution of Language Complexity to Second Language Video Lectures Difficulty Assessment. *The Modern Language Journal*, 106 (2), pp.393-410.

Allen, W. (1952). Readability of instructional film commentary. *Journal of Applied Psychology*, 36 (3), pp.164-168.

Anderson, J. R. (1980). *Cognitive psychology and its implications*. San Francisco: W.H. Freeman.

Anderson, R. C. (1984). Role of Reader's Schema in Comprehension, Learning and Memory. In Anderson, R. C., Osborn, J., & Tierney, R. J. (eds.), *Learning to Read in American Schools: Basal readers and content texts*, London: Routledge, pp.243-272.

Anderson, A., & Lynch, T. (1988) *Listening*. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. New York: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation*, 8, pp.47-89.

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, 4 (11), pp.417-423.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of communication disorders*, 36 (3), pp.189-208.

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, pp.12449-12460.

Bigi, B. (2012). SPPAS: a tool for the phonetic segmentations of Speech. *The eighth international conference on Language Resources and Evaluation*, pp.1748-1755.

Bigi, B., & Meunier, C. (2018). Automatic segmentation of spontaneous speech. *Revista de Estudos da Linguagem*, 26 (4), pp.1489-1530.

Blau, E. K. (1990). The effect of syntax, speed, and pauses on listening comprehension. *TESOL quarterly*, 24 (4), pp.746-753.

Byrnes, H. (1984). The role of listening comprehension: A theoretical base. *Foreign Language Annals*, 17 (4), pp.317-29.

Call, M. E. (1985). Auditory short-term memory, listening comprehension, and the input hypothesis. *Tesol Quarterly*, 19 (4), pp.765-781.

Candea, M. (2000). *Contribution à l'étude des pauses silencieuses et des phénomènes dits" d'hésitation" en français oral spontané. Étude sur un corpus de récits en classe de français*. Thèse de doctorat, Université de la Sorbonne nouvelle-Paris III.

Carroll, J. B. (1971). Learning from Verbal Discourse in Educational Media: A Review of the Literature. *The Educational Testing Service*, Princeton, N.J.

Cartier, F. A. (1952). The Social Context of Listenability Research. *Journal of Communication*, 2 (1), pp.44-47.

Cartier, F. A. (1955). II. Listenability and "human interest". *Communications Monographs*, 22 (1), pp.53-57.

Chall, J. S., & Dial, H. E. (1948). Predicting listener understanding and interest in newscasts. *Educational Research Bulletin*, 27 (6), pp.141-168.

Charrow, V. R., & Charrow, R. P. (1979). Characteristics of the language of jury instructions. In Alatis, J., & Tucker, G. R. (eds.), *Language in public life*, pp.163-185, Washington D.C.: Georgetown University Press.

Conseil de l'Europe. (2001). *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*. Paris: Hatier.

Cook, V. (1977). Cognitive processes in second language learning. *International Review of Applied Linguistics*, XV (1), pp.1-20.

Cook, V. (1996). *Second language learning and language teaching 2nd edition*. New York: Arnold.

Cornaire, C. (1998). *La compréhension orale*. Paris: CLE international.

Crook, C., & Schofield, L. (2017). The video lecture. *The Internet and Higher Education*, 34, pp.56-64.

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), pp.2862-2873.

Cutler, A., & Clifton, C. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown., & P. Hagoort. (eds.), *The neurocognition of language*, pp.123-166, Oxford: Oxford University Press.

Dale, E. (1941). A comparison of two word lists. *Educational Research Bulletin*, 10, pp.484-489.

Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational research bulletin*, 27 (1), pp.37-54.

Denbow, C. J. (1975). Listenability and readability: An experimental investigation. *Journalism Quarterly*, 52 (2), pp.285-290.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Ducoffe, M., Mayaffre, D., Precioso, F., Lavigne, F., Vanni, L., & Tre-Hardy, A. (2016). Machine Learning under the light of Phraseology expertise: use case of presidential speeches, De Gaulle-Hollande (1958-2016). *JADT 2016 – Statistical Analysis of Textual Data*, 1, pp.157-168.

Eastwood, J., & Snook, B. (2012). The effect of listenability factors on the comprehension of police cautions. *Law and human behavior*, 36 (3), pp.177-183.

Ericsson, K.A., & Simon, H. (1984). *Protocol Analysis: Verbal Reports as Data*. Cambridge: MIT Press.

Ericsson, K.A., & Simon, H. (1987). Verbal reports on thinking. In Færch, C., & Kasper, G. (eds.), *Introspection in Second Language Research*. Clevedon: Multilingual Matters, pp.24-53.

Fang, I. E. (1967). The "Easy listening formula. *Journal of Broadcasting & Electronic Media*, 11 (1), pp.63-68.

Feyten, C. M. (1991). The power of listening ability: An overlooked dimension in language acquisition. *The modern language journal*, 75 (2), pp.173-180.

Field, J. (1998). Skills and strategies: Towards a new methodology for listening. *ELT Journal*, Volume 52/2, pp.110-118.

Field, J. (1999). Bottom-up and top-down. *ELT Journal*, 53 (4), pp.338-339.

Field, J. (2002). The changing face of listening. In J. C. Richards., & W. A. Renandya. (eds.), *Methodology in language teaching: An anthology of current practice*, pp.242-247, Cambridge: Cambridge University Press.

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

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32 (3), pp.221-233.

Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice*. Cambridge: Cambridge University Press.

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., & Miltsakaki, E. (2012). Do NLP and machine learning improve traditional readability formulas?. *Proceedings of the First Workshop on Predicting and Improving Text Readability for target reader populations*, pp.49-57.

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

Fukuda, M. (2002). A review of the studies on listening comprehension in the second language (in Japanese). *Bulletin of the Graduate School of Education*, 51, pp.367-374.

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, pp.379-399.

Glasser, T. L. (1975). On readability and listenability. *ETC: A Review of General Semantics*, 32 (2), pp.138-142.

Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28 (1), pp.55-75.

Gray, W. S., & Leary, B. E. (1935). *What makes a book readable*. University of Chicago Press, Chicago.

Gremmo, M. J., & Holec, H. (1990). La compréhension orale : un processus et un comportement. In Gaonac'h, D., Mac Nally, D., & Ballaire, M-F. (eds.), *Acquisition et utilisation d'une langue étrangère : l'approche cognitive*, pp.30-40, Paris: Hachette.

Griffiths, R. (1992). Speech rate and listening comprehension: Further evidence of the relationship. *TESOL quarterly*, 26 (2), pp.385-390.

Harwood, K. A. (1950). A concept of listenability. *Western Speech*, 14 (2), pp.10-12.

Harwood, K., & Cartier, F. (1952). On definition of listenability. *Southern Journal of Communication*, 18 (1), pp.20-23.

Harwood, K. A. (1955a). I. Listenability and readability. *Communications Monographs*, 22 (1), pp.49-53.

Harwood, K. A. (1955b). III. Listenability and rate of presentation. *Communications Monographs*, 22 (1), pp.57-59.

Ikeda, H. (2003). A Study of Listening Strategy Instruction on Japanese EFL Learners (in Japanese). *Kyoto Sosei University Review*, 3, pp.71-78.

Ivan, M. (2006). La méthode structuro-globale audio-visuelle (SGAV). *Dialogos*, 7 (14), pp.16-21.

Kiany, G. R., & Shiramiry, E. (2002). The Effect of Frequent Dictation on the Listening Comprehension Ability of Elementary EFL Learners. *TESL Canada Journal*, 20 (1), pp.57-63.

Kikuchi, K., & Nakayama, K. (2006). Listening to English Movies: Effects on Junior High School Students' Intrinsic Interest in Learning English (in Japanese). *Japanese journal of educational psychology*, 54 (2), pp.254-264.

Kita, Y. (2005). A study on Japanese Students' English Listening Comprehension Abilities (in Japanese). *Journal of tourism studies*, 4, pp.113-126.

Kiyokawa, H. (1976). A New List of Basic Spoken English Words (in Japanese). *Senshu Language Laboratory bulletin*, 5, pp.28-38.

Kiyokawa, H. (1977). A Review of the Research Literature on Listenability Studies (in Japanese). *Senshu Language Laboratory bulletin*, 6, pp.49-57.

Kiyokawa, H. (1985). A Revised List of Basic Spoken English Words (in Japanese). *The journal of Wayo Women's University*, pp.209-224.

Kiyokawa, H. (1987). A List of Basic Spoken English Words: Final Report (in Japanese). *Wayo Women's University Language and Literature*, 21, pp.43-62.

Kiyokawa, H. (1989). Study on Listenability (I) (in Japanese). *Wayo Women's University Language and Literature*, 23, pp.45-57.

Kiyokawa, H. (1990). A Formula for Predicting Listenability: The Listenability of English Language Materials 2. *Wayo Women's University Language and Literature*, 24, pp.57-74.

Kobayashi, T. (2008). Eigo listening ni okeru gakushusha ga ryûi subeki otohenka to "ruiongo" no kokufuku ni muketa shidô (Changements sonores dont les apprenants doivent être conscients lors de l'écoute de l'anglais et stratégies pédagogiques pour surmonter les « mots homophones ») (in Japanese). *Language Studies*, 16, pp.3-34.

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*, pp.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), pp.7-11.

Kotani, K., & Yoshimi, T. (2019). Listenability Measurement Based on Learners' Transcription. *The Journal of Information and Systems in Education*, 18 (1), pp.1-6.

Larsen-Freeman, D. (2000). *Techniques and Principles in Language Teaching Second edition*. Oxford: Oxford University Press.

Lennon, P. (1990). Investigating fluency in EFL: a quantitative approach. *Language Learning*, 40 (3), pp.387-417.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.

Lorge, I. (1944). Predicting readability. *Teachers college record*, 45 (6), pp.404-419.

Loukina, A., Yoon, S. Y., Sakano, J., Wei, Y., & Sheehan, K. (2016). Textual complexity as a predictor of difficulty of listening items in language proficiency tests. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp.3245-3253.

Martin, L., Muller, B., Suárez, P. J. O., Dupont, Y., Romary, L., de la Clergerie, É., Seddah, D., & Sagot, B. (2020). CamemBERT : a tasty French language model. arXiv preprint arXiv:1911.03894.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63 (2), pp.81-97.

Morley, J. (1990). Trends and developments in listening comprehension: Theory and practice. *Georgetown University round table on languages and linguistics*, 1220, pp.317-337.

Morriss, E. C., & Holversen, D. (1938). *Idea Analysis Technique*. Unpublished Ms., Teachers College, Columbia University, New York.

Nakamura, K., & Hasegawa, T. (1995). *Furansugo wo donoyoni oshieruka (Comment enseigner le français ?)* (in Japanese). Tokyo: Surugadai Shuppansha.

Newmark, G., & Diller, E. (1964). Emphasizing the Audio in the Audio-Lingual Approach. *The Modern Language Journal*, 48 (1), pp.18-20.

Noble, W. S. (2006). What is a support vector machine?. *Nature biotechnology*, 24 (12), pp.1565-1567.

Nunan, D. (1989). *Designing Tasks for the Communicative Classroom*. Cambridge: Cambridge University Press.

Ogawa, M. (2018). The Benefits of Dictation and Shadowing for Teaching Listening (in Japanese), *Senshu journal of foreign language education*, 46, pp.77-91.

Oishi, H., & Kinoshita, T. (2008). Daiichigengo shori to dainigengo shori ni okeru nôkasseijôtai no chigai – nihongo to eigo no listening ni oite – (Différences dans les états d'activation du cerveau dans le traitement de la première et de la deuxième langue – en écoutant du japonais et de l'anglais –) (in Japanese). *Studia linguistica*, 21, pp.143-154.

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

O'ki, T., & Izumi, Y. (2015). Shadowing vs. Accelerated Speech Dictation: Which Improves Learner's Decoding Skill? (in Japanese). *Hakuoh journal of the faculty of education*, 9 (1), pp.227-243.

O'ki, T., Maeda, H., & Oka, H. (2016). What Makes Listening to English Difficult? (in Japanese). *Hakuoh journal of the faculty of education*, 10 (2), pp.511-530.

O'Malley, J. M., Chamot, A. U., & Küpper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied linguistics*, 10 (4), pp.418-437.

Onaha, H. (2014). How Human Memory and Working Memory Work in Second Language Acquisition (in Japanese). *Ryudai Review of Euro-American Studies*, 58, pp.1-25.

Ozawa, M., Wilkens, R. S., Sugiyama, K., & François, T. (2024). Modéliser la facilité d'écoute en FLE : vaut-il mieux lire la transcription ou écouter le signal vocal ?. *Actes de la 31ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN2024), volume 1 : articles longs et prises de position*, pp.549-566.

Peltonen, P. (2017). Temporal fluency and problem-solving in interaction: An exploratory study of fluency resources in L2 dialogue. *System*, 70 (C), pp.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*, pp.134-154.

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

Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, 17 (2), pp.219-39.

Richards, J. C., & Rodgers, T. S. (1986). *Approaches and Methods in Language Teaching*. Cambridge: Cambridge University Press.

Richards, J. C. (1993). Beyond the Text Book: the Role of Commercial Materials in Language Teaching. *RELC Journal*, 24 (1), pp.1-14.

Rogers, J. R. (1962). A formula for predicting the comprehension level of material to be presented orally. *The journal of educational research*, 56 (4), pp.218-220.

Rost, M. (1991). *Listening in action*. New Jersey: Prentice Hall.

Rost, M. (2002). *Teaching and researching listening*. Harlow: Pearson Education.

Rubin, J. (1994). A review of second language listening comprehension research. *The modern language journal*, 78 (2), pp.199-221.

Rubin, D. L. (2012). Listenability as a tool for advancing health literacy. *Journal of health communication*, 17 (sup3), pp.176-190.

Ruggia, S. (2019). Le deep learning : un outil pour la didactique du FLE ?. *Dialettica pedagogica*, 1, pp.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*, pp.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), pp.103-109.

Satori, M. (2021). Effects of working memory on L2 linguistic knowledge and L2 listening comprehension. *Applied Psycholinguistics*, 42 (5), pp.1313-1340.

Sauzedde, B. (2016). Étude comparative des pauses silencieuses en français lu par des natifs et par des apprenants japonais. *Ritsumeikan studies in language and culture*, 27 (2/3), pp.295-308.

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), pp.90-116.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49, pp.93-120.

Skehan, P. (2003). Task-based instruction. *Language teaching*, 36 (1), pp.1-14.

Snook, B., Luther, K., Eastwood, J., Collins, R., & Evans, S. (2016). Advancing legal literacy: The effect of listenability on the comprehension of interrogation rights. *Legal and Criminological Psychology*, 21 (1), pp.174-188.

Stempleski, S. (1987). Short Takes: Using Authentic Video in the English Class. *Paper presented at the 21st Annual Meeting of the International Association of Teachers of English as a Foreign Language*. Westende, Belgium: IATEFL.

Sugai, K. (2020). How Native Japanese Learners Listen to Spoken English: Applying Research on the Bottom-up Process to Education (in Japanese). *LET Kyushu-Okinawa BULLETIN*, 20, pp.1-10.

Sugiura, M., Takeuchi, S., & Baba, K. (2002). Autonomous learning for the development of listening skills: The effects of dictation (in Japanese). *Studies in Language and Culture*, 23 (2), pp.105-121.

Tavakoli, P., & Skehan, P. (2005). Strategic planning task structure, and performance testing. In R. Ellis (ed.), *Planning and task performance in a second language*, Amsterdam: John Benjamins.

Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics and Language Teaching*, 54 (2), pp.133-150.

Thorndike, E. L. (1932). *A teacher's word book of the twenty thousand words*. Bureau of publications, Teachers college, Columbia University, New York.

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, pp.410-413.

Ueda, S., Yoshimi, T., Nanjo, H., & Kotani, K. (2014). A Listenability Formula Using Listening Proficiency Level Information of Japanese Learners of English as a Foreign Language (in Japanese). *Transactions of Japanese Society for Information and Systems in Education*, 31 (2), pp.203-207.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40, pp.191-210.

Voss, B. (1979). Hesitation phenomena as sources of perceptual errors for non-native speakers. *Language and Speech*, 22 (2), pp.129-144.

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*, pp.1217-1233.

Xu, B., Tao, C., Feng, Z., Raqui, Y., & Ranwez, S. (2021). A benchmarking on cloud based speech-to-text services for french speech and background noise effect. arXiv preprint arXiv:2105.03409.

Yin, S. (2002). A review of teaching second/foreign language listening comprehension research (in Japanese). *Japanese Language Education*, pp.279-288.

Yonezu, A., & Ogura, M. (2017). A Language Activity in a Senior High School Focused on Listening Activity – Theory and Practice – (in Japanese). *The journal of University Educational Center*, 5, pp.13-22.

Yoon, S. Y., Cho, Y., & Napolitano, D. (2016). Spoken text difficulty estimation using linguistic features. *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, pp.267-276.

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, pp.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), pp.44-49.