



République Algérienne Démocratique et Populaire  
Ministère de l'enseignement supérieur et de la recherche scientifique  
Université des Sciences et de la Technologie Houari Boumediene



**Faculté d’Informatique**  
**Département Systèmes Informatiques**  
**Mémoire de Licence**

Filière : Informatique  
Spécialité : Ingénierie des Systèmes d’Information et des Logiciels

---

**Thème**  
Recommandation d’activités pédagogiques pour l’apprentissage de la langue anglaise

---

Proposé et encadré par :

**Dr. BERKANI Lamia**

**Prof. GUESSOUM Ahmed**

Présenté par :

**M. ARRIS Yanis**

**Mlle. MEKKI Sarah Maria**

Soutenu le : ../06/2024

Devant le jury composé de :

**Dr. SEDDIKI Manel (Présidente)**

**Dr. ABERBOUR Rima (Membre)**

**Binôme N°: 043/2024**

# **Remerciements**

Nous tenons à exprimer notre gratitude profonde et remerciements à nos encadreurs: Dr. Lamia BERKANI et Pr. Ahmed GUESSOUM pour leur disponibilité, leurs orientations précieuses et leur contribution inestimable sans laquelle ce travail n'aurait pas pu être accompli. Nous leur sommes également reconnaissants pour leur confiance et leur patience à notre égard.

Nous souhaitons remercier tout particulièrement Dr. Mohamed HADJ AMEUR, membre de l'équipe TALAA, Laboratoire LRIA de l'USTHB et enseignant à l'ENSIA, pour son aide et ses conseils précieux.

Nous adressons également nos remerciements les plus sincères aux membres du jury, Dr.Seddiki Manel et Dr.Aberbour Rima pour l'attention qu'elles ont accordée à notre travail en le passant en revue et en l'améliorant grâce à leurs conseils.

Nous souhaitons également remercier toutes les personnes qui ont contribué, directement ou indirectement, au bon déroulement de ce projet. Leur soutien, leurs conseils et leur encouragement ont été d'une aide précieuse tout au long de cette aventure.

Enfin, nos remerciements vont à nos familles et amis pour leur soutien moral et leur compréhension pendant cette période intense de travail.

# Résumé

Ce mémoire de licence s'intéresse à la conception et développement d'une application web dédiée à l'apprentissage de l'anglais. Reconnaissant l'importance croissante de cette langue dans notre pays et vu la décision de son adoption dans nos écoles et universités, nous avons développé une plateforme nommée "TaalimNet", permettant un apprentissage adaptatif et personnalisé des apprenants de différents niveaux.

Les contributions majeures de ce travail concernent principalement l'élaboration d'un dictionnaire par niveaux de difficultés comprenant 144 997 mots et le développement de la plateforme d'apprentissage. Le dictionnaire a été construit sur la base de mots extraits de Princeton WordNet et classés automatiquement selon la classification adoptée par le Cadre Européen Commun de Référence pour les langues connu sous le nom de CEFR (Common European Framework of Reference). La classification a été effectuée en utilisant le modèle BERT et un corpus de mots collectés de ressources linguistiques crédibles (Cambridge, Oxford), initialement classés par niveaux de difficultés. La deuxième contribution concerne la conception et développement d'une plateforme proposant diverses activités pédagogiques personnalisées selon le profil de l'apprenant. Cette plateforme vise à développer leur niveau de maîtrise de cette langue sur tous les aspects (prononciation, écriture, écoute et lecture), à travers une variété d'exercices (tests de niveaux et d'apprentissage). Différentes ressources et outils ont été exploités pour la génération d'exercices dont le dictionnaire par niveaux et le LLM Gemini de Google. Ces activités sont adaptées au contexte Algérien, puisqu'elles proposent des traductions en arabe et en français pour faciliter la compréhension des apprenants.

**Mots clés :** Dictionnaire par niveaux de difficulté, WordNet, LLM, BERT, Activités pédagogiques.

## ملخص

يتناول هذا البحث الجامعي تصميم وتطوير تطبيق ويُبْ مخصص لتعلم اللغة الإنجليزية. ومع الاعتراف بالأهمية المترادفة لهذه اللغة في بلادنا ونظرًا لقرار اعتمادها في مدارسنا وجامعتنا، قمنا بتطوير منصة تسمى "TaalimNet"، تُسَمِّح بتعلم متكيٍّ وشخصيٍّ للمتعلمين من مختلف المستويات.

تتمثل المساهمات الرئيسية لهذا العمل في إنشاء قاموس مصنف حسب مستويات الصعوبة ويشمل 144997 كلمة وتطوير منصة التعلم. تم بناء هذا القاموس بناءً على كلمات مستخرجة من Princeton WordNet وتم تصنيفها تلقائيًا وفقًا للتصنيف المعتمد من قبل الإطار الأوروبي المرجعي المشترك للغات المعروف باسم CEFR . تم التصنيف باستخدام نموذج BERT ومجموعة كلمات مُجمعة من موارد لغوية موثوقة (كامبريدج، أكسفورد)، تم تصنيفها مبدئيًّا حسب مستويات الصعوبة.

أما المساهمة الثانية، فتتعلق بتصميم وتطوير منصة تقدم أنشطة تعليمية متنوعة مخصصة وفقًا لملف التعلم الخاص بكل متعلم. تهدف هذه المنصة إلى تطوير مستوى إتقانهم لهذه اللغة في جميع الجوانب (النطق، الكتابة، الاستماع والقراءة)، من خلال مجموعة متنوعة من التمارين (اختبارات مستوى وتعلم). تم استغلال موارد وأدوات مختلفة لتوليد التمارين بما في ذلك القاموس المصنف ونموذج LLM Gemini من Google. تم تكييف هذه الأنشطة مع السياق الجزائري، حيث تُقدم ترجمات إلى العربية والفرنسية لتسهيل فهم المتعلمين.

**الكلمات المفتاحية :** قاموس حسب مستويات الصعوبة، WordNet، LLM، BERT، أنشطة تعليمية.

# Abstract

This bachelor's thesis focuses on the design and development of a web application dedicated to learning English. Recognizing the growing importance of this language in our country and considering its adoption in our schools and universities, we have developed a platform named "TaalimNet" that enables adaptive and personalized learning for students of different levels.

The major contributions of this work primarily involve the creation of a level-based dictionary comprising 144,997 words And the development of the learning platform. This dictionary was built using words extracted from Princeton WordNet and automatically classified according to the classification adopted by the Common European Framework of Reference for Languages, known as CEFR. The classification was performed using the BERT model and a corpus of words collected from credible linguistic resources (Cambridge, Oxford), initially classified by difficulty levels. The second contribution concerns the design and development of a platform offering various personalized educational activities according to the learner's profile. This platform aims to enhance their proficiency in all aspects of the language (pronunciation, writing, listening, and reading) through a variety of exercises (level and learning tests). Different resources and tools were utilized for generating exercises, including the level-based dictionary and Google's LLM Gemini. These activities are tailored to the Algerian context, as they provide translations in Arabic and French to facilitate learners' understanding.

**Keywords:** Level-based dictionary, WordNet, LLM, BERT, Educational activities.

# Table des matières

<b>Table des figures</b>	<b>7</b>
<b>Liste des tableaux</b>	<b>8</b>
<b>Introduction générale</b>	<b>9</b>
<b>1 État de L'art</b>	<b>10</b>
1.1 Introduction . . . . .	10
1.2 Spécificité de la Langue Anglaise . . . . .	10
1.2.1 Généralité . . . . .	10
1.2.2 Niveaux de difficulté . . . . .	11
1.2.3 Test de niveau . . . . .	12
1.3 Plateformes d'apprentissage en ligne . . . . .	13
1.3.1 Description de Duolingo . . . . .	13
1.3.2 Description de Babbel . . . . .	15
1.4 Dictionnaire par niveaux . . . . .	15
1.4.1 Dictionnaire vs Dictionnaire par niveaux . . . . .	15
1.4.2 Travaux liés . . . . .	16
1.5 Large Language Models . . . . .	16
1.5.1 Le Traitement Automatique du Langage Naturel . . . . .	16
1.5.2 Large Language Models . . . . .	17
1.5.3 Exemples de LLMs . . . . .	17
1.5.4 Fine-tuning des LLMs . . . . .	17
1.5.5 Prompt Engineering . . . . .	18
1.6 Conclusion . . . . .	18
<b>2 Conception</b>	<b>19</b>
2.1 Introduction . . . . .	19
2.2 Architecture Générale du système . . . . .	19
2.3 Construction du dictionnaire par niveaux de difficultés . . . . .	21
2.3.1 Sélection du corpus . . . . .	21
2.3.2 Organisation . . . . .	21
2.3.3 Prétraitement et analyse du corpus . . . . .	22
2.3.4 Structure du dictionnaire par niveaux . . . . .	22
2.4 Collecte des données . . . . .	24
2.5 Proposition d'activités d'apprentissage . . . . .	24
2.5.1 Test de niveau . . . . .	24
2.5.2 Activités d'apprentissage . . . . .	25
2.5.3 Recommandation d'exercices . . . . .	26

2.5.4	Consultation du dictionnaire . . . . .	27
2.6	Partie ingénierie du logiciel de notre plateforme . . . . .	27
2.6.1	Modélisation des fonctionnalités . . . . .	27
2.6.2	Diagramme de classes . . . . .	28
2.6.3	Modèle relationnel des Données . . . . .	28
2.7	Conclusion . . . . .	29
<b>3</b>	<b>Implémentation</b>	<b>30</b>
3.1	Introduction . . . . .	30
3.2	Environnement de développement . . . . .	30
3.2.1	Environnement matériel . . . . .	30
3.2.2	Architecture technique . . . . .	30
3.2.3	Langages de programmation, outils et formats de données . . . . .	32
3.3	Création du dictionnaire par niveaux . . . . .	32
3.3.1	Partitionnement en niveaux . . . . .	32
3.3.2	Analyse de corpus . . . . .	33
3.3.3	Collecte de données . . . . .	33
3.4	Description de la plateforme d'apprentissage "TaalimNet" . . . . .	35
3.4.1	Analyse des couleur de l'interface . . . . .	35
3.4.2	Présentation de quelques fonctionnalités . . . . .	36
3.4.3	Présentation de quelques activités d'apprentissage . . . . .	39
3.5	Conclusion . . . . .	40
<b>Conclusion générale</b>		<b>41</b>
<b>Bibliographie</b>		<b>42</b>
<b>Annexe</b>		<b>45</b>
A.	Activité d'apprentissage (suite) . . . . .	45
B.	Business Model Canvas . . . . .	50

# Table des figures

1.1	Deux interfaces de la plateforme Duolingo . . . . .	13
1.2	Quelques interfaces de la plateforme Babbel . . . . .	15
2.1	Architecture générale de notre système . . . . .	20
2.2	Analyse Du Corpus . . . . .	22
2.3	Modélisation du dictionnaire par niveaux . . . . .	23
2.4	Modélisation des fonctionnalités . . . . .	27
2.5	Diagramme De Classes . . . . .	28
3.1	Architecture Technique. . . . .	31
3.2	Page d'authentification et d'inscription . . . . .	36
3.3	Page d'accueil d'un apprenant . . . . .	36
3.4	Page de profil de l'apprenant . . . . .	37
3.5	Page du Dictionnaire de l'apprenant . . . . .	37
3.6	Page du Dictionnaire d'un enseignant. . . . .	38
3.7	Page importer un dictionnaire d'un enseignant. . . . .	38
3.8	Exemple d'activité du Pendu . . . . .	39
3.9	Exemple d'activité de Correspondance Texte-Texte . . . . .	39
3.10	Exemple d'activité de prononciation . . . . .	40
3.11	Exemple d'activité de Coupler les mots . . . . .	40
3.12	Exemple d'activité de Correspondance Image Texte . . . . .	45
3.13	Exemple d'activité de Mot en carte . . . . .	45
3.14	Exemple d'activité de Lire un texte et répondre aux questions . . . . .	46
3.15	Exemple d'activité de Phrases Verbes . . . . .	46
3.16	Exemple d'activité de Choisir la bonne traduction . . . . .	47
3.17	Mots de la même famille . . . . .	47
3.18	Exemple d'activité d'écrire la phrase que vous entendez . . . . .	47
3.19	Exemple d'activité de Entendre un texte et répondre aux questions . . . . .	48
3.20	Exemple d'activité d'expression écrite . . . . .	48
3.21	Exemple d'activité de Conjugaison au présent simple . . . . .	49
3.22	Exemple d'activité de conjugaison au passé simple . . . . .	49
3.23	Business Model Canvas . . . . .	50

# Liste des tableaux

1.1	Epreuves De L'examen TOEFL . . . . .	12
2.1	Types d'exercices du test de notre système . . . . .	24
3.1	Caractéristiques des postes de travail. . . . .	30
3.2	Distribution des types de mots par niveau CECR . . . . .	34
3.3	Nombre de mots par niveau CECR . . . . .	34
3.4	Nombre de Texte par niveau CECR . . . . .	34
3.5	Nombre de mots par niveau CECR . . . . .	35
3.6	Nombre de mots par niveau CECR et catégories grammaticales . . . . .	35

# Introduction générale

L'apprentissage de la langue anglaise occupe une place centrale dans le contexte mondial actuel, vu l'importance de cette langue devenue la première langue au niveau international. Par conséquent, la maîtrise de l'anglais est devenue un atout indispensable dans divers domaines, qu'il s'agisse du monde professionnel, académique ou même personnel. Face à cette réalité, de nombreux outils et plateformes d'apprentissage de l'anglais ont émergé, offrant une variété d'approches et de méthodes pour accompagner les apprenants dans leur parcours linguistique.

Cependant, malgré cette abondance d'outils, il demeure essentiel de développer des solutions adaptées aux besoins spécifiques des apprenants de notre pays, en tenant compte de leur niveau de compétence linguistique, de leurs préférences d'apprentissage et de leur environnement culturel. C'est dans cette optique que s'inscrit le présent projet de fin d'études de licence, qui se propose d'explorer la recommandation d'activités pédagogiques pour l'apprentissage de la langue anglaise dans un contexte Algérien, en mettant l'accent sur la conception et le développement d'un outil innovant et personnalisé.

Le sujet abordé dans ce mémoire se situe à la croisée de plusieurs domaines d'étude, allant de la linguistique et de la pédagogie à l'informatique et à l'intelligence artificielle. En effet, la conception d'un tel outil requiert une compréhension approfondie des mécanismes d'apprentissage linguistique, ainsi que des compétences techniques en développement logiciel et en traitement automatique du langage naturel (TALN).

Après cette introduction, ce mémoire sera structuré en trois chapitres, décrits comme suit :

1. **État de l'art** : présente les outils et plateformes existants pour l'apprentissage de la langue anglaise, en mettant en lumière leurs forces. Nous explorerons également les concepts clés tels que les niveaux de difficulté des mots, les tests de niveaux et les grandes tendances en matière d'apprentissage des langues.
2. **Conception** : propose notre démarche de conception pour le développement d'une plateforme d'apprentissage d'anglais, en mettant l'accent sur la création d'un dictionnaire par niveau de difficultés qui sera exploité comme ressource essentielle à la recommandation personnalisée d'activités pédagogiques.
3. **Implémentation** : décrit les étapes d'implémentation de notre plateforme, en mettant en œuvre les concepts et les méthodes définis dans la phase de conception.

Une conclusion avec quelques perspectives futures sera présentée à la fin de ce mémoire.

# Chapitre 1

## État de L'art

### 1.1 Introduction

Ce chapitre s'intéresse à la présentation de notions liées au développement de plates-formes d'apprentissage de la langue anglaise. De prime abord, quelques généralités seront abordées telles que les niveaux de difficulté des mots et les tests de niveaux avant de présenter quelques plates-formes d'apprentissage existantes. Ensuite, nous présenterons les dictionnaires par niveaux avec quelques travaux liés et techniques utilisées pour la construction de ces dictionnaires. Finalement nous donnerons quelques notions de base liés aux Large Language Models (LLMs) en vue de les exploiter dans la construction de nos activités d'apprentissage.

### 1.2 Spécificité de la Langue Anglaise

#### 1.2.1 Généralité

La langue anglaise présente des aspects linguistiques et culturels spécifiques, qu'il serait intéressant d'explorer en vue de développer une plateforme d'apprentissage portant sur un ensemble d'activités pédagogiques dédié à des apprenants de la langue anglaise. Voici quelques éléments clés qui caractérisent la langue anglaise :

1. **Richesse lexicale et variété des accents** :L'anglais se distingue par sa vaste gamme de vocabulaire et la diversité des accents présents dans les pays anglophones. Les outils pédagogiques doivent prendre en compte cette diversité pour offrir une expérience d'apprentissage adaptée à différents contextes et accents régionaux.
2. **Complexité grammaticale** :Bien que la grammaire anglaise soit relativement simple dans sa structure de base, elle comporte des nuances et des règles spécifiques pouvant être délicates pour les apprenants. Les exercices proposés doivent aborder progressivement ces aspects, en fonction du niveau de compétence des apprenants.
3. **Phrase idiomatique et expressions courantes** :L'anglais est réputé pour ses expressions idiomatiques et ses phrases courantes qui ne se traduisent pas toujours littéralement. Les outils pédagogiques doivent inclure des activités spécifiques pour familiariser les apprenants de cette langue avec ces expressions et les aider à les utiliser de manière appropriée dans différents contextes.
4. **Culture anglophone** :Une compréhension approfondie de la culture anglophone est essentielle pour maîtriser la langue. Les activités pédagogiques devraient donc

intégrer des éléments culturels tels que les coutumes, la littérature, la musique et les traditions, offrant ainsi une immersion culturelle aux apprenants.

5. **Adaptabilité aux niveaux de compétence** : L'enseignement de l'anglais se fait souvent selon des niveaux de compétence allant de A1 à C2 (A1, A2, B1, B2, C1, C2) [9] ; A1 étant le niveau débutant, et C2 niveau avancé. Les outils pédagogiques doivent être en mesure de s'adapter à ces niveaux en proposant des activités appropriées, tout en prenant en compte les progrès individuels des apprenants.
6. **Importance mondiale de l'anglais** : En raison de sa prédominance mondiale, l'anglais est devenu un moyen de communication universel et une compétence essentielle dans un monde de plus en plus interconnecté. L'anglais est devenu une langue indispensable à l'échelle mondiale, vu sa position centrale dans les domaines professionnel, académique et diplomatique. Parlée comme langue officielle dans plus de 400 millions de personnes et comme langue étrangère par 750 millions de personnes, l'anglais joue un rôle crucial dans les échanges internationaux et le succès professionnel.

### 1.2.2 Niveaux de difficulté

Les niveaux de difficulté pour l'apprentissage de la langue anglaise, en utilisant le Cadre européen commun de référence pour les langues CEFR aussi connu sous le nom de CEFR (Common European Framework of Reference) :

- **A1 – Débutant** : Capable de réaliser des transactions simples sur des sujets et dans des situations familiaires.
- **A2 – Élémentaire** : Dispose d'un vocabulaire suffisant pour des interactions quotidiennes et pour exprimer des besoins de base en communication.
- **B1 – Intermédiaire** : Possède un vocabulaire permettant de s'exprimer sur la plupart des sujets de la vie quotidienne, bien que des périphrases puissent être nécessaires.
- **B2 – Intermédiaire supérieur** : Dispose d'un large éventail de mots pour des sujets généraux et spécifiques, malgré quelques lacunes lexicales pouvant entraîner des hésitations.
- **C1 – Avancé** : Maîtrise un vaste répertoire lexical, permettant de pallier facilement les manques par des périphrases, avec une bonne compréhension des expressions idiomatiques et du langage familier.
- **C2 – Maîtrise** : Possède une maîtrise exceptionnelle d'un lexique étendu, incluant les expressions idiomatiques et le langage familier, avec une sensibilité aux nuances de sens.

Le Cadre européen commun de référence pour les langues (CEFR) est un système international de notation utilisé pour évaluer la maîtrise d'une langue étrangère. Il offre une échelle claire et détaillée pour décrire les compétences linguistiques des apprenants, en se basant sur des critères objectifs et des descripteurs de compétences pour chaque niveau. Comparé à d'autres systèmes de classification, le CEFR est préféré pour sa reconnaissance mondiale, sa structure détaillée et sa compatibilité avec les pratiques d'enseignement et d'évaluation des langues.

### 1.2.3 Test de niveau

Différents tests de niveaux peuvent être distingués ; nous les exposons comme suit :

#### 1. Test of English as a Foreign Language (TOEFL) :

Le TOEFL est l'un des tests de langue anglaise les plus reconnus au monde. Il évalue les compétences en lecture, en écriture, en compréhension orale et en expression orale de l'anglais. Le TOEFL iBT (Internet-Based Test) est le format le plus courant, bien qu'il existe également des versions papier et informatisées. Il est souvent utilisé pour l'admission dans les universités anglophones, pour l'obtention de bourses d'études et pour l'immigration dans certains pays. Le tableau 1.1 répertorie les différentes épreuves de l'examen TOEFL, accompagnées de leur catégorie respective[4].

Catégorie d'exercices	Exercices
Reading (60-80mn)	Lire 3-4 passages de texte (600-700 mots) puis répondre un nombre de questions de 12-14 pour chaque passage.
Listening (60-90mn)	Écouter une conversation entre 2 personnes (2-3mn) accompagné de 5 questions pour chaque conversation. Ensuite, écouter un enseignant parler et répondre à six questions.
Speaking (20mn)	Entreprendre une conversation avec une personne en temps réel.
Writing (50mn)	Lire un passage académique, ensuite écouter une conférence sur le même sujet. Pour enfin rédiger ensuite un essai les résumant. Lire une consigne (un sujet) et répondre par écrit.

TABLE 1.1 – Epreuves De L'examen TOEFL

#### 2. International English Language Testing System (IELTS) :

L'IELTS est un autre test largement reconnu qui évalue les compétences en anglais. Il existe deux versions de l'IELTS : l'IELTS Academic, utilisé pour l'admission à l'université, et l'IELTS General Training, utilisé à des fins d'immigration et professionnelles. L'IELTS évalue les compétences en compréhension écrite, expression écrite, compréhension orale et expression orale[10].

#### 3. Cambridge English Exams (FCE, CAE, CPE) :

Ces examens sont administrés par l'Université de Cambridge et évaluent différents niveaux de compétence de l'anglais. Le First Certificate in English (FCE) est destiné aux apprenants de niveau intermédiaire supérieur, le Certificate in Advanced English (CAE) est destiné aux apprenants de niveau avancé et le Certificate of Proficiency in English (CPE) est destiné aux apprenants de niveau maîtrise de la langue. Les examens Cambridge évaluent les compétences en compréhension écrite et orale, expression écrite et expression orale[11].

#### 4. Test of English for International Communication (TOEIC) :

Le TOEIC est un test axé sur les compétences en anglais dans un contexte professionnel et commercial. Il évalue les compétences en compréhension écrite et orale à

travers des situations professionnelles courantes. Le TOEIC est souvent utilisé par les entreprises et les organisations pour évaluer le niveau d'anglais de leurs employés ou de leurs candidats[12].

## 1.3 Plateformes d'apprentissage en ligne

Nous présentons ci-dessous deux plateformes d'apprentissage de langue qui sont probablement les plus connues et utilisées : Duolingo et Babbel.

### 1.3.1 Description de Duolingo

- Duolingo : est une plateforme d'apprentissage en ligne des langues .Elle offre une variété de cours dans différentes langues, y compris l'anglais. Duolingo présente une gamme d'activités variées pour renforcer les compétences linguistiques, avec des cours adaptés au niveau individuel de chaque apprenant. Un système de points est mis en place pour garantir un meilleur suivi de la progression de l'apprentissage et maintenir la motivation des apprenants. La Figure 1.1 présente quelques interfaces de Duolingo.

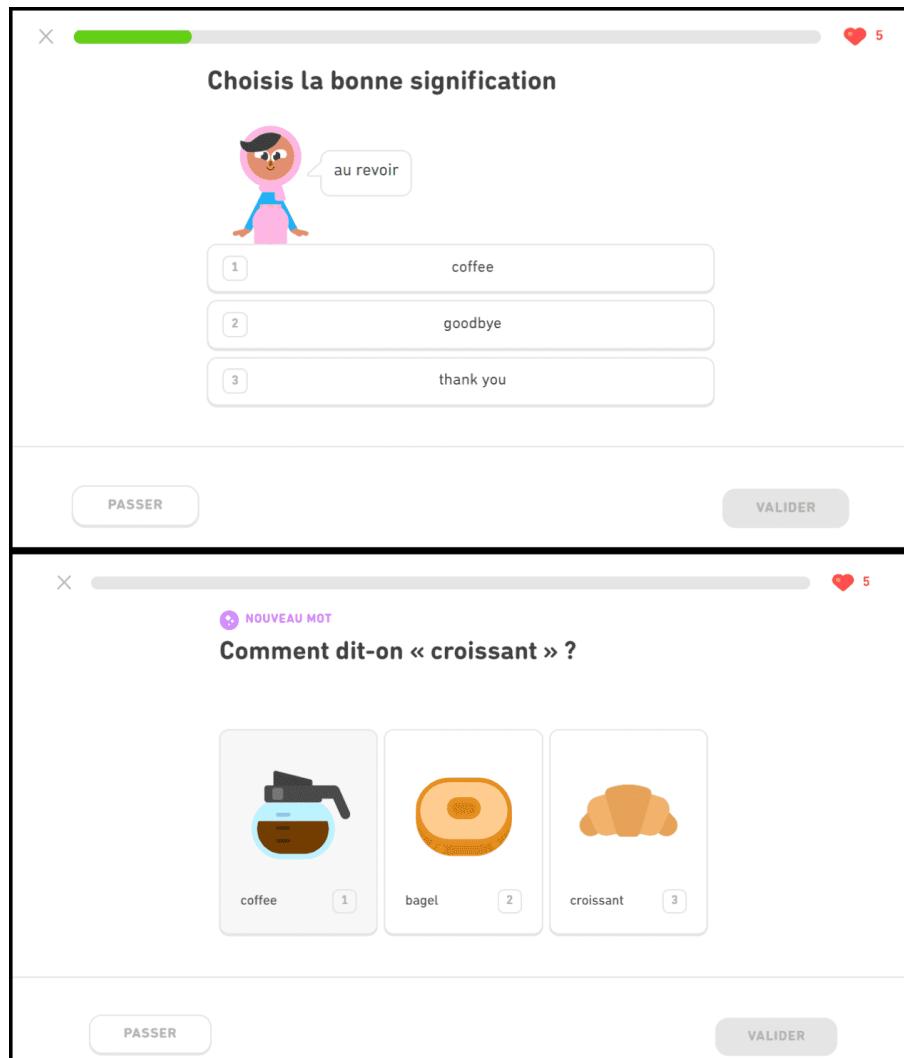


FIGURE 1.1 – Deux interfaces de la plateforme Duolingo

Le système existant de Duolingo utilise une combinaison d'intelligence artificielle, notamment le modèle de langage large (LLM) et un modèle appelé "Birdbrain"[3], pour accélérer la création de cours et personnaliser l'expérience d'apprentissage. Le LLM prédit la manière la plus probable de compléter une phrase, ce qui permet de générer rapidement du contenu pour les exercices. Le système Birdbrain ajuste le niveau de difficulté des exercices en fonction des forces et des faiblesses de chaque utilisateur. Ces outils permettent à Duolingo de produire du contenu de qualité plus rapidement, d'explorer des concepts linguistiques avancés et de mieux répondre aux besoins de ses millions d'utilisateurs quotidiens. En utilisant l'IA pour la création de cours, Duolingo peut également allouer ses ressources à des cours plus spécialisés, contribuant ainsi à améliorer l'efficacité et la productivité de l'ensemble de la plateforme.

Les activités principales proposées par Duolingo sont décrites comme suit[5] :

- Organisez les mots : Une phrase mélangée dans la langue cible de l'apprenant lui est présentée ,et il/elle lui est demandé de réorganiser les mots pour former une phrase grammaticalement correcte.
- Complétez la traduction : Une phrase est présentée et sa traduction manque d'un mot que l'apprenant doit taper.
- Indiquez la signification correcte : L'apprenant est demandé de choisir laquelle parmi plusieurs phrases dans une langue est une traduction valide d'une phrase dans l'autre langue. Souvent, les alternatives incorrectes ne sont même pas grammaticalement correctes. Voir les révélations flagrantes.
- Association de cartes flash d'images : L'apprenant se voit présenter plusieurs mots et images correspondantes dans une langue, et il lui est demandé de choisir celle qui correspond à un mot dans l'autre langue.
- Traduction de carte flash d'image (étiquetée : Lequel de ceux-ci est le 'mot'?) : L'apprenant est présenté avec un mot dans une langue et on lui demande de le traduire dans l'autre langue avec l'aide de quelques images pertinentes.
- Sélectionnez le mot manquant : L'apprenant est présenté avec une phrase comportant un espace vide et est demandé de choisir le mot ou la phrase correcte pour le compléter.
- Reconstituez la phrase (étiquetée : Écrivez ceci dans la langue) : L'apprenant est présenté avec le texte d'une phrase dans une langue et une phrase mélangée dans l'autre. Il/elle devra réorganiser les mots pour former la traduction correcte.
- Prononcez cette phrase : L'apprenant est présenté avec le texte d'une phrase et est demandé de la prononcer dans votre microphone.
- Associez les paires : Un nombre pair de boîtes, moitié dans la langue de base de l'apprenant et moitié dans sa langue cible, sont affichées. Il/elle devra sélectionner les paires. L'apprenant peut sélectionner une boîte dans l'une ou l'autre langue, puis sa traduction.
- Traduction : L'apprenant est présenté avec le texte d'une phrase dans une langue et est demandé de le traduire dans l'autre.
- Écrivez ce que vous entendez : L'apprenant est présenté avec l'audio d'une phrase et est demandé de la transcrire. Un bouton portant une tortue répète lentement la phrase.
- Qu'est ce que vous entendez ? : L'apprenant est présenté avec l'audio d'un mot ou d'une phrase et plusieurs transcriptions en alternatives.

### 1.3.2 Description de Babbel

-Babbel : est une plateforme d'apprentissage en ligne des langues proposant des cours dans diverses langues, avec un accent mis sur la pratique de la conversation et de la grammaire. Les leçons sont adaptées au niveau de l'utilisateur et utilisent des exercices interactifs pour renforcer les compétences linguistiques de l'apprenant. Un système de suivi des progrès permet aux utilisateurs de suivre leur évolution et de rester motivés. La Figure 1.2 présente quelques interfaces de la plateforme Babbel.

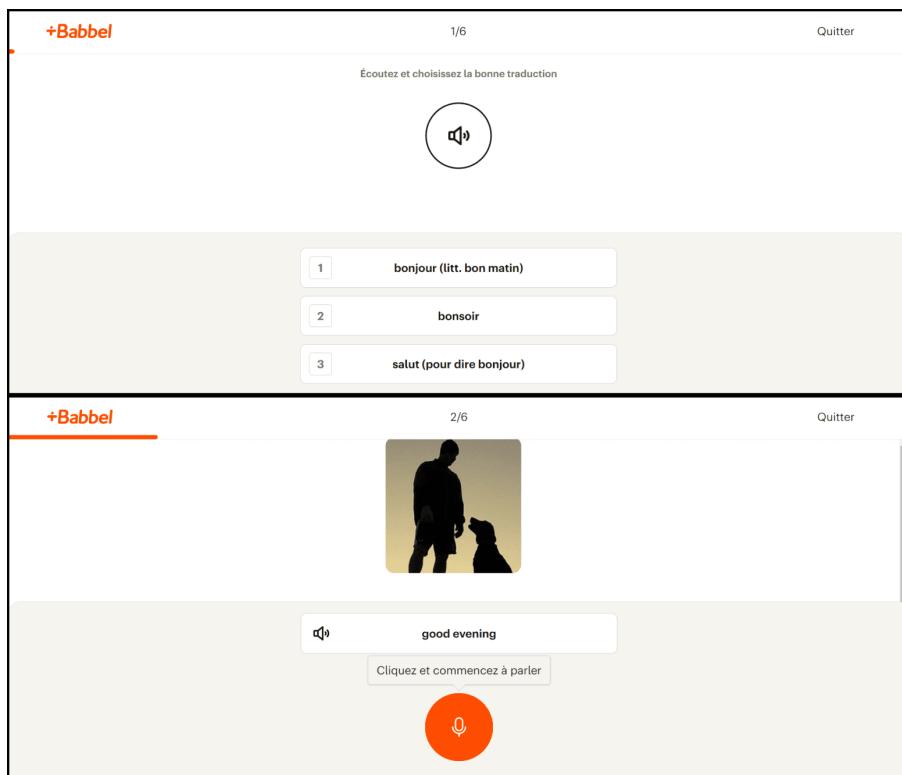


FIGURE 1.2 – Quelques interfaces de la plateforme Babbel

Après avoir examiné de près les plateformes les plus populaires d'apprentissage des langues existantes, nous avons relevé plusieurs points forts, dont les suivants :

- La diversité des exercices proposés, permettant une pratique variée des compétences linguistiques.
- L'adaptabilité des cours en fonction du niveau de l'apprenant, favorisant ainsi une progression personnalisée.
- La mise en place de mécanismes de motivation tels que des récompenses et des défis, encourageant ainsi la régularité et l'engagement dans l'apprentissage de la langue.

## 1.4 Dictionnaire par niveaux

### 1.4.1 Dictionnaire vs Dictionnaire par niveaux

Un dictionnaire représente un précieux ouvrage de référence qui répertorie et explique les mots d'une langue, dans un ordre alphabétique, en fournissant diverses informations telles que la prononciation, la signification, l'origine étymologique, et parfois

des exemples d'utilisation. En outre, les dictionnaires peuvent comporter des indications grammaticales, des synonymes, des antonymes, des expressions idiomatiques et d'autres données linguistiques utiles. Ils servent d'outils incontournables pour l'apprentissage, l'enseignement, l'écriture et la compréhension de la langue. Parmi les dictionnaires les plus renommés, on reconnaît le dictionnaire anglais de Cambridge[13]. Par ailleurs, il existe des dictionnaires classés par niveaux (CECR), allant du niveau A1 (débutant) au niveau C2 (avancé), afin d'adapter les définitions et les exemples à chaque niveau de compétence linguistique. Parmi ces dictionnaires, on reconnaît l'Oxford 5000 ou 3000[14]. Contrairement aux dictionnaires traditionnels, ces ouvrages catégorisent les mots selon le niveau de maîtrise requis, facilitant ainsi l'apprentissage et le perfectionnement des compétences linguistiques pour les apprenants.

### 1.4.2 Travaux liés

Diverses recherches ont exploré plusieurs métriques et méthodes afin de déterminer la complexité des mots en anglais.

- Jagoda et Boi nski[1] : ont conçu un modèle de traitement du langage naturel (TALN) qui organise des mots provenant de la base lexicale WordNet<sup>1</sup> en cinq niveaux de difficulté, allant du niveau 1 au niveau 5. La classification des mots s'est basée sur plusieurs critères, notamment la fréquence d'occurrence des mots dans des extraits de Wikipédia, la longueur des mots et le score de lisibilité Flesch-Kincaid<sup>2</sup>.
- Curto, Mamede et Baptista[2] : ont proposé un système qui utilise le TALN pour extraire des caractéristiques linguistiques des textes en Portugais européen, puis forme un classificateur automatique de difficulté, dans le but de classifier des textes par niveaux de difficulté. Les caractéristiques extraites incluent les parties du discours, les syllabes et les mots. Le classificateur développé a atteint une précision de 75,11%. Testé avec trois niveaux de difficulté, la précision de ce classificateur était de 81,44%, montrant sa capacité à classifier avec une précision raisonnable les textes en fonction de leur niveau de difficulté.

## 1.5 Large Language Models

Le Traitement Automatique du Langage Naturel est profondément influencé par l'émergence et le développement des Large Language Models. Dans cette section, nous explorerons ces deux notions, TALN et LLM, avec quelques exemples notables de LLMs. Finalement, nous aborderons les notions de fine-tuning et de prompt engineering.

### 1.5.1 Le Traitement Automatique du Langage Naturel

Le TALN est une discipline de l'intelligence artificielle qui vise à doter les ordinateurs de la capacité de comprendre, interpréter et générer le langage humain de manière naturelle[6]. Il englobe un large éventail de techniques et d'outils conçus pour permettre

---

1. WordNet est une base de données lexicale de relations sémantiques entre les mots qui relie les mots en relations sémantiques incluant les synonymes, les hyponymes et les méronymes.

2. Le score de lisibilité Flesch-Kincaid est conçu pour évaluer la complexité de compréhension d'un passage en anglais selon une formule spécifique.

aux machines de traiter efficacement le langage écrit et parlé, ouvrant ainsi la voie à de nombreuses applications pratiques dans divers domaines.

### 1.5.2 Large Language Models

Les LLMs ont bouleversé le TALN en fournissant une compréhension inégalée du langage humain. Leur architecture basée sur les transformers leur permet d'analyser de vastes corpus textuels pour capturer les nuances linguistiques, les rendant exceptionnellement précis dans diverses tâches linguistiques. Les transformers sont une classe d'architectures de réseaux neuronaux qui ont introduit une analyse bidirectionnelle du contexte, améliorant ainsi la compréhension globale des énoncés en tenant compte du contexte environnant.[7]

Ces modèles ne se contentent pas de comprendre le langage, leur architecture flexible leur permet également de générer du texte de manière fluide et naturelle. Cela ouvre la voie à des applications telles que la génération automatique de contenu, la traduction avancée et la création de dialogues conversationnels, ouvrant ainsi de nouvelles perspectives pour l'intelligence artificielle basée sur le langage. On parle aujourd'hui d'IA générative[15].

En outre, les LLMs peuvent être “fine-tunés” (affinés) pour des tâches spécifiques[8], ce qui implique de ré-entraîner les dernières couches du modèle sur des ensembles de données annotés. Cette technique permet d'adapter le modèle à des tâches particulières tout en préservant les connaissances générales acquises lors du pré-entraînement, optimisant ainsi les performances pour des applications spécifiques dans divers domaines, tels que le domaine médical, financier, ou autres.

### 1.5.3 Exemples de LLMs

Parmi les LLMs les plus remarquables, Bidirectional Encoder Representations from Transformers (BERT)[16] se distingue particulièrement. Développé par Google AI en 2018, BERT a été une avancée majeure dans le domaine du TALN, ouvrant la voie à une nouvelle ère de compréhension du langage par les machines. Son architecture basée sur les transformers, a permis une compréhension plus approfondie du contexte dans lequel les mots apparaissent.. Contrairement aux modèles précédents, BERT prend en compte le contexte dans les deux directions, ce qui lui permet de saisir plus efficacement la signification des mots dans une phrase en tenant compte de leur environnement. Cependant, BERT n'est pas le seul acteur dans cette révolution du TALN. Aux côtés de BERT, d'autres LLMs sont disponibles dont Gemini[17], ChatGPT[18], Falcon[19], RoBERTa[20], DistilBERT[21], ALBERT[22] et ELECTRA[23]. Chaque modèle apporte ses propres améliorations spécifiques, qu'il s'agisse d'une plus grande efficacité, d'une réduction de la taille des modèles ou d'une approche différente de l'entraînement. ouvrant ainsi la voie à de nouvelles possibilités et applications dans ce domaine en constante évolution.

### 1.5.4 Fine-tuning des LLMs

Une caractéristique essentielle des LLMs est leur capacité à être fine-tunés pour des tâches spécifiques. Le fine-tuning implique de ré-entraîner les dernières couches du modèle sur des ensembles de données annotés, ce qui permet d'adapter le modèle à des tâches particulières tout en préservant les connaissances générales acquises lors du pré-entraînement. Cette technique permet d'optimiser les performances des LLMs pour des

applications spécifiques dans divers domaines.

### 1.5.5 Prompt Engineering

Le "prompting" est une méthode utilisée dans le TALN pour guider les LLMs, en leur fournissant des instructions ou des exemples de texte. Cela permet aux modèles de produire des résultats plus précis et pertinents pour diverses tâches, telles que la génération de textes, la classification de texte ou la réponse à des questions. En utilisant des prompts appropriés, les modèles peuvent être orientés pour accomplir des tâches spécifiques, ce qui améliore leurs performances et leur adaptabilité dans différentes situations.

## 1.6 Conclusion

En conclusion, ce chapitre a posé les fondements nécessaires à la compréhension de l'apprentissage de l'anglais en ligne. Nous avons exploré les niveaux de difficulté des mots, les tests de niveaux, ainsi que des plateformes existantes pour l'apprentissage des langues. De plus, nous avons abordé des concepts clés tels que les dictionnaires, notamment les dictionnaires par niveaux, le TALN, les LLMs , BERT. Ces éléments fournissent une base cruciale pour notre étude conceptuelle sur le développement d'un portail d'apprentissage d'anglais innovant ; que nous présentons dans le chapitre suivant.

# Chapitre 2

## Conception

### 2.1 Introduction

Ce chapitre présente la conception de notre plateforme d'apprentissage de la langue anglaise dédiée à des élèves des écoles algériennes. Un ensemble d'activités d'apprentissage sera décrit ainsi que notre proposition pour la création d'un dictionnaire par niveaux de difficultés qui sera exploité dans la génération de ces activités.

### 2.2 Architecture Générale du système

La figure 2.1 décrit l'architecture générale de notre système, composé de deux parties essentielles. La première partie concerne la création d'un dictionnaire par niveaux de difficultés de la langue anglaise, classant les mots selon leur niveau de difficulté, conformément aux six niveaux du CECR. La seconde partie de notre système consiste à développer une plateforme offrant une variété d'activités et de tests visant à évaluer le niveau de compétence en langue anglaise des apprenants. L'objectif est d'aider les élèves à progresser dans leur apprentissage et à améliorer leur maîtrise de la langue. En outre, les enseignants peuvent soumettre leur propre dictionnaire de niveaux via cette plateforme, présenté sous forme d'un fichier CSV incluant les mots accompagnés de leur niveau CECR et de leur classe grammaticale. Cette fonctionnalité leur permet de personnaliser le contenu pédagogique selon les besoins spécifiques de leurs élèves, tout en contribuant à enrichir constamment les ressources disponibles.

La plateforme offre également la description des mots, leurs traductions, synonymes et conjugaisons dans différents temps verbaux, fournissant ainsi un support complet pour l'enseignement et l'apprentissage de la langue. En outre, l'apprenant aura le choix de choisir une autre langue, l'arabe ou le français, comme support d'accompagnement pour une meilleure compréhension des différentes activités qui lui seront proposées en anglais.

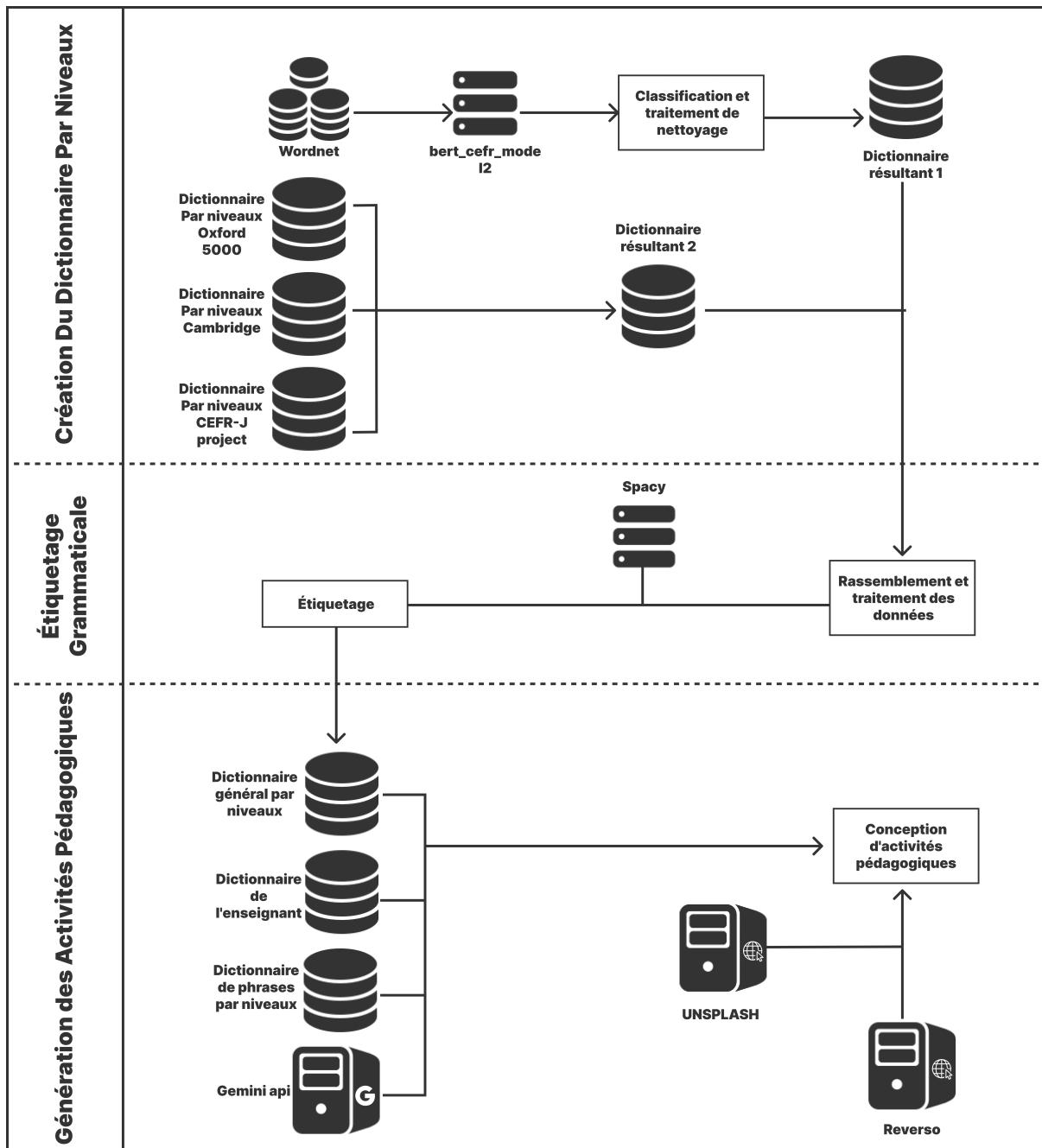


FIGURE 2.1 – Architecture générale de notre système

## 2.3 Construction du dictionnaire par niveaux de difficultés

### 2.3.1 Sélection du corpus

Pour enrichir notre corpus, nous avons sélectionné quatre ressources principales, et nous avons utilisé modèle BERT pour la labellisation des mots de WordNet :

- **Oxford 5000 [14]** : Réputé pour son exhaustivité et sa précision, l’Oxford 5000 offre une large gamme de mots couramment utilisés en anglais, ce qui en fait une ressource précieuse pour les apprenants de l’anglais comme langue étrangère de tous niveaux.
- **Cambridge Dictionary [24]** : Le Cambridge Dictionary est une référence bien établie pour les apprenants de l’anglais, offrant des définitions claires et des exemples d’utilisation pour chaque mot, ce qui en fait une source idéale pour notre corpus.
- **CEFR-j Project [25]** : Ce projet, conduit par une université japonaise, présente un dictionnaire spécifiquement élaboré en fonction des niveaux CECR. Il contient 9937 mots classifiés selon leur niveau de difficulté, ce qui le rend particulièrement approprié pour notre objectif d’ajuster les activités d’apprentissage au niveau de chaque apprenant.
- **Kaggle dataset [26]** : Ce dataset comprend plus de 1400 textes classifiés par niveaux CECR et extraits de diverses sources telles que The British Council<sup>3</sup>, ESLFast<sup>4</sup> et le dataset cnn-daily mail<sup>5</sup>, offrant ainsi une variété de contenus pour enrichir le corpus que nous voulons développer pour les besoins de notre système.

**Modèle BERT (bert\_cefr\_model2) [27]**, un modèle pré-entraîné basé sur BERT a été utilisé pour la labellisation de tous les mots de WordNet, permettant ainsi une classification précise des mots en fonction de leur niveau de difficulté. Afin d’évaluer l’exactitude de ce modèle, nous avons comparé les niveaux CECR prédits par le modèle BERT avec les niveaux réels (obtenus à partir des dictionnaires que nous avons fusionnés). Les résultats sont les suivants : une précision de 40,84 % a été obtenue pour une classification exacte du niveau, et une précision de 83,58 % en appliquant une marge de tolérance permettant une erreur d’un niveau. Par exemple, lorsque le niveau réel est A2 et que le modèle propose A1 ou B1, ceci est toléré dans la deuxième évaluation.

### 2.3.2 Organisation

Après avoir sélectionné les sources de données, nous avons procédé à l’organisation des données extraites. Nous avons utilisé pour catégoriser grammaticalement chaque mot la bibliothèque Spacy<sup>6</sup>. Ceci nous permet d’étiqueter correctement les mots en fonction de leur nature grammaticale (verbe, adjetif, etc.), Part-of-speech en Anglais.

Le processus d’extraction des données a été réalisé avec soin pour garantir la qualité et la fiabilité du corpus. Nous avons transformé les données trouvées dans chaque source en

---

3. [www.learnenglish.britishcouncil.org](http://www.learnenglish.britishcouncil.org)

4. [www.eslfast.com](http://www.eslfast.com)

5. [www.paperswithcode.com/dataset/cnn-daily-mail-1](http://www.paperswithcode.com/dataset/cnn-daily-mail-1)

6. Spacy, une bibliothèque Python gratuite et open-source, offre des fonctionnalités avancées en NLP. Elle permet la création de systèmes d’extraction d’informations, de compréhension du langage naturel, et de prétraitement de texte pour l’apprentissage en profondeur.

fichiers CSV, ce qui facilite leur manipulation ultérieure. Pour ce faire, nous avons utilisé des outils tels qu'Excel, Table Capture un outil de web scraping (BeautifulSoup)<sup>7</sup>[28] et Pandas. Les données ont été regroupées dans deux fichiers distincts : un fichier CSV contenant les mots et un autre fichier CSV contenant les textes complets.

### 2.3.3 Prétraitement et analyse du corpus

L'analyse du corpus est divisée en deux principales étapes.

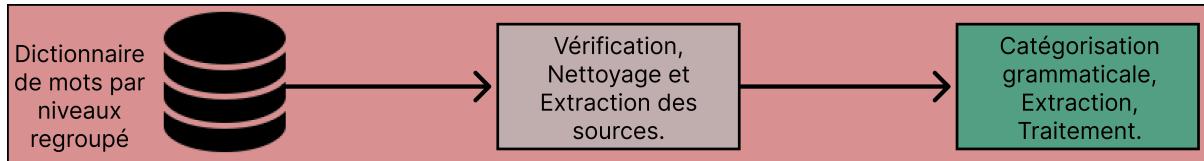


FIGURE 2.2 – Analyse Du Corpus

#### Vérification et nettoyage des mots,suppression des doublons

Après la fusion de toutes les bases de données, nous avons entrepris une opération de nettoyage visant à supprimer les doublons de mots, en ne conservant qu'une seule occurrence de chaque mot accompagnée de son niveau CECR .

#### Catégorisation grammatical , extraction , traitement

Les ensembles de données collectés étaient déjà étiquetés grammaticalement. Cependant, pour tout classifier de manière similaire, nous avons opté pour Spacy[29] , une riche bibliothèque de traitement du langage naturel en Python, connue pour sa précision dans l'étiquetage des parties du discours et dans l'analyse syntaxique des textes.

#### Résultats de l'analyse du corpus

Après avoir organisé tous les mots de WordNet par niveaux CECR et avoir attribué la classe grammaticale à chaque mot, nous les avons fusionnés avec d'autres dictionnaires collectés comme mentionné dans la section 3.1 de ce chapitre. Au total, nous avons rassemblé un ensemble de 144 997 mots classifiés par niveaux CECR.

### 2.3.4 Structure du dictionnaire par niveaux

La figure 2.3 décrit les différentes classes et relations de notre modèle dédié à la sauvegarde du dictionnaire par niveaux de difficulté.

Notre modèle se concentre principalement sur les mots. Chaque mot est associé à un niveau de difficulté CECR spécifique, ainsi qu'à une traduction en français et en arabe, accompagnée d'une brève description avec les deux langues extraites de WordNet. De plus, chaque mot est étiqueté selon sa position grammaticale, en incluant les différentes conjugaisons associées (présent et passé simple), en cas de verbe. En complément aux mots, notre dictionnaire inclut également des phrases classées par niveaux CECR.

7. Web scraping implique de récupérer et d'extraire des informations d'une page web puis les enregistrer dans une base de données ou un fichier pour une utilisation ultérieure.

Les mots du dictionnaire sont classés selon le CECR, allant du niveau A1 au niveau C2. Le niveau A1 représente les bases avec des mots simples et couramment utilisés adaptés aux débutants, tandis que le niveau C2 comprend des termes sophistiqués et spécialisés adaptés aux locuteurs très compétents. Cette hiérarchie permet une progression systématique des apprenants, commençant par les bases et avançant vers des compétences linguistiques plus avancées.

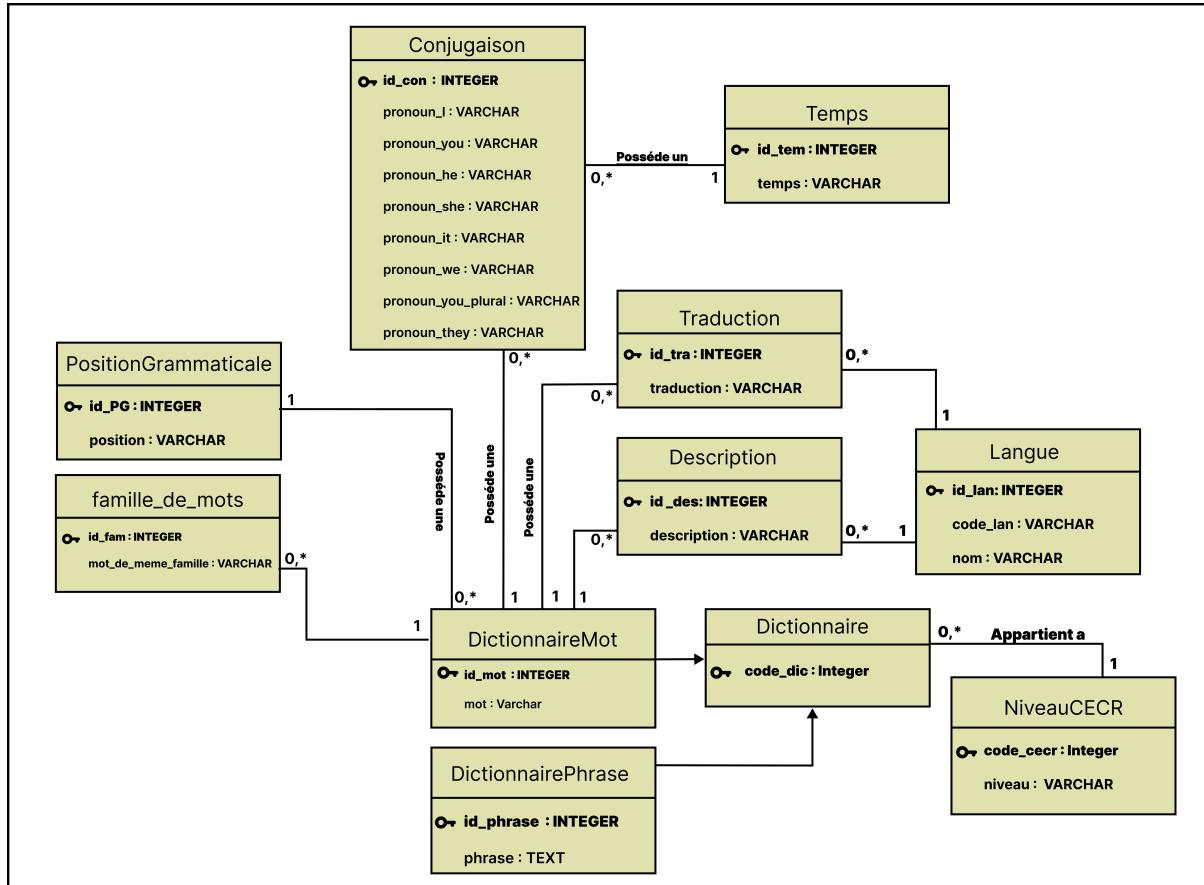


FIGURE 2.3 – Modélisation du dictionnaire par niveaux

- Chaque instance du Dictionnaire possède un code unique "code\_dic", appartient à un niveau de difficulté CECR (NiveauCECR). Un niveau CECR a plusieurs instances du dictionnaire.
- Les instances du dictionnaire peuvent être soit un mot(DictionnaireMot) soit une phrase(DictionnairePhrase).
- Chaque mot du dictionnaire a une seule position grammaticale (PositionGrammaticale). Une position grammaticale englobe plusieurs mots.
- Chaque mot du dictionnaire a une seule traduction par langue, sachant qu'un mot traduit correspond à plusieurs mots du dictionnaire.
- Chaque mot est lié à une description par langue et une description correspond à plusieurs mots du dictionnaire..
- Une langue peut être soit l'anglais, l'arabe ou le français. Une langue possède plusieurs traductions et description de mots.
- Chaque mot a un seul synonyme ou mot du même champ lexical(famille\_de\_mots).

- Les mots dont la classe grammaticale est "VERB" ne possèdent qu'une conjugaison par temps. Un temps possède plusieurs conjugaisons par verbe.

## 2.4 Collecte des données

Pour bâtir un dictionnaire complet, nous devons récolter une multitude d'informations pour chaque mot. Cela englobe la recherche de sa traduction en français et en arabe, ainsi que sa description disponible dans les deux langues. De plus, il est crucial de trouver ses synonymes ou les mots de sa même famille (champ lexical) dans le cas échéant. En ce qui concerne les verbes, nous devons obtenir leur conjugaison dans tous les temps. En combinant tous ces éléments, nous pourrons créer un dictionnaire précis qui répondra aux besoins de nos utilisateurs.

## 2.5 Proposition d'activités d'apprentissage

### 2.5.1 Test de niveau

Les évaluations de niveau disponibles seront utilisées pour catégoriser chaque utilisateur en fonction du niveau approprié selon le CECR. Ensuite, l'apprenant se lancera dans une série d'exercices où les mots seront sélectionnés à partir du niveau déterminé. En cas de réussite dans la plupart des exercices de son niveau, l'apprenant pourra alors se soumettre au test permettant de passer au niveau supérieur. En outre, l'utilisateur aura la faculté de passer le test de niveau à tout moment. Ces activités visent à tester les compétences de l'apprenant dans toutes les catégories tout en offrant une expérience ludique.

Le test de passage au niveau supérieur est noté sur 240 points, en considérant les différentes formes de questions portant sur l'écoute, la lecture, le parler et l'écrit, évaluées chacune sur 60 points. Si l'apprenant obtient un score de 90% ou plus, il passera automatiquement au niveau supérieur.

Le tableau 2.1 explique l'organisation des activités du test de niveau en fonction de la catégorie.

Catégorie d'exercices	Exercices
Listening	Écouter un texte en audio puis répondre à des questions : 60 points
Reading	Lire un texte puis répondre à des questions : 30 points Pour chaque mot, associez-le à sa traduction : 30 points
Speaking	Lire et écouter une phrase, puis la relire à haute voix : 60 points
Writing	Écrire une expression écrite : 20 points Conjuguer un verbe au présent simple : 20 points Conjuguer des verbes au passé simple : 20 points

TABLE 2.1 – Types d'exercices du test de notre système

## 2.5.2 Activités d'apprentissage

1. **Correspondance image texte** : Cette activité permet de renforcer l'association entre images et mots en Anglais. Pour chaque exercice , l'apprenant doit faire correspondre trois mots à trois images sélectionnées à partir d'Unsplash<sup>8</sup>.
2. **Correspondance texte texte** : L'apprenant devra organiser des mots, décomposés à partir d'une phrase générée par le modèle de langage (en fonction du niveau du CECR de l'apprenant), qui sont désordonnés afin de former une phrase correcte sur le plan sémantique.
3. **Carte de mots** : Cette fonctionnalité offre une approche visuelle et interactive pour apprendre de nouveaux mots. Chaque carte affiche un mot anglais avec une brève description ainsi qu'une traduction en Arabe et en Français.
4. **Lire un texte et répondre aux questions** : Dans cette activité, l'apprenant est encouragé à lire un texte puis répondre à des questions. Deux versions sont disponibles : l'une destinée aux enfants et l'autre à un public plus âgé. Les différences entre les deux résident dans les sujets des textes. Pour les enfants, des histoires ainsi que des contes sont proposés, contrairement au public plus large où une variété de sujets est abordée, ainsi que le nombre de questions posées (trois pour les enfants, tandis que le nombre de questions pour le large public est le nombre maximal de questions généré par le LLM). Cette activité est également disponible dans la section Examen. Elle vise à évaluer la capacité de compréhension de texte de l'apprenant.
5. **Prononcez cette phrase** : L'activité utilise un système de reconnaissance vocale. L'apprenant est invité à prononcer à haute voix une phrase générée par le LLM adaptée à son niveau CECR, qu'il pourra lire et écouter. Le système fournit ensuite un retour instantané sur la qualité de la prononciation, surlignant les mots mal prononcés. De plus, il évalue et retourne une note globale sur la prononciation.
6. **Choisissez la bonne traduction** : L'apprenant doit choisir la traduction appropriée d'un mot anglais vers une langue qu'il choisit.
7. **Associez les paires** : Dans cette activité, les apprenants doivent associer des mots anglais à leurs équivalents français ou arabes.
8. **Le pendu** : Le but de ce jeu éducatif est de tester le vocabulaire anglais. L'apprenant essaie de deviner un mot en proposant des lettres. Chaque lettre correcte est révélée dans le mot, mais chaque lettre incorrecte entraîne l'affichage d'une partie d'un dessin représentant un personnage. L'objectif est de deviner le mot avant que le dessin du personnage ne soit complet, i.e. toutes ses parties complétant sa pendaison.
9. **Associez un verbe à son expression** : L'apprenant devra associer des verbes à leurs actions correspondantes, par exemple "paint" à "masterpiece". Les verbes utilisés seront sélectionnés à partir de notre dictionnaire, en s'assurant que leur classe grammaticale est "VERB". Ensuite, le LLM devra proposer une expression qui complète ce verbe. Cette activité a pour but de renforcer la compréhension des relations entre les verbes et les expressions en Anglais.
10. **Synonymes ou mots de la même famille** : Dans cette activité, l'apprenant sera présenté avec des mots en Anglais et devra identifier leurs synonymes ou un mot du même champ lexical. Cette activité vise à enrichir le vocabulaire de l'apprenant.

8. Unsplash est une plateforme en ligne offrant une vaste collection d'images de haute qualité, souvent utilisées pour des projets éducatifs.

**11. Écrire cette phrase :** Cette activité mettra à l'épreuve la capacité de l'apprenant à écrire correctement des phrases en Anglais. Une phrase en Anglais sera présentée à l'apprenant via un audio, générée par le LLM de la même manière que dans l'activité "prononcez cette phrase" et il devra ensuite la réécrire.

```
convo.send_message(  
    f"Generate me a short phrase of 5 words in English with a meaning.  
    I'm going to take the phrase and jumble the words to create a game. You have to reorder them.  
    Please provide a phrase suitable for CEFR level {cefr}.")
```

En plus de ces activités, nous avons ajouté quatre activités suivantes dédiées seulement à la section examens :

- Expression écrite :** Grâce au LLM Gemini de Google, un sujet est généré pour l'apprenant en fonction de son niveau CECRL. Ensuite, l'apprenant rédigera son expression écrite en réponse au sujet et enverra sa réponse. Gemini attribuera alors une note sur 20 en fonction de plusieurs critères tels que la grammaire, le vocabulaire, etc., tout en ajoutant un commentaire indiquant ce qui était bon et ce qui devrait être amélioré dans son expression écrite.
- Conjugaison au “present simple” :** l'apprenant devra conjuguer un verbe au présent simple pour tous les pronoms personnels.
- Conjugaison au “past simple” :** À partir du dictionnaire du niveau correspondant, cinq verbes sont choisis , et l'apprenant devra conjuguer chacun au passé simple.
- Entendre un texte puis répondre à des questions :** Dans cette activité, l'apprenant est invité à écouter un texte audio, qu'il pourra réécouter plusieurs fois. Ensuite, il lui sera présenté des questions de type Vrai ou Faux auxquelles il devra répondre. Cette activité testera la capacité d'écoute et de compréhension de l'apprenant.

### 2.5.3 Recommandation d'exercices

Notre système de recommandation d'exercices repose sur une classification précise des mots en fonction de leur niveau de difficulté selon le CECR. Cette classification nous permet d'offrir aux apprenants des activités adaptées à leurs niveaux de compétence linguistique, favorisant ainsi un apprentissage progressif et efficace. Nous avons pour cela mis en place une stratégie de sélection rigoureuse, où chaque mot est attribué à un niveau CECR. Cette classification nous permet de recommander des mots variés et appropriés à chaque activité, garantissant ainsi une expérience d'apprentissage optimale pour chaque apprenant.

En résumé, notre approche de recommandation d'exercices basée sur les niveaux CECR offre aux apprenants une expérience d'apprentissage individualisée et efficace, en proposant des activités adaptées à leur niveau de compétence linguistique. Cette personnalisation des recommandations, combinée à une stratégie de sélection rigoureuse, constitue un élément clé de notre système, visant à soutenir et à faciliter le processus d'apprentissage de la langue anglaise pour tous les apprenants.

## 2.5.4 Consultation du dictionnaire

Dans cette partie, deux solutions s'offrent à l'apprenant : il peut rechercher n'importe quel mot dans la partie dictionnaire ou bien explorer les mots aléatoirement avec word card. Il peut également voir les mots qu'il a déjà rencontrés durant ses exercices. Les enseignants peuvent également ajouter leurs propres dictionnaires dans cette partie.

## 2.6 Partie ingénierie du logiciel de notre plateforme

### 2.6.1 Modélisation des fonctionnalités

La figure 2.4 présente le diagramme de cas d'utilisations, en considérant les trois acteurs principaux, l'apprenant, l'enseignant, et l'administrateur :

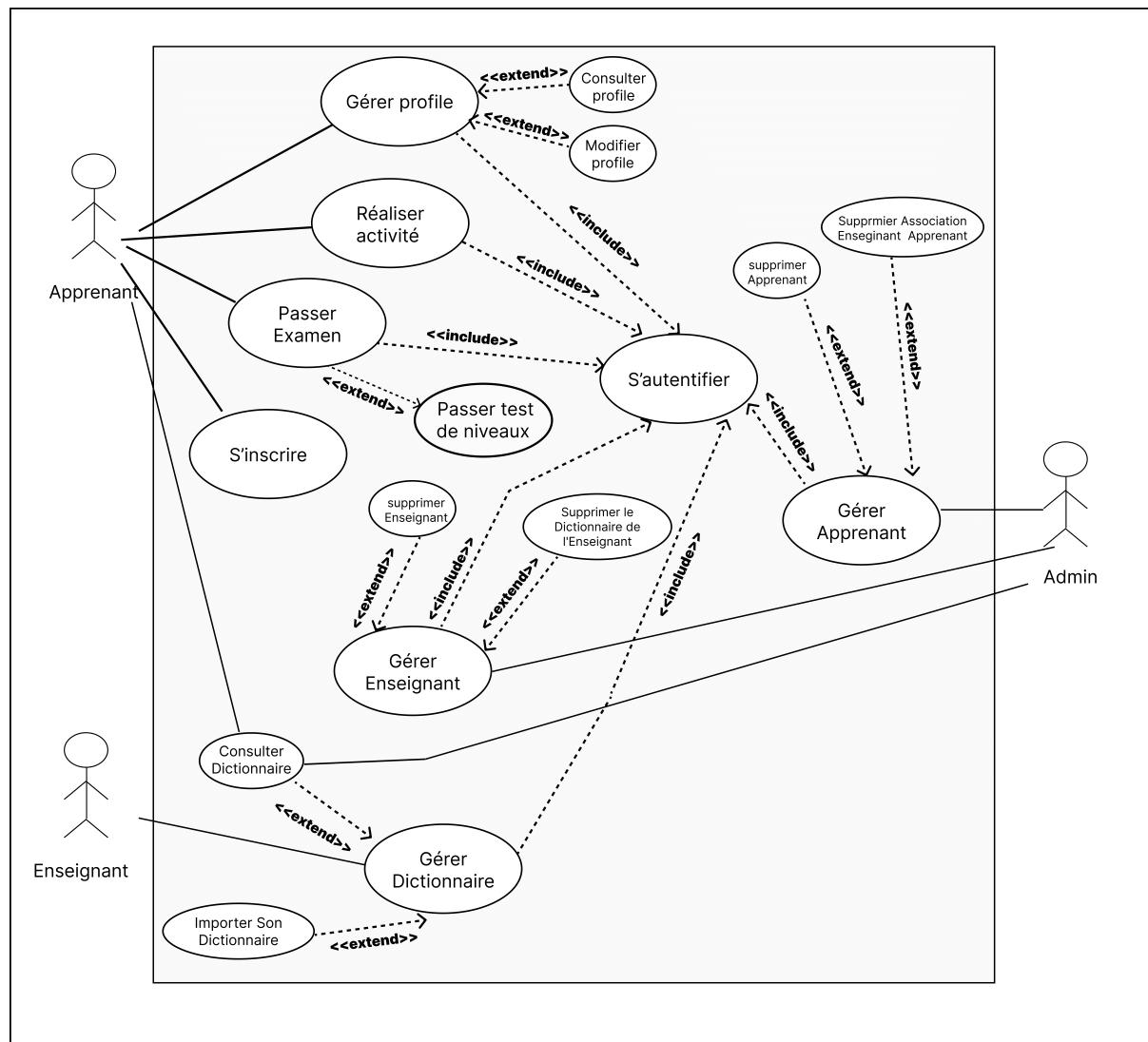


FIGURE 2.4 – Modélisation des fonctionnalités

## 2.6.2 Diagramme de classes

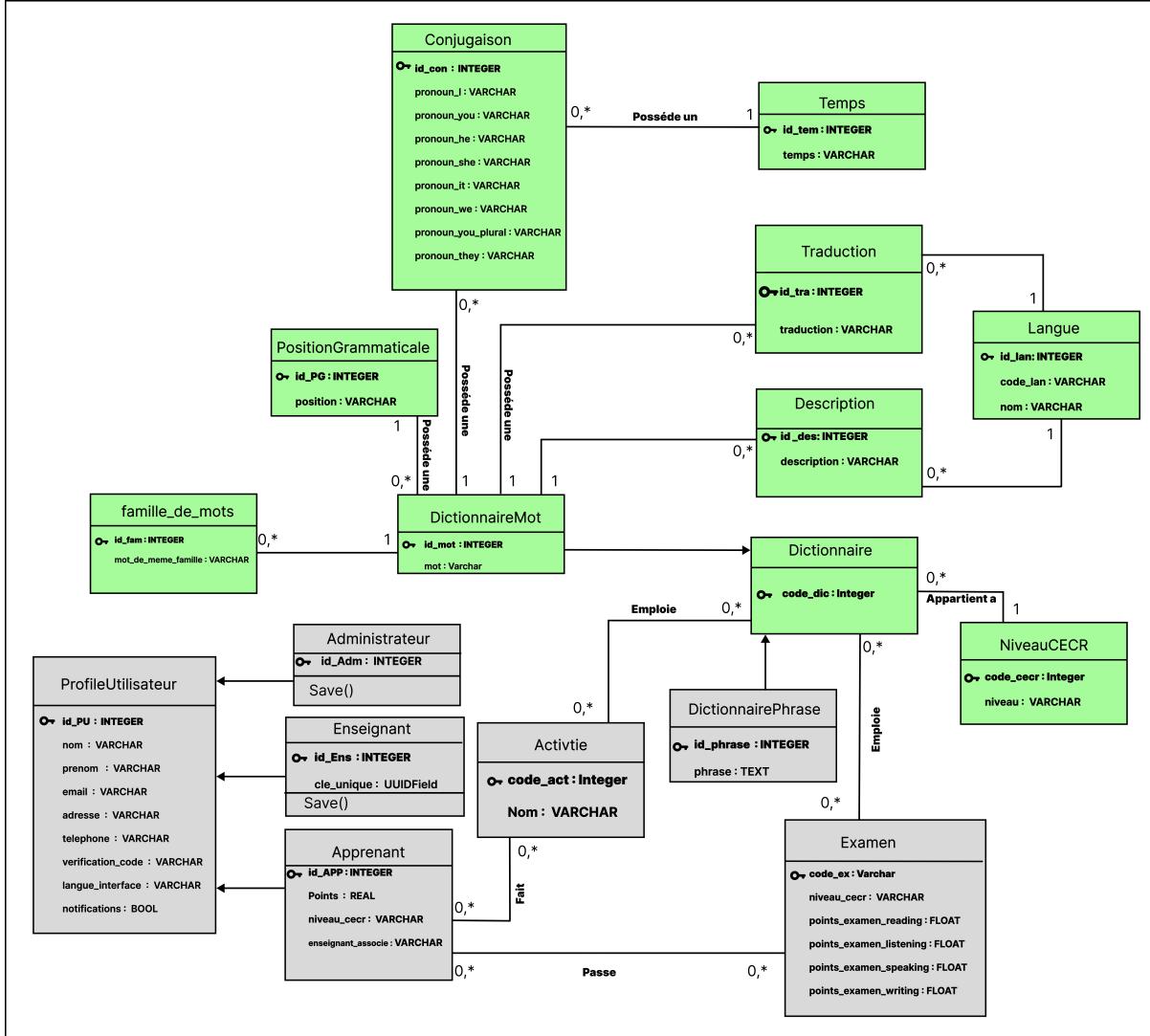


FIGURE 2.5 – Diagramme De Classes

## 2.6.3 Modèle relationnel des Données

- NiveauCECR(code\_cecr, niveau)
- Langue(id\_lan, code\_lan, nom)
- Temps(id\_tem, temps)
- PositionGrammaticale(id\_PG, position)
- Dictionnaire(code\_dic, niveau\_cecr\*)
- DictionnaireMot(id\_mot, mot, id\_PG\*)
- DictionnairePhrase(code\_phrase, phrase)
- Description(id\_des, mot\*, langue\*, description)
- Traduction(id\_tra, mot\*, langue\*, traduction)
- famille\_de\_mots(id\_fam, mot\*, mot\_de\_meme\_famille)

- Conjugaison(id\_con, mot\*, temps\*, pronom\_I, pronom\_you, pronom\_he, pronom\_she, pronom\_it, pronom\_we, pronom\_you\_pluriel, pronom\_they)
- ProfileUtilisateur(id\_PU, nom, prenom, email, adresse, telephone, verification\_code, langue\_interface, notifications)
- Administrateur(id\_Adm)
- Enseignant(id\_Ens, cle\_unique)
- Apprenant(id\_APP, niveau\_cecr, points, enseignant\_associe)
- Activite(code\_act, nom)
- Examen(code\_ex, niveau\_cecr, points\_examen\_reading, points\_examen\_writing, points\_examenSpeaking, points\_examen\_listening)
- Examen-Dictionnaire(code\_ex\*,code\_dic\*)
- Passer-examen(code\_ex\*,id\_APP\*)
- Realiser-activite(id\_APP\*,code\_act\*)
- Activite-Dictionnaire(code\_act\*,code\_dic\*)

## 2.7 Conclusion

Dans ce chapitre, nous avons examiné la conception de notre plateforme d'apprentissage, en soulignant la modélisation du dictionnaire par niveaux et la description des activités offertes.

Le chapitre suivant sera dédié à la présentation des aspects techniques liés au développement de notre système.

# Chapitre 3

## Implémentation

### 3.1 Introduction

Ce chapitre présente les aspects techniques liés à la mise en œuvre de notre plateforme d'apprentissage de la langue anglaise appelée "TaalimNet". Il sera structuré en trois sections : une description de l'environnement de développement, la création du dictionnaire par niveaux et une description de la plateforme d'apprentissage.

### 3.2 Environnement de développement

#### 3.2.1 Environnement matériel

Deux ordinateurs personnels ont été utilisés dans la réalisation de ce projet, chacun offrant les capacités suivantes :

Poste de travail	Système d'exploitation	Type	Processeur	RAM
1	Windows 11 Familiale	Base x64	Intel(R) Core(TM) i5-10210U CPU @ 1.60GHz	16 GO
2	Windows 11 Familiale	Base x64	AMD Ryzen 5 4600H with Radeon Graphics @ 3.00 GHz	16 GO

TABLE 3.1 – Caractéristiques des postes de travail.

Nous avons également utilisé Google Colab, un environnement de développement collaboratif en ligne qui offre des ressources de calcul gratuites basées sur le cloud et se révèle puissant et efficace pour l'exécution de tâches de traitement de données et d'apprentissage automatique.

#### 3.2.2 Architecture technique

Cette partie aborde les éléments techniques concernant les outils, les langages et les bibliothèques employés dans notre processus de développement.

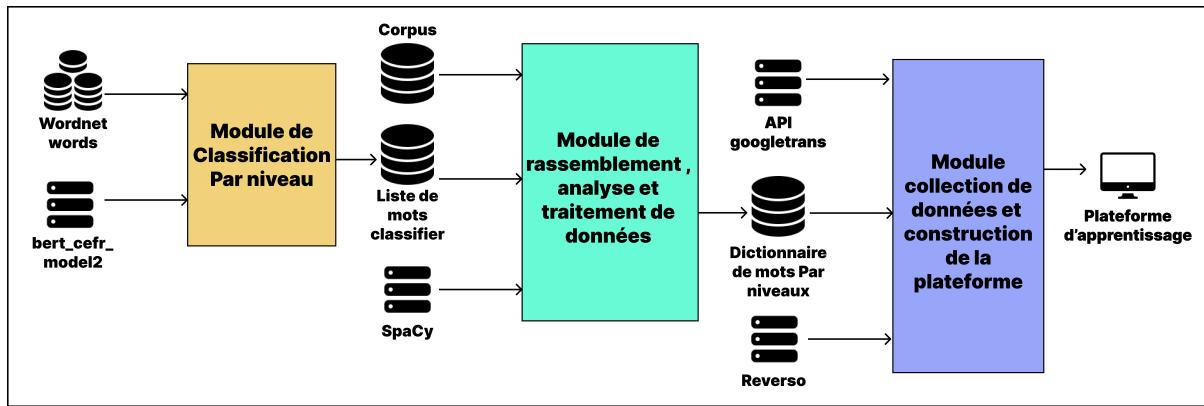


FIGURE 3.1 – Architecture Technique.

## Ressources utilisées

- **Corpus** : Les trois dictionnaires par niveau déjà collectés.
- **WordNet** : Une base de données lexicale de la langue anglaise, où les noms, les verbes, les adjectifs et les adverbes sont regroupés en ensembles de synonymes cognitifs (synsets), chacun exprimant un concept distinct.
- **SpaCy** : Une bibliothèque Python pour le traitement du langage naturel, qui offre des fonctionnalités telles que l'étiquetage grammatical, etc.
- **Reverso** : Une plateforme en ligne utilisée pour extraire les conjugaisons des verbes, ce qui peut être utile pour l'analyse linguistique et la génération de synonymes, d'antonymes ...
- **API Translation** : Une interface de programmation d'application qui permet aux développeurs d'intégrer les fonctionnalités de traduction automatique de Google.
- **Modèle BERT\_CEFR\_Model2** : Un modèle Hugging Face de TALN basé sur BERT pour la classification selon le CEFR.

## Modules

- **Module de classification Par niveau**
  - Ce module vise à classer les mots de WordNet selon les niveaux du CEFR.
- **Module de rassemblement, analyse et traitement de données**
  - Ce module a pour but de collecter les mots classifiés, les analyser avec le corpus et les nettoyer. Ensuite, il procède à un étiquetage selon leur classe grammaticale.
- **Module collection de données et construction de la plateforme**
  - Ce dernier module est dédié à la collecte des mots traités dans le module précédent, afin de les exploiter dans des activités pédagogiques nécessaires à l'apprentissage de la langue anglaise. Nous utiliserons des outils externes tels que le LLM Gemini et une API de traduction, etc.

### 3.2.3 Langages de programmation, outils et formats de données

Dans le cadre de ce projet, une stratégie polyvalente en termes de langages de programmation, d'outils et de formats de données a été déployée pour assurer le succès des analyses et le développement d'une plateforme robuste et fonctionnelle.

- **Langage de programmation** : Pour commencer, **Python** a été privilégié comme langage de développement principal en raison de sa polyvalence et de son riche écosystème de bibliothèques. Le choix du framework Django pour la création de l'interface web (Back-end) s'est imposé, bénéficiant de sa réputation de robustesse et de documentation complète. Par ailleurs, le Front-end de l'application a été élaboré en utilisant les technologies HTML, CSS et JavaScript, garantissant ainsi une expérience utilisateur fluide et interactive.
- **Bibliothèques TALN** : Dans le domaine du TALN, les bibliothèques **NLTK** et **Spacy** ont été exploitées pour leurs fonctionnalités avancées d'analyse linguistique. Pour diversifier le contenu, Gemini a été sélectionné pour sa capacité à générer des textes naturels et cohérents, contribuant ainsi à une expérience utilisateur de qualité.
- **Outils** : Pour répondre aux besoins de traduction, l'**API Google Translate** a été intégrée. De même, pour faciliter le scraping de données sur le web, l'outil **Beautifulsoup4** a été utilisé pour extraire les informations nécessaires à partir des sources en ligne.

Pour l'extraction de données, des ressources telles que **Reverso** ont été exploitées pour obtenir des informations détaillées sur les mots, incluant leur traduction, leur conjugaison, ainsi que leurs synonymes et antonymes. Parallèlement, l'intégration de la banque d'images **Unsplash** a permis d'obtenir une sélection d'images de haute qualité.

- **Sauvegarde et représentation des données** : En ce qui concerne la gestion des données, **SQLite** a été privilégiée pour sa légèreté et son efficacité dans le stockage de données structurées. Pour la manipulation et l'analyse des données, divers formats tels que **CSV** et **XLSX** ont été utilisés, ainsi que des outils avancés comme **Pandas**, particulièrement adaptés pour le traitement de données complexes.
- **Autres ressources** : L'utilisation de **Google Colab** a également été envisagée pour profiter de ses ressources de calcul étendues. Enfin, pour extraire efficacement des données à partir de pages web, l'extension **Chrome Table Capture** a été déployée, simplifiant ainsi le processus d'extraction à partir de sources variées.

## 3.3 Crédit du dictionnaire par niveaux

### 3.3.1 Partitionnement en niveaux

Le processus de partitionnement par niveaux débute par l'agrégation d'un vaste éventail de mots provenant de WordNet, visant à couvrir une large diversité lexicale. Cette démarche permet d'obtenir une représentation exhaustive du vocabulaire à étudier. Ensuite, cette compilation exhaustive de mots est soumise à une classification minutieuse grâce au modèle Hugging Face BERT\_CEFER\_model2, intégré dans un pipeline spécialisé. Ce pipeline exploite les capacités avancées du TALN du modèle pour attribuer à chaque mot un niveau CEFR prédit avec une précision remarquable. Cette approche

permet une segmentation précise des mots en différentes catégories de compétence linguistique, facilitant ainsi leur analyse et leur utilisation dans divers contextes.

### 3.3.2 Analyse de corpus

Après avoir segmenté la wordlist, une étape fondamentale consiste à fusionner les données issues des trois dictionnaires pré-classifiés avec la wordlist initiale. Cette fusion vise à enrichir la base de données en y intégrant des informations supplémentaires sur le vocabulaire. Ensuite, pour assurer la fiabilité des données, un processus rigoureux de tri est mis en œuvre afin d'éliminer les doublons et les termes indésirables qui pourraient fausser les résultats de l'analyse. Cette étape de nettoyage garantit la qualité des informations utilisées pour alimenter la plateforme.

Pour aller encore plus loin dans l'analyse linguistique, une classification grammaticale est réalisée grâce à l'utilisation de la bibliothèque Spacy. Cette étape permet d'identifier le rôle de chaque mot dans une phrase, que ce soit en tant que nom, verbe, adjectif, etc. En fournissant une analyse fine de la structure grammaticale, cette classification grammaticale enrichit notre compréhension du texte et facilite l'extraction de sens et d'informations pertinentes.

### 3.3.3 Collecte de données

La collecte de données s'appuie sur plusieurs sources, notamment Reverso, l'API de traduction et WordNet. Reverso fournit des informations précieuses telles que les traductions, les conjugaisons, les synonymes et les antonymes des mots. L'API de traduction, utilisant Google Translate sans clé d'API, permet de traduire les descriptions et les mots nécessaires à la compréhension et à l'enrichissement du corpus. Enfin, WordNet est utilisé pour obtenir des descriptions détaillées à partir des mots, contribuant ainsi à une meilleure compréhension du vocabulaire et de ses nuances.

#### Statistiques

Les tableaux suivants donnent des statistiques collectées sur les nombres de mots par niveaux et selon leurs catégories grammaticales (voir Tableau 3.2), les nombres de mots par niveau CECR (voir Tableau 3.3) pour les trois Dictionnaires précédemment mentionnés, et les nombres de Texte par niveaux CECR (voir Tableau 3.4) du kaggle dataset.

Niveau CECR	A1	A2	B1	B2	C1	C2
Noun	574	974	1514	2033	1435	1111
Coordinating Conjunction	4	5	6	6	1	0
Subordinating Conjunction	6	12	17	10	4	3
Adposition	29	38	38	31	19	18
Numeral	43	13	1	4	1	0
Adverb	75	161	254	380	302	234
Particle	2	2	2	1	0	0
Verb	253	489	827	1048	727	604
Proper Noun	73	105	170	240	158	149
Adjective	149	288	555	737	584	512
Auxiliary	17	11	10	6	1	3
Determiner	5	2	4	4	3	0
Interjection	16	13	8	8	3	4
Autre	49	36	24	18	3	3

TABLE 3.2 – Distribution des types de mots par niveau CECR

Niveau CECR	Nombre de mots
A1	1295
A2	2149
B1	3430
B2	4526
C1	3241
C2	2641
Total	17282

TABLE 3.3 – Nombre de mots par niveau CECR

Niveau CECR	Nombre de Texte
A1	287
A2	286
B1	272
B2	241
C1	205
C2	202
Total	1493

TABLE 3.4 – Nombre de Texte par niveau CECR

Ayant combiné les différentes ressources mentionnées dans la section 3.1 du chapitre “Conception”, nous avons obtenu des dictionnaires par niveaux CECR avec les statistiques mentionnées dans le Tableau 3.5.

Niveau CECR	Nombre de mots
A1	872
A2	9060
B1	3973
B2	52110
C1	66148
C2	12834
Total	144997

TABLE 3.5 – Nombre de mots par niveau CECR

Le tableau 3.6 détaille les différentes classes grammaticales des mots du dictionnaire. Ces mots ont été obtenus grâce au traitement effectué par la bibliothèque SpaCy pour les mots du corpus de WordNet, pour les mots provenant de dictionnaires préexistants, ils ont été intégrés directement sans traitement par SpaCy, et organisés selon leurs niveaux CECR.

Niveau CECR	A1	A2	B1	B2	C1	C2
Verb	161	1686	846	9070	12000	1905
Adjective	101	697	420	4581	5096	2722
Adverb	39	407	209	2171	3262	926
Interjection	8	66	16	137	229	12
Adposition	5	101	15	272	425	21
Noun	435	3786	1705	22607	28026	5926
Proper Noun	77	1757	726	11	9261	1310
Numeral	29	230	8	588	1547	1
Particle	0	0	1	2	0	0
Auxiliary	4	8	2	9	16	4
Coordinating Conjunction	2	14	1	9	106	1
Subordinating Conjunction	0	4	3	24	2	0
Determiner	0	2	0	8	8	0
AUTRE	11	302	21	1523	6170	6

TABLE 3.6 – Nombre de mots par niveau CECR et catégories grammaticales

## 3.4 Description de la plateforme d'apprentissage “Taa-limNet”

### 3.4.1 Analyse des couleurs de l'interface

Les choix de couleurs pour l'interface utilisateur de notre plateforme ont été méticuleusement sélectionnés pour optimiser l'expérience utilisateur. Le blanc (#FFFFFF) offre une base claire et lumineuse, tandis que le orange clair (#FFBB00) injecte de l'énergie et de l'enthousiasme. Le turquoise (#60E4C4) apaise l'esprit et favorise la concentration, et le bleu foncé (#30344C) offre un contraste subtil pour une expérience visuelle équilibrée. Cette palette vise à créer un environnement propice à l'apprentissage efficace et à l'engagement des utilisateurs.

### 3.4.2 Présentation de quelques fonctionnalités

Nous présentons quelques fonctionnalités de notre plateforme “TaalimNet” dédiée à l’apprentissage de langue anglaise.

La figure 3.2 montre la page d’authentification et d’inscription (Accueil)



FIGURE 3.2 – Page d’authentification et d’inscription

Les figures 3.3-3.5 montrent respectivement la page d’accueil de l’apprenant, la page liée à son profil et la page de son dictionnaire.

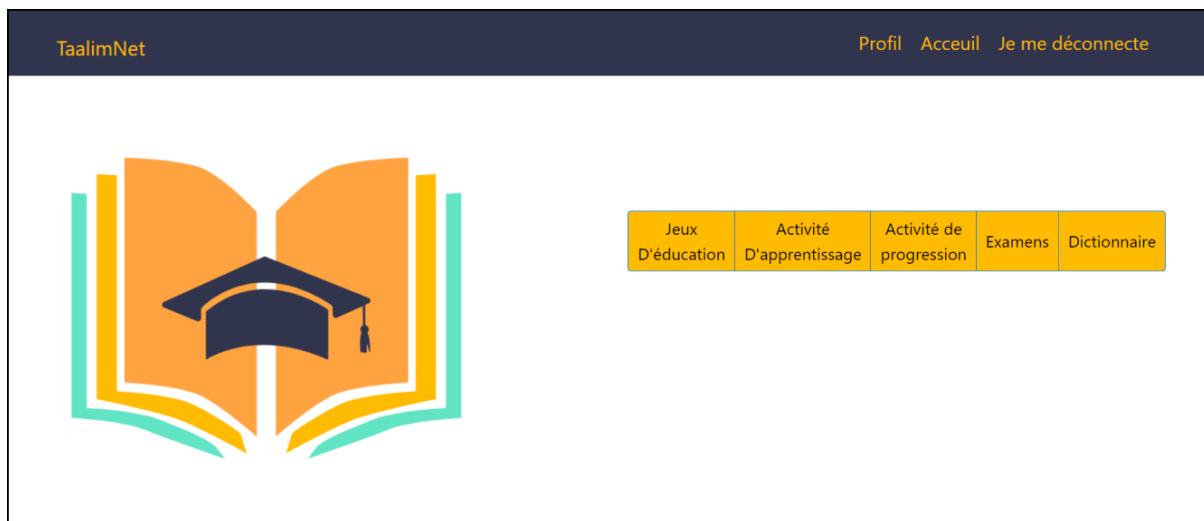


FIGURE 3.3 – Page d’accueil d’un apprenant

TaalimNet

Profil Acceuil Je me déconnecte

### Profile utilisateur : etudx

Nom: haddad

Prénom: Leila

Email: sartzy1869@gmail.com

Adresse: CITE 720

Téléphone: 18698691869

code professeur: None

Langue Interface: Français

Notifications Activées:

### Etudiant Information

Points: 1.27

Niveau CEFR: A1

**Enregistrer**

FIGURE 3.4 – Page de profil de l'apprenant

Apprends tes mots à partir du dictionnaire

Quitter

### Dictionnaire

Mots	Niveau CEFR
able	B1
unable	B1
abaxial	C1
dorsal	C1
adaxial	C1

FIGURE 3.5 – Page du Dictionnaire de l'apprenant

Les figures 3.6 et 3.7 permettent respectivement la consultation du dictionnaire et l'importation du dictionnaire d'un enseignant.

The screenshot shows a web application titled "Apprends tes mots à partir du dictionnaire". At the top right is a "Quitter" button. Below the title is a search bar with placeholder text "Search...", a dropdown menu for "Filtre par niveau CEFR" (with "Tous les mots" selected), a yellow "Filtré" button, and a blue "Importer Votre Dictionnaire" button. The main content area is titled "Dictionnaire". It contains a table with two columns: "Mots" and "Niveau CEFR". The data is as follows:

Mots	Niveau CEFR
last	A1
cut	A1
living	A1
ideal	A1
real	A1

FIGURE 3.6 – Page du Dictionnaire d'un enseignant.

The screenshot shows a web application with a dark header bar containing the logo "TaalimNet" on the left and "Profile Acceuil Je me déconnecte" on the right. The main content area is titled "Importer un fichier CSV". It features a text input field labeled "Csv file:" with a placeholder "Choisir un fichier" and a message "Aucun fichier choisi". Below the input field is a large yellow "Importer" button.

FIGURE 3.7 – Page importer un dictionnaire d'un enseignant.

### 3.4.3 Présentation de quelques activités d'apprentissage

Nous présentons quelques exemples d'activités d'apprentissage (voir figures 3.8 , 3.9 ,3.10 et 3.11). Le reste des activités développées sera représenté en annexe (voir Annexe A).

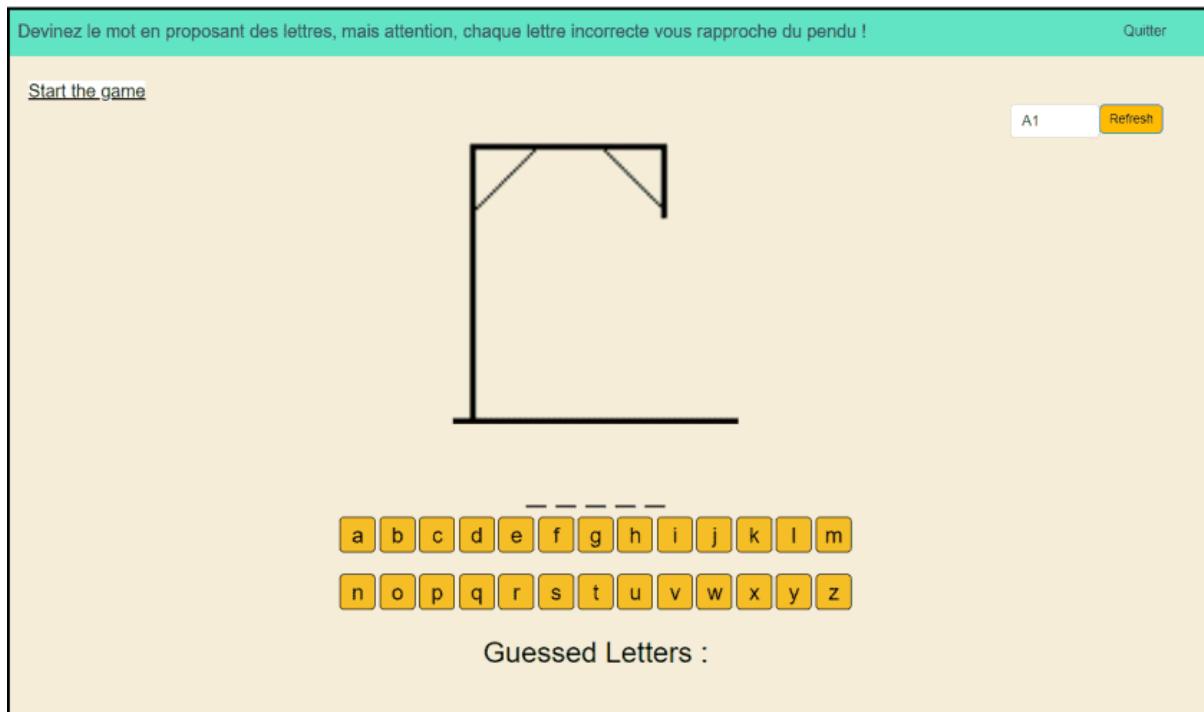


FIGURE 3.8 – Exemple d'activité du Pendu



FIGURE 3.9 – Exemple d'activité de Correspondance Texte-Texte



FIGURE 3.10 – Exemple d’activité de prononciation



FIGURE 3.11 – Exemple d’activité de Coupler les mots

### 3.5 Conclusion

Dans ce chapitre, nous avons présenté en détail les aspects techniques liés au développement de notre plateforme d’apprentissage d’anglais. Nous avons également illustré quelques-unes des fonctionnalités développées.

# Conclusion générale

Dans le but de créer une application d'apprentissage de la langue anglaise adaptée au contexte algérien, nous avons cherché à répondre à la problématique de l'importance et de la prévalence croissante de l'anglais. Après avoir examiné les ressources existantes et les spécificités de la langue anglaise, nous avons d'abord élaboré un dictionnaire de mots. En utilisant le modèle BERT, nous avons classé les mots de WordNet par niveaux du CECR. Ces mots ont été par la suite intégrés avec ceux de dictionnaires reconnus comme Oxford et Cambridge, puis étiquetés grammaticalement grâce à la bibliothèque spaCy.

Les contributions majeures de ce projet de fin d'études sont la construction d'un dictionnaire structuré par niveaux CECR, comprenant un total de 144 997 mots. Ce dictionnaire a été intégré dans notre plateforme d'apprentissage TaalimNet, qui propose des activités pédagogiques et un test de niveaux. Cette plateforme permet aux apprenants de développer tous les aspects de la langue, notamment la prononciation, l'écriture, l'écoute et la lecture, à travers une variété d'exercices qui leur sont recommandés automatiquement. Ces activités sont générées grâce à l'utilisation de LLM de Google Gemini selon leurs niveaux. Elles sont adaptées au contexte Algérien, puisqu'elles proposent des traductions en arabe et en français pour faciliter la compréhension des apprenants. Cette approche bilingue permet aux utilisateurs de mieux saisir les nuances de la langue anglaise en se référant à leurs langues maternelles ou d'usage courant, ce qui rend l'apprentissage plus accessible et moins intimidant.

Comme perspectives futures, nous prévoyons la validation humaine du dictionnaire par des experts et la collecte du feedback d'utilisation de la plateforme, en la mettant à disposition d'un groupe d'apprenants et d'enseignants pour une évaluation initiale, avant de la déployer plus largement auprès de la communauté algérienne. De plus, nous envisageons de développer d'autres types d'activités, telles que des activités de groupe, afin d'adapter notre solution à un environnement social, telles que des débats virtuels, des groupes de lecture interactive, des quiz et jeux interactifs, des ateliers de rédaction, etc.

# Bibliographie

- [1] Jakub Jagoda and Tomasz Boiński, 2018, *Assessing Word Difficulty for Quiz-like Game*.
- [2] Pedro Curto, Nuno Mamede and Jorge Baptista, 2015, *Automatic Text Difficulty Classifier Assisting the Selection Of Adequate Reading Materials For European Portuguese Teaching*.
- [3] Klinton Bicknell and Claire Brust, 2020, *Learning how to help you learn : Introducing Birdbrain!*, <https://blog.duolingo.com/learning-how-to-help-you-learn-introducing-birdbrain/>.
- [4] Guide des épreuves de l'examen TOEFL, consulté en mars 2024, <https://www.bestmytest.com/toefl/practice-test>.
- [5] Activités proposées par Duolingo, consulté en mars 2024, <https://duolingo.fandom.com/wiki/Exercise>.
- [6] Daniel Jurafsky and James H. Martin, 2009, *Speech and Language Processing (second edition)*, <https://aclanthology.org/J09-3005.pdf>.
- [7] Enkelejda Kasneci, Kathrin Sessler and many others, 2023, *ChatGPT for good ? On opportunities and challenges of large language models for education*, <https://www.sciencedirect.com/science/article/abs/pii/S1041608023000195>.
- [8] Guide sur le fine-tuning des LLMs, consulté en février 2024, <https://learn.microsoft.com/en-us/ai/playbook/technology-guidance/generative-ai/working-with-l1ms/fine-tuning>.
- [9] Cadre européen commun de référence pour les langues (CECR), consulté en janvier 2024, <https://www.cambridgeenglish.org/fr/exams-and-tests/cefr/>.
- [10] Informations sur le test de niveau IELTS, consulté en mars 2024, <https://ielts.org/about-ielts>.
- [11] Informations sur les tests de niveau de Cambridge, consulté en mars 2024, <https://www.cambridgeenglish.org/exams-and-tests/qualifications/new-exam-names/>.
- [12] Informations sur le test de niveau TOEIC, consulté en mars 2024, <https://www.iibc-global.org/english/toeic/test/lr/about.html#:~:text=The%20TOEIC%20L%26R%20is%20an,one%20a%20separate%20answer%20sheet..>

- [13] Dictionnaire anglais de Cambridge, consulté en février 2024, <https://dictionary.cambridge.org/fr/dictionnaire/anglais/>.
- [14] Dictionnaire anglais par niveaux d’Oxford, consulté en janvier 2024, <https://www.oxfordlearnersdictionaries.com/wordlists/oxford3000-5000>.
- [15] Pr. Cécile Godé, 2024, *Développer des compétences avec les IA génératives*, <https://hal.science/hal-04550169/document>.
- [16] Jacob Devlin, Ming-Wei Chang and many others, 2019, *BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding*, <https://arxiv.org/pdf/1810.04805.pdf>.
- [17] Pethigamage Perera, Madhushan Lankathilake, 2023, *Preparing to Revolutionize Education with the Multi-Model GenAI Tool Google Gemini ? A Journey towards Effective Policy Making*, [https://www.researchgate.net/publication/373066196\\_Preparing\\_to\\_Revolutionize\\_Education\\_with\\_the\\_Multi-Model\\_GenAI\\_Tool\\_Google\\_Gemini\\_A\\_Journey\\_towards\\_Effective\\_Policy\\_Making](https://www.researchgate.net/publication/373066196_Preparing_to_Revolutionize_Education_with_the_Multi-Model_GenAI_Tool_Google_Gemini_A_Journey_towards_Effective_Policy_Making).
- [18] Tianyu Wu, Shizhu He and many others, 2023, *A Brief Overview of ChatGPT : The History, Status Quo and Potential Future Development*, <https://ieeexplore.ieee.org/abstract/document/10113601>.
- [19] Ebtesam Almazrouei, Hamza Alobeidli and many others, 2023, *The Falcon Series of Open Language Models*, <https://arxiv.org/pdf/2311.16867.pdf>.
- [20] Yinhan Liu, Myle Ott and many others, 2019, *RoBERTa : A Robustly Optimized BERT Pretraining Approach*, <https://arxiv.org/abs/1907.11692>.
- [21] Victor Sanh, Lysandre Debut and many others, 2020, *DistilBERT, a distilled version of BERT : smaller, faster, cheaper and lighter*, <https://arxiv.org/abs/1910.01108>.
- [22] Zhenzhong Lan, Mingda Chen and many others, 2020, *ALBERT : A Lite BERT for Self-supervised Learning of Language Representations*, <https://arxiv.org/abs/1909.11942>.
- [23] Kevin Clark, Minh-Thang Luong and many others, 2020, *ELECTRA : Pre-training Text Encoders as Discriminators Rather Than Generators*, <https://arxiv.org/abs/2003.10555>.
- [24] English Vocabulary Profile Online - British English-Cambridge, consulté en janvier 2024, <https://www.englishprofile.org/wordlists/evp>.
- [25] Open Language Profiles — English datasets from CEFR-J.Tono Laboratory at TUFS (Tokyo University of Foreign Studies), consulté en février 2024, <https://github.com/openlanguageprofiles/olp-en-cefrj>.
- [26] Adam Montgomerie, 2021, *CEFR Levelled English Texts*, consulté en janvier 2024, <https://www.kaggle.com/datasets/amontgomerie/cefr-levelled-english-texts>.

- [27] Pretrained model that predicts CEFR level of english words “bert\_cefr\_model2”, consulté en février 2024, [https://huggingface.co/kalobiralo/bert\\_cefr\\_model2](https://huggingface.co/kalobiralo/bert_cefr_model2).
- [28] Ayat Abodayeh, Reem Hejazi and many others, 2023, *Web Scraping for Data Analytics : A BeautifulSoup Implementation*, <https://ieeexplore.ieee.org/abstract/document/10145369>.
- [29] Spacy documentation, consulté en février 2024, <https://spacy.io/api/textcategorizer>.

# Annexe

## A. Activité d'apprentissage (suite)

Voici la suite des exemples d'activités d'apprentissage

Correspondez chaque texte à son image Quitter

**Correspondance Image-Texte**

Coffee	dog	Desk

↻

FIGURE 3.12 – Exemple d'activité de Correspondance Image Texte

Lisez le mot, sa définition dans les trois langues Quitter

A1 English

become

↻

FIGURE 3.13 – Exemple d'activité de Mot en carte

### The Little Boy and the Star

Once upon a time, there was a little boy named Tommy who lived in a small village. One night, as he was looking up at the stars, he saw a bright star fall from the sky. He ran over to where it had landed and found a tiny, sparkling star lying in the grass. Tommy picked up the star and took it home. He kept it in a jar by his bed and talked to it every night. He told the star all about his hopes and dreams, and the star listened patiently. One day, Tommy's father told him that he had to move to a new town. Tommy was heartbroken to leave the star, but he knew he couldn't take it with him. So, he went to the meadow where he had found it and placed it gently on a flower.

#### True/False Questions about the story

1. Tommy found a star in the forest.

True  False

2. The star was bigger than Tommy's hand.

True  False

3. Tommy kept the star in a drawer.

True  False

**Envoyer les Réponses**

FIGURE 3.14 – Exemple d'activité de Lire un texte et répondre aux questions

liked

it now.

call

money with care.

buy

me anytime

borrow

it very much

A1 ▾



FIGURE 3.15 – Exemple d'activité de Phrases Verbes

choisissez la bonne traduction parmi ces trois choix.

Quitter

A1 Filtré

**prefer**

- attacher
- fendu
- préférer

soumettre

FIGURE 3.16 – Exemple d’activité de Choisir la bonne traduction

Associez chaque mot avec son synonyme ou avec un mot qui a le même champ lexical

Quitter

A1

Words	synonyms
interesting	good
foreign	excited
delicious	out
hungry	excellent
dangerous	low

↻

FIGURE 3.17 – Mots de la même famille

Ecrivez la phrase que vous entendez.

Quitter

Phrase aléatoire:

Write your sentence here...

Vérifier la réponse

↻

FIGURE 3.18 – Exemple d’activité d’écrire la phrase que vous entendez

Ecoutez le texte puis repondez par Vrai ou Faux

Quitter



**True/False Questions about the story**

1. He adored his new piece of outerwear.

True  False

2. The jacket's hue was a shade of brown.

True  False

FIGURE 3.19 – Exemple d'activité de Entendre un texte et répondre aux questions

Expression écrite

Quitter

**Your Topic**

Write a simple greeting card message to a family member or friend. Express your thoughts and feelings using basic English. The message should be at least 2-3 lines long.

Write down your expression here

Write your expression here...

Submit

FIGURE 3.20 – Exemple d'activité d'expression écrite

Conjuguez le verbe "to cook" au présent simple

Quitter

I:

you:

he/she/it:

we:

you:

they:

**Submit**

FIGURE 3.21 – Exemple d'activité de Conjugaison au présent simple

Conjuguez chaque verbe au passé simple.

Quitter

rose:

believe:

watching:

return:

fishing:

**Submit**

FIGURE 3.22 – Exemple d'activité de conjugaison au passé simple

## B. Business Model Canvas

Ci-dessous, nous vous présentons notre Business Model Canvas (BMC) pour notre plateforme d'apprentissage de la langue anglaise, spécialement conçue pour les Algériens, appelée "TaalimNet". Dans un monde de plus en plus globalisé, la maîtrise de l'anglais est devenue essentielle pour les opportunités professionnelles et personnelles. Notre projet vise à combler ce besoin en offrant des solutions innovantes et accessibles pour apprendre l'anglais. À travers cette présentation, nous allons explorer les neuf blocs constitutifs de notre modèle d'affaires, en détaillant la proposition de valeur, les segments de clients, les canaux de distribution, et bien plus encore. Notre objectif est de démontrer comment cette plateforme peut transformer l'apprentissage linguistique en Algérie, en s'adaptant aux besoins spécifiques de notre public cible et en tirant parti des dernières technologies éducatives.



Business Model Canevas 2023/2024					Nom du projet : Recommandation d'activités pédagogiques pour l'apprentissage de la langue anglaise (possibilité de proposer ce thème comme projet startup)
Partenaires	Activités clés	Proposition de valeur	Relation Client	Segment Client	
<b>Établissements privés:</b> Partenariats avec des écoles, universités et centres de formation.  <b>Parents :</b> Cherchent des outils supplémentaires pour aider leurs enfants à apprendre l'anglais.	<b>Développement et Maintenance:</b> Crédit et mise à jour de la plateforme web et mobile.  <b>Création de Contenus :</b> Développement de nouveaux exercices et activités adaptées aux différents niveaux.	<b>Contenus Éducatifs Personnalisés :</b> Exercices et activités adaptées aux niveaux A1 à C2 du CEFR.  <b>Traduction Multilingue :</b> Traduction automatique en arabe et en français pour faciliter la compréhension.  <b>Accessibilité :</b> Plateforme disponible sur le web donc pas de type d'appareil obligatoire.	<b>Support Technique :</b> Assistance en ligne pour aider les utilisateurs avec des problèmes techniques ou questions sur l'utilisation de l'outil.  <b>Canaux de distribution</b>	<b>Élèves et Étudiants :</b> Apprenants de différents niveaux scolaires, du primaire à l'université.  <b>Enseignants :</b> Utilisation de l'outil pour améliorer les méthodes d'enseignement et suivre les progrès des élèves.  <b>Écoles et Institutions Éducatives :</b> Intégration de l'outil dans les programmes éducatifs pour renforcer l'apprentissage de l'anglais.	<b>Segment Client</b>
<b>Ressources Clés</b>	<b>Équipe de Développement :</b> Développeurs web et mobile pour la création et la maintenance de la plateforme.  <b>Experts en Éducation :</b> Linguistes, enseignants et pédagogues pour la création de contenus éducatifs.	<b>Partenariats Éducatifs :</b> Collaboration avec des écoles et centres de formation pour intégrer l'outil dans les programmes d'études.  <b>Médias Sociaux :</b> Utilisation des réseaux sociaux pour la promotion et l'engagement des utilisateurs.	<b>Partenariats Éducatifs :</b> Collaboration avec des écoles et centres de formation pour intégrer l'outil dans les programmes d'études.  <b>Médias Sociaux :</b> Utilisation des réseaux sociaux pour la promotion et l'engagement des utilisateurs.	<b>Parents :</b> Cherchent des outils supplémentaires pour aider leurs enfants à apprendre l'anglais.	
<b>Structure des coûts</b>		<b>Source de Revenus</b>			
<b>Développement Technologique :</b> Coûts de développement et de maintenance de la plateforme web.  <b>Création de Contenus :</b> Coûts liés à la création et à la mise à jour des ressources éducatives.  <b>Marketing et Ventes :</b> Dépenses pour les campagnes de promotion et l'acquisition de nouveaux utilisateurs.  <b>Support Client :</b> Coûts liés à l'assistance technique et au support utilisateur.  <b>Infrastructure :</b> Hébergement, serveurs et autres coûts technologiques.		<b>Abonnements :</b> Modèle d'abonnement pour les écoles et les utilisateurs individuels, avec différents niveaux d'accès (gratuit, premium).  <b>Publicité :</b> Revenus provenant de la publicité intégrée pour les utilisateurs gratuits.  <b>Vente de Contenus :</b> Vente de cours avancés et de contenus supplémentaires pour les utilisateurs premium.			

FIGURE 3.23 – Business Model Canvas

# Business Model Canevas

## 2023/2024

		Nom du projet : Recommandation d'activités pédagogiques pour l'apprentissage de la langue anglaise (possibilité de proposer ce thème comme projet startup)		
<b>Partenaires</b>	<b>Activités clés</b>	<b>Proposition de valeur</b>	<b>Relation Client</b>	<b>Segment Client</b>
<b>Établissements privés:</b> Partenariats avec des écoles, universités et centres de formation.  <b>Parents :</b> Cherchent des outils supplémentaires pour aider leurs enfants à apprendre l'anglais.	<b>Développement et Maintenance:</b> Crédit et mise à jour de la plateforme web et mobile.  <b>Création de Contenus :</b> Développement de nouveaux exercices et activités adaptées aux différents niveaux.	<b>Contenus Éducatifs</b> <b>Personnalisés :</b> Exercices et activités adaptés aux niveaux A1 à C2 du CEFR.  <b>Traduction Multilingue :</b> Traduction automatique en arabe et en français pour faciliter la compréhension.  <b>Accessibilité :</b> Plateforme disponible sur le web donc pas de type d'appareil obligatoire.	<b>Support Technique :</b> Assistance en ligne pour aider les utilisateurs avec des problèmes techniques ou questions sur l'utilisation de l'outil.	<b>Élèves et Étudiants :</b> Apprenants de différents niveaux scolaires, du primaire à l'université.  <b>Enseignants :</b> Utilisation de l'outil pour améliorer les méthodes d'enseignement et suivre les progrès des élèves.
			<b>Canaux de distribution</b>	<b>Écoles et Institutions</b>
			<b>Plateforme Web :</b> Site internet où les utilisateurs peuvent accéder aux ressources et activités.	<b>Éducatives :</b> Intégration de l'outil dans les programmes éducatifs pour renforcer l'apprentissage de l'anglais.
			<b>Partenariats Éducatifs :</b> Collaboration avec des écoles et centres de formation pour intégrer l'outil dans les programmes d'études.	<b>Parents :</b> Cherchent des outils supplémentaires pour aider leurs enfants à apprendre l'anglais.
			<b>Médias Sociaux :</b> Utilisation des réseaux sociaux pour la promotion et l'engagement des utilisateurs.	
			<b>Source de Revenus</b>	
			<b>Abonnements :</b> Modèle d'abonnement pour les écoles et les utilisateurs individuels, avec différents niveaux d'accès (gratuit, premium).	<b>Publicité :</b> Revenus provenant de la publicité intégrée pour les utilisateurs gratuits.
			<b>Vente de Contenus :</b> Vente de cours avancés et de contenus supplémentaires pour les utilisateurs premium.	
<b>Structure des coûts</b> <ul style="list-style-type: none"> <li><b>Développement Technologique :</b> Coûts de développement et de maintenance de la plateforme web.</li> <li><b>Création de Contenus :</b> Coûts liés à la création et à la mise à jour des ressources éducatives.</li> <li><b>Marketing et Ventes :</b> Dépenses pour les campagnes de promotion et l'acquisition de nouveaux utilisateurs.</li> <li><b>Support Client :</b> Coûts liés à l'assistance technique et au support utilisateur.</li> <li><b>Infrastructure :</b> Hébergement, serveurs et autres coûts technologiques.</li> </ul>				