République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Chahid Cheikh Larbi Tbessi – Tébessa

Faculté des Lettres et des Langues

Département de Lettres et Langue Françaises

Mémoire de Master

Filière : Lettres et Langue Française

Option : Science du langage

Analyse sociolinguistique de l'adaptation de ChatGPT aux variations linguistiques dans différentes situations d'échange conversationnel écrit avec des internautes

Réalisé par : **BADJI Houcem**

GHOUL Hicham

Membres de Jury :

Président(e) : Dre. HADJAR

Rapporteur(e) : Mme : BELAIDI

Examinateur : Dr. TALBI

2023/2024

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Chahid Cheikh Larbi Tbessi – Tébessa

Faculté des Lettres et des Langues

Département de Lettres et Langue Françaises

Mémoire de Master

Filière : Lettres et Langue Française

Option : Science du langage

Analyse sociolinguistique de l'adaptation de ChatGPT aux variations linguistiques dans différentes situations d'échange conversationnel écrit avec des internautes

Réalisé par : **BADJI Houcem**

GHOUL Hicham

Membre de Jury :

Président(e) : Dre. HADJAR

Rapporteur(e) : Mme : BELAIDI

Examinateur : Dr. TALBI

2023/2024

Remerciements

Nous tenons tout d'abord, à remercier notre Dieu le tout-puissant, de nous avoir donné la volonté et le courage de réaliser ce modeste travail.

Nous exprimons nos sincères remerciements à notre directrice de recherche Mme BELAIDI d'avoir accepté de diriger notre projet scientifique, pour son soutien et ses recommandations qui nous ont permis de bien mener cette recherche.

Nos chaleureux remerciements vont également aux membres du jury qui nous ont fait l'honneur de lire, d'évaluer ce mémoire et de participer à cette soutenance.

Nous adressons nos remerciements également à nos familles pour l'encouragement et le soutien inconditionnel qu'elles nous ont apporté.

Un grand merci à toutes les personnes qui nous ont aidées durant toute la période de ce travail, de près ou de loin.

Dédicaces

C'est avec une grande modestie et un immense plaisir que je dédie ce modeste travail à la mémoire de mes chers, père et frère

Que Dieu les accueille dans son vaste paradis

Avec tous mes sentiments d'amour et de respect,

À ma source de tendresse, de douceur et de bienveillance, ma chère mère,

À toute ma famille.

Ainsi qu'à tous mes camarades et amis

Et à tous ceux qui m'ont vivement encouragé à terminer ce mémoire.

Ghoul Hicham

Dédicaces

À la femme qui a toujours rêvé que je le réalise pour elle, ma chère maman.

À mon père, qui a toujours tenu à me voir parmi les hommes les plus élevés.

À mes frères et sœurs.

À mes amis et mes collègues, chacun par son nom.

À ceux qui ont parié que je n'arriverais pas.

A moi-même … Si tu veux, tu peux.

Badji Houcem

Résumé

Ce mémoire de recherche met l'accent sur les compétences interactionnelles de ChatGPT, dans des situations d'échange langagier avec des êtres humains, Il vise à étudier essentiellement, la faculté adaptative qui lui permet de s'intégrer efficacement dans ces conversations, et de s'ajuster face aux divers registres de la langue française.

En appliquant une étude quantitative et qualitative pour analyser les données collectées, la méthode mixte est le meilleur choix qui sert à réaliser cette analyse. Avec ces deux approches descriptive et analytique en s'appuyant sur les captures d'écran comme un moyen convenable de collecte de données.

Les résultats montrent que le registre courant est le plus dominant dans toutes les conversations, notamment les énoncés produits par ChatGPT, avec l'implication plus ou moins des deux autres registres, cela signifie que ChatGPT préfère l'usage d'un langage standard qui peut être compris par toutes les catégories des interlocuteurs. Tout en soulignant sa capacité de s'adapter de façon performante, avec les autres registres, en fonction, de diverses situations contextuelles.

L'importance de ce travail scientifique réside dans le fait que l'influence de ces logiciels interactifs est devenue, de plus en plus, marquante dans notre vie quotidienne, et la nécessité de les étudier pour mieux comprendre, leur fonctionnement devient indispensable.

Abstract:

This master graduation focuses on the interactional skills of ChatGPT, in situations of language exchange with human beings. It aims to essentially study the adaptive faculty which allows it to integrate effectively into these conversations, and to adjust to the various registers of the French language.

By applying a quantitative and qualitative study to analyze the data collected, the mixed method is the best choice used to carry out this analysis. With both descriptive and analytical approaches relying on screenshots as a suitable means of data collection. The results show that the current register is the most dominant in all conversations, notably the utterances produced by ChatGPT, with the more or less involvement of other registers, this means that ChatGPT prefers the use of a standard language which can be understood by all categories of interlocutors. While emphasizing these abilities, to adapt effectively, with others registers, depending on various contextual situations.

The importance of this scientific work lies in the fact that the influence of this interactive software has become more and more significant in our daily lives, and the need to study it to better understand it becomes essential.

خلاصة :

تركز مذكرة الماستر هذه على المهارات التفاعلية لـ**ChatGPT** أثناء تبادل أطراف الحديث مع بعض مستعملي الانترنت من الناس، وتهدف بشكل أساسي إلى دراسة القدرة التكيفية التي تسمح له بالاندماج بشكل فعال في هذه المحادثات، والتكيف مع الأساليب اللغوية المختلفة للغة الفرنسية.

ومن خلال تطبيق دراسة كمية ونوعية من اجل تحليل البيانات التي تم جمعها، فإن الطريقة المزدوجة **(la méthode mixte)** بمنهجيها الوصفي والتحليلي هي الخيار الأفضل لإجراء هذا النوع من التحليل, معتمدين في ذلك على لقطات الشاشة (**captures d'écran**) كوسيلة مثلى لجمع البيانات، حيث أظهرت النتائج أن الأسلوب اللغوي الصحيح (registre courant)هو الأكثر استخداما في جميع المحادثات، ولا سيما الخطاب الذي ينتجه **ChatGPT** ، مع إدراج الأساليب الأخرى**(registre soutenu, registre familier)** إلى حد ما. وهذا يعني أن **ChatGPT** يفضل استخدام لغة قياسية يمكن فهمها من طرف جميع فئات المحاورين. مع التأكيد على قدرته على التكيف بشكل فعال، مع الأساليب اللغوية المشار إليها سابقا، اعتمادا على ما يقتضيه سياق المحادثة.

تكمن أهمية هذا البحث العلمي في أن تأثير مثل هذه البرنامج التفاعلي, أصبح ذا أهمية متزايدة في حياتنا اليومية، وأن الحاجة لدراستها لفهمها بشكل أفضل أصبحت أمرًا ضروريًا.

Mots-clés : traitement automatique du langage, variations linguistiques, registres de langue, ChatGPT, intelligence artificielle.

Table des matières :

Introduction générale.....11

La partie théorique : Concepts théoriques.....16

Chapitre I : La sociolinguistique et ses variations (registres de langue).....17

I.1-Définition de la sociolinguistique 18

I.2- La Communauté linguistique 19

I.3- La variation linguistique 20

I.3.1- Définition 20

I.3.2-Les types de la variation linguistique 21

I.3.2.1- La variation diachronique (historique).....21

I.3.2.2- La variation diatopique (géographique). …..…….23

I.3.2.3- La variation diastratique (sociale)....…….……..23

I.3.2.4- La variation diaphasique (stylistique).…...…….……...…24

I.3.2.4.1- *Définition*……...……24

I.3.2.4.2- *Les registres de langue*…….…….24

I.3.2.4.2.1-Définition…….…….24

I.3.2.4.2.2-Les trois principaux registres de la langue française……..…….25

I.3.2.4.2.2.1-Le registre familier……...……26

I.3.2.4.2.2.2-Le registre courant..…….……27

I.3.2.4.2.2.3-Le registre soutenu...……29

I.4- La variété linguistique 30

I.4.1- Définition 30

I.4.2- les types de variété linguistique 30

I.4.2.1-Les variétés géographique(régionales)..…..……32

I.4.2.2-Les variétés sociales……..……33

I.4.2.3-Les variétés historique...…….……..33

I.5- Distinguer la variation linguistique de la variété linguistique :..... 32

Chapitre II : Présentation du ChatGPT..... 33

II.1. Linguistique cognitive 34

II.1.1- La théorie de Chomsky 35

II.1.2- La métaphore de l'ordinateur 36

II.1.3- La Métaphore de l'Automate 37

II.2- La linguistique informatique (LI) 38

II.2.1- Définition de la Linguistique Informatique (LI) 39

II.2.2- Le fonctionnement de la Linguistique Informatique (LI) 39

II.2.3- Traitement Automatique des Langues (TAL) 39

II.3- L'Intelligence Artificielle (IA) 40

II.3.1- Définition de l'Intelligence Artificielle (IA) 40

II.3.2- L'apprentissage profond (Deep Learnig) 41

II.3.3- Les Réseaux de Neurones 42

II.4- Les Chatbots conversationnels 43

II.4.1- L'Historique de Chatbots 43

II.4.2- ChatGPT **44**

II.4.2.1-OpenAI……46

II.4.2.2-Définition de ChatGPT……..47

II.4.2.3-Fonctionnement de ChatGPT……..…….48

La Partie Pratique.....**48**

Chapitre III : Choix méthodologique et analyse des données **49**

III.1- Choix méthodologique **50**

III.1.1- la nature de l'étude **50**

III.1.2- La méthode d'analyse **51**

III.1.3-Présentation du corpus……..…….....51

III.2- Étude quantitative des données et vérification de l'hypothèse n°1 .....**52**

III.2.1- Analyse des conversations **52**

*III.2.2-Discussion et interprétation de résultats obtenus par approchequantitative..***64**

III.2.3- vérification de la première hypothèse **66**

III.3- Étude qualitative des données et vérification de l'hypothèse n° 2 .....**67**

III.3.1-Analyse des conversations **67**

III.3.2-Discussion et interprétation de résultats obtenus par approche qualitative ..**83**

III.3.3- vérification de la deuxième hypothèse **84**

Conclusion générale **84**

Références bibliographiques.....**88**

Annexes :.....**…91**

Introduction générale

Analyse sociolinguistique de l'adaptation de ChatGPT aux variations linguistiques dans différentes situations d'échange conversationnel écrit avec des internautes

Notre sujet s'inscrit dans le cadre d'une étude sociolinguistique, qui met l'accent sur la capacité de ChatGPT, comme un modèle de l'intelligence artificielle capable de communiquer avec l'homme, à s'adapter aux variations linguistiques de la langue française. Notamment à la variation diaphasique, l'un de ses quatre types. Cette étude s'intéresse essentiellement, à examiner l'impact de l'usage varié des registres de langue sur le comportement linguistique de ChatGPT dans des situations variées de l'échange conversationnel avec des internautes.

La croissance rapide de la science et de la technologie, en particulier dans le domaine en pleine expansion de l'intelligence artificielle, a vu l'émergence de logiciels intelligents tels que ChatGPT. Ces outils de nouvelle génération sont désormais en mesure d'engager indépendamment des dialogues avec les êtres humains et d'accomplir une multitude de tâches complexes. Face à ces avancées, il devient primordial de s'interroger sur le fonctionnement du langage naturel chez ChatGPT, et particulièrement sur la capacité d'adaptation de ces systèmes aux diverses variations diaphasiques (registres de langue).

Cette compétence adaptative peut avoir une importance capitale, notamment lors des situations de communication impliquant des interlocuteurs provenant de contextes socioculturels hétérogènes. En effet, la maîtrise des différents registres de langue et la faculté à s'ajuster au parler de chacun, sont des prérequis indispensables à l'établissement d'une communication fluide et efficace.

Évaluer la capacité interactionnelle de ChatGPT avec l'humain, c'est savoir si cette compétence, lui permet de prendre en considération, les différentes particularités linguistiques, culturelles et contextuelles propres à chaque langue.

Afin d'étudier ces compétences linguistiques machinales, des recherches ont été menées par des spécialistes dans les domaines concernés, nous citons parmi eux :

Franck TB (2006) dans sa thèse de doctorat estime que l'ingénierie de l'interaction homme-machine doit se concentrer sur l'adaptation des interfaces homme-machine au contexte d'utilisation dans toutes ses dimensions. Il propose un cadre théorique modélisant l'interaction adaptative, en présentant de nouvelles techniques d'adaptation.

Nadjar Insaf (2020), propose dans son mémoire de master une nouvelle approche qui soutient l'interaction homme-machine à travers des systèmes normatifs multi-agents (SMA) qui vise à établir des interactions intelligentes entre l'homme et la machine. Le SMA sert à

interpréter les échanges conversationnels avec les utilisateurs, et de réadapter l'interaction homme-machine (IHM) à leurs comportements interactionnels.

Dans un article scientifique Ana, PC avec ses collaborateurs (2021), établissent un cadre pour analyser l'effet des choix linguistiques sur les perceptions des utilisateurs et prend une première étape vers le développement d'une base prescriptive pour adapter les choix linguistiques des chatbots à des situations interactionnelles spécifiques. Cette recherche analyse comment les différences de registre jouent un rôle dans la perception de l'utilisateur de l'interaction homme-chatbot. Les auteurs ont produit des corpus parallèles de conversations dans le domaine du tourisme avec un contenu similaire et des caractéristiques de registre variables, puis ont évalué les préférences des utilisateurs concernant les choix linguistiques des chatbots en termes d'adéquation, de crédibilité et d'expérience utilisateur.

Les résultats montrent que les caractéristiques de registre sont de forts prédicteurs des préférences des utilisateurs, soulignant ainsi la nécessité de concevoir des chatbots avec un langage approprié au registre pour améliorer l'acceptation et les perceptions des utilisateurs des interactions avec les chatbots.

À la lumière de cette réflexion, et à partir de ce qui précède, nous tenterons de répondre à la question principale suivante :

- Dans quelle mesure ChatGPT s'adapte-t-il aux variations diaphasiques (registre de langue) ?

Cette question centrale peut être explorée à travers deux questions secondaires complémentaires :

1- Quels sont les aspects de la variation diaphasique (les registres de langue) dans des conversations écrites entre ChatGPT et des internautes ?
2- Quel est l'impact de l'interaction avec des internautes sur la faculté adaptative de ChatGPT aux variations diaphasiques ?

Afin de répondre aux questions précédentes nous allons proposer deux hypothèses, qui constituent des réponses provisoires, comme suit :

1- ChatGPT communiquerait avec divers internautes en utilisant une langue standard.
2- ChatGPT interagirait efficacement face aux différents registres de langue.

La finalité de cette étude scientifique, consiste d'un côté à comprendre en profondeur, comment ChatGPT s'adapte aux variations linguistiques (diaphasiques), notamment en ce qui concerne, son interaction avec les diverses implications de différents registres de langue, dans divers échanges linguistiques ou dans une même conversation, et d'un autre côté, à analyser de manière critique les interactions de ChatGPT avec les internautes, afin de mesurer ses

compétences communicationnelles, et d'évaluer son efficacité dans divers contextes d'échange conversationnel,

La justification de ce choix vient de notre curiosité scientifique motivée par une volonté insistante d'approfondir nos compétences en recherche dans deux domaines liés, d'une part, les sciences du langage, et leurs applications dans de nombreux secteurs de recherche, et d'une autre part, l'intelligence artificielle, surtout les chatbots conversationnels, de plus, nous souhaitons contribuer à l'avancement des connaissances sur l'adaptation de ChatGPT aux registres de langue en apportant une perspective personnelle et critique à cette étude.

En appliquant une approche hypothético-déductive, qui consiste à valider nos deux hypothèses par l'observation des faits linguistiques qui composent le corpus de ce mémoire. L'analyse scientifique de notre projet de recherche sera basée sur une étude quantitative et qualitative à la fois, qui comporte une comparaison entre les productions langagières des deux interlocuteurs. Nous nous appuierons sur la méthode mixte qui combine deux parties, descriptive et analytique.

L'étude quantitative aura pour but, d'analyser des statistiques des données linguistiques, en recensant le nombre du vocabulaire propre à chaque registre, ainsi que calculer le taux d'utilisation de différents registres de langue par les deux interlocuteurs dans chaque conversation. Cette étude quantitative vise à déterminer d'une manière précise le registre dominant de chacune des conversations qui constituent notre corpus.

Tandis que l'étude qualitative, aura pour objet, de décrire les caractéristiques syntaxiques, lexicales et stylistiques des données linguistiques qui caractérisent chaque registre. En mesurant l'impact de ces variables sur le comportement linguistique de ChatGPT, (le sens des mots, les constructions syntaxiques, etc.). Elle vise surtout à évaluer ses capacités interactionnelles en comparaison avec des interlocuteurs humains.

Afin de mener cette recherche, le corpus de cette étude sera fondé sur un recueil de six conversations variées, soit par le niveau du sujet de la discussion, soit par le type de registre utilisé. Entre d'un côté des interlocuteurs humains, appelés tous (You) et d'un autre côté ChatGPT version 3,5. Pour ce faire, nous allons choisir les captures d'écran comme un instrument pertinent de collecte des données.

Notre mémoire de recherche sera divisé en deux parties essentielles, une première partie théorique et une seconde partie pratique :

La partie théorique comportera une introduction générale et deux chapitres : Dans l'introduction générale nous allons présenter le sujet du mémoire, la problématique, la question centrale, les questions secondaires, et les hypothèses,

Le premier chapitre sera consacré aux définitions des principaux concepts tels que, la définition de la sociolinguistique, la variation linguistique, la différence entre celle-ci et la variété linguistique, les types de ces variations, et enfin l'étude plus ou moins détaillée des registres de langue

Le deuxième chapitre abordera, si nous pouvons dire, l'aspect technique de ce mémoire c'est-à-dire tout ce qui concerne les automates, et le traitement automatique du langage ainsi que des théories scientifiques dans le même contexte, comme la théorie de la grammaire générative et transformationnelle de Noam Chomsky. En terminerons par l'émergence de ChatGPT et son fonctionnement, en s'appuyant sur des figures illustratives.

Le troisième chapitre de la deuxième partie comportera essentiellement l'analyse statistique et linguistique des données, accompagnées également par des tableaux et des diagrammes explicatifs. Nous conclurons ce chapitre par la discussion et l'interprétation des résultats obtenus et la vérification des deux hypothèses, puis par une conclusion générale.

La partie théorique

Concepts théoriques

Chapitre I

La sociolinguistique et ces variations

(Registres de langues)

La sociolinguistique s'est caractérisée par une réaction marquée contre le structuralisme linguistique à partir de la fin des années soixante, et qui se concentrait essentiellement sur la structure interne des langues en négligeant leur utilisation dans un contexte social, dans ce sens Labov (1976) déclare qu'ils « S'obstinent à rendre compte des faits linguistiques par d'autres faits linguistiques, et refusent toute explication fondée sur des données extérieures tirées du comportement social » (p.259).

Ce nouveau domaine d'étude a été développé par un groupe de chercheurs influents tels que (Dell Hymes, Fishman, Gumperz, Labov, Ferguson, etc.) leur approche peut se résumer comme suit : « Étudier qui parle quoi, comment, où et à qui ». (Fishman, 1971, pp. 286-299)

La sociolinguistique a placé le rapport social entre les individus au cœur de ses préoccupations, mettant ainsi en lumière l'importance de considérer la langue dans son contexte social et culturel, ainsi que l'impact des facteurs sociaux sur la variation linguistique. Autrement dit : « La sociolinguistique prend en compte tous les phénomènes liés à l'homme parlant au sein d'une société » (Boyer, 1996, p. 36)

Ce champ disciplinaire est devenu un domaine important et actif qui a grandement contribué au renouvellement de nos catégories linguistiques, notamment à travers le développement de la linguistique de contact. Ainsi, les langues, jadis perçues comme des systèmes autonomes, sont désormais envisagées comme des systèmes fluides et variables, etc.

I.1-Définition de la sociolinguistique :

La sociolinguistique est une branche des sciences du langage, il s'agit en effet, d'un domaine d'étude qui représente simplement un aspect de la linguistique elle-même. Pour William Labov, l'un des fondateurs de la discipline, ainsi pour d'autres, le facteur social est un élément primordial, pour mieux comprendre les phénomènes linguistiques. Ce qui marque une rupture avec les linguistes qui suivent la tradition saussurienne, telle que décrite dans le Cours de linguistique générale de Ferdinand de Saussure.

Il critique ainsi, ces linguistes pour leur insistance à expliquer les phénomènes linguistiques uniquement par d'autres phénomènes linguistiques, en rejetant toute explication fondée sur des données sociales extérieures, ce qui contredit Boyer (1996), en déclarant que : « La sociolinguistique prend en compte tous les phénomènes liés à l'homme parlant au sein d'une Société. » (p.36)

Pour résumer, il est important de souligner que, cette discipline, d'après Labov et d'autres linguistes, prend en compte tous les aspects des pratiques langagières humaines au

sein d'une société sociolinguistique. Et que la naissance de la sociolinguistique s'est réalisée dans un contexte de critique envers les orientations théoriques et méthodologiques de la linguistique structurale, mettant en avant l'importance d'incorporer des données sociales dans l'étude du langage.

I.2- La Communauté linguistique :

Une communauté linguistique est formée par un groupe d'êtres humains, qui utilisent la même langue ou dialecte, et peuvent communiquer entre eux. Elle se caractérise par une communication constante entre ses membres en partageant les mêmes normes linguistiques. L'importance de l'échange implique que les membres se communiquent plus entre eux qu'avec des étrangers, ce qui sert à créer des frontières marquées dans les échanges conversationnels. Les normes partagées se réfèrent à un ensemble de jugements évaluatifs et à une compréhension commune de ce qui est estimé comme approprié dans diverses situations sociales.

Cela signifie qu'une communauté linguistique se forme dès lors que ses membres partagent au moins une variété linguistique, commune ainsi que les règles pour son usage approprié. Fishman et Labov proposent deux définitions différentes, pour la première : « Une communauté née d'une communication intensive et/ou d'une intégration symbolique en relation avec la possibilité de communication, sans tenir compte du nombre de langues ou de variétés employées. » (Fishman, 1971, pp. 46,47).

Et pour la deuxième : « La communauté linguistique se définit moins par un accord explicite quant à l'emploi des éléments du langage que par une participation conjointe à un ensemble de normes » (Labov, 1976, p. 187).

Les deux définitions partagent deux caractéristiques clés : l'intensité des interactions et le partage de normes communes. En termes, d'intensité de communication, cela signifie essentiellement que les membres d'une communauté linguistique ont plus d'interactions entre eux qu'avec des étrangers. Ainsi, les frontières entre les communautés linguistiques se traduisent par des zones moins actives ou des lacunes dans le schéma de communication.

Finalement, une communauté linguistique existe dès lors que tous ses membres partagent les mêmes pratiques linguistiques communes, de ce fait, les normes régissant son utilisation correcte.

Une communauté linguistique peut, donc, se limiter à un groupe de personnes qui choisissent de vivre ensemble, comme c'est le cas chez la communauté Kabyle.

I.3- La variation linguistique :

I.3.1- Définition :

Dans le domaine de la sociolinguistique, la variation linguistique se réfère aux différences observées, en fonction, de divers facteurs internes ou externes. Cette notion a été conceptualisée à partir des premières recherches de William Labov entre 1966 et 1972, mettant en lumière comment des significations similaires peuvent être exprimées à travers des formes linguistiques distinctes. En effet, les locuteurs ne recourent pas uniformément aux mêmes expressions linguistiques. Selon le contexte, la situation de communication ou l'intention, un locuteur adopte un style spécifique, utilisant, par conséquence, une variété linguistique, une variante ou un registre particulier. Cette diversité linguistique découle des conditions de vie des interlocuteurs, des intentions communicatives, des circonstances de production et de réception du message, de l'environnement social, géographique et historique du locuteur, ainsi que du registre linguistique visé. Dans ce contexte Labov (1976), souligne :

> Qu'il n'est pas de langue que ses locuteurs ne manient sous des formes diversifiées « … » les sociolinguistes saisissent cette différenciation en parlant de variétés pour désigner différentes façons de parler, et de variation pour les phénomènes diversifiés en synchronie, et de changement pour la dynamique en diachronie. (p.259)

Pour les sociolinguistes, chaque langue est manipulée sous des formes diverses par ses locuteurs. Ces spécialistes décrivent ces différences en utilisant les termes de "variétés" pour désigner les différents modes de parole synchroniques, de "variation" pour les phénomènes linguistiques variés dans le temps présent, et de "changement" pour la dynamique diachronique. La langue n'est donc pas homogène, mais transmet des variations qui sont influencées par les groupes d'utilisateurs et qui suivent certaines régularités. Comme l'a souligné Moreau, (1997, p. 283) : « Aucune langue ne se présente comme un ensemble unique de règles, toutes présentent une multitude de variantes ou de "lectes", masquées par des étiquettes au singulier (comme "le français", "le turc", etc.), même si ces désignations sont abstraites et réductrices. »

I.3.2-Les types de la variation linguistique :

En outre des travaux de William Labov, considéré comme le pionnier de l'approche variationniste en sociolinguistique, Moreau, ajoute que toutes les langues du monde montrent des variations, ce qui signifie qu'elles ne sont pas caractérisées par un ensemble unique et stable de règles linguistiques.(1997, p.236). Les locuteurs au sein d'une même communauté linguistique, n'ont pas nécessairement les mêmes pratiques linguistiques en tout temps, en distinguant quatre principaux types de variation linguistique.

- La variation diachronique (ou historique)
- La variation diatopique (ou géographique)
- La variation diastratique (ou sociale)
- La variation diaphasique (ou stylistique)

I.3.2.1- la variation diachronique (historique). Nous nous situons dans une perspective historique, car nous abordons ici l'évolution de la langue à travers les différentes époques qu'elle traverse. Toutes les langues subissent des changements, qu'ils soient brusques ou imperceptibles, touchant les aspects phonétiques, morphosyntaxiques, lexicaux ou sémantiques. Les facteurs temporels se manifestent dans des situations courantes où un locuteur plus âgé ou plus jeune peut percevoir des différences dans le langage de l'autre. La variation diachronique est liée au temps, permettant de contraster les traits linguistiques en fonction de leur perception comme étant plus anciens ou plus récents. En d'autre terme, la « variation diachronique est liée au temps ; elle permet de contraster les traits selon qu'ils sont perçus comme plus ou moins anciens ou récents » (Moureau, 1997, p. 284).

Pour finir, La langue est un système vivant en constante évolution qui change au fil du temps. Nous observons ainsi l'émergence de nouveaux mots, la disparition totale de certains autres, ainsi que l'acquisition de nouveaux sens ou le développement d'archaïsmes. Ce processus de transformation constitue le changement que peut subir la langue. Lorsque nous examinons les traits linguistiques anciens et récents, nous abordons ce qui est appelé la variation diachronique.

I.3.2.2- La variation diatopique (géographique). Ce que nous appelons également variation régionale, se situe sur un axe géographique, où la langue se diversifie selon les différents usages observés d'une région à une autre. En d'autres termes, cela concerne les particularismes régionaux au sein d'une même langue. C'est-à-dire la « diversité des usages à l'intérieur d'une aire linguistique géographiquement circonscrite » (Moureau,

1997, p. 236). Ce qui donne naissance à ce que nous désignons par des dialectes, des régiolectes ou des topolectes.

Les facteurs spatiaux nous montrent que la manière de parler, peut varier non seulement d'un endroit à un autre sur un territoire donné, mais aussi que chaque lieu porteur d'identité peut engendrer ces mêmes variations, (comme les quartiers d'une ville). La variation diatopique opère sur l'axe géographique, elle consiste en la différenciation de la langue selon les régions. Les usages qui en découlent sont désignés par des termes tels que régiolectes, topolectes ou géolectes. À ce propos Moreau, (1997, p.284) indique que : « La variation diatopique joue sur l'axe géographique ; la différenciation d'une langue suivant les régions relève de cette variation, pour désigner les usages qui en résultent, on parle de régiolectes, de topolectes ou de géolectes.»

I.3.2.3-La variation diastratique (sociale). Nous abordons ici un axe social, où nous observons les différences dans les usages linguistiques des locuteurs, en fonction des classes sociales auxquelles ils appartiennent.

La variation diastratique décrit ces différences d'usage entre les différentes classes sociales. Nous parlons alors de sociolectes pour désigner ces variations linguistiques liées au statut social des locuteurs. Ce qui peut être traduit par la citation suivante : « La variation diastratique explique les différences entre les usages pratiqués par les diverses classes sociales, il est question en ce cas de sociolectes. » (Moureau, 1997, p. 284).

Une langue donnée peut présenter une diversité de réalisations étroitement liée aux différentes couches sociales. Chaque individu appartient à une classe sociale spécifique et est donc enclin à utiliser sa propre variété linguistique. Ce qui explique que :

> Les linguistes étudient la variation sociale et les dialectes sociaux, ou sociolecte, qui naissent des inégalités dans la société. Les individus qui composent une société ne sont pas égaux : Le général et le bidasse, le proviseur d'un lycée et le maitre auxiliaire, l'enfant prodige et l'enfant handicapé moteur. (Baylon & Mignot, 1991, p. 227)

Cela implique que la variation diastratique est particulièrement marquée entre le discours d'un ouvrier et celui d'un médecin, amplifiant ainsi la représentation de leur identité sociale et de leur statut respectif.

I.3.2.4- La variation diaphasique (stylistique) :

I.3.2.4.1-Définition. Ce que nous pouvons appeler également variation situationnelle, ne divise pas la société elle-même, mais plutôt le locuteur qui, selon les différentes situations de communication auxquelles il est confronté, utilise différents styles ou registres de la même langue. La variation diaphasique désigne cette différenciation des usages linguistiques en fonction des situations de discours.

Ainsi, la manière dont la langue est utilisée est influencée par le degré de formalité du contexte d'énonciation, ce qui se traduit par l'emploi de registres ou de styles différents. À cet égards Moreau (1997), déclare qu'«on parle de variation diaphasique lorsqu'on observe une différenciation des usages selon les situations de discours ; ainsi la production langagière est-elle influencée par le caractère plus ou moins formel du contexte d'énonciation et se coule-t-elle en des registres ou des styles différents.»(P. 284).

En règle générale, nous employons différents termes pour décrire cette variation, tels que le « registre linguistique » ou le « style de langage ». Il est également important de prendre en compte d'autres facteurs qui contribuent à expliquer ce phénomène, notamment l'âge, le sexe, la profession, la religion, et ainsi de suite.

I.3.2.4.2-Les registres de langue :

I.3.2.4.2.1-Définition. Nous adaptons notre langage selon notre interlocuteur et selon le contexte. Ainsi, nous utilisons différents registres de langue pour exprimer nos idées, nos sentiments, etc.

Un registre de langue, nous disons aussi niveau de langue ou style, représente une manière d'exprimer adaptée à une situation d'énonciation spécifique. Il influence notamment les choix de vocabulaire et de structure grammaticale, ainsi que l'attitude générale adoptée dans la communication.

Il est généralement accepté qu'une langue n'est pas homogène, mais plutôt composée de divers sous-systèmes, comme l'ont souligné (Baylon & Fabre, 1990, p. 148) :« C'est à la stylistique de les démêler et de rendre compte de leur utilisation. » La stylistique s'emploie à identifier ces éléments et à en rendre compte dans leur utilisation.

Ces chercheurs notent également que les langues reflètent des distinctions sociales et professionnelles, classées par les sociolinguistes sous le terme de "registres de langue" ou "registres de discours", en particulier pour décrire le degré de formalité entre interlocuteurs.

Cuq (2003) définit le registre de langue comme l'adaptation de la langue en fonction des situations de communication. Il s'agit pour lui de« l'utilisation de la langue en fonction de diverses situations de communication » (p.78). Le mode d'expression varie en fonction des caractéristiques socioculturelles telles que l'âge, le niveau d'instruction, l'activité professionnelle et le milieu social de la personne qui communique, ainsi que les circonstances de la communication, qu'elle soit orale, écrite, solennelle, ou décontractée. La façon de parler diffère selon à qui nous nous adressons, à un enfant ou à un adulte, à un ami ou à un inconnu, à un parent ou à un supérieur hiérarchique.

Selon l'auteur (Lubinda, s. d.). En sociolinguistique française, ces variations sont désignées sous divers termes tels que "registre de discours", "registre de langue", "registre de parole", "registre d'expression", ou "niveau de langue". L'appellation préférée aujourd'hui est "registre de discours" ou "registre de langue", car elle est moins chargée de connotations que "niveau de langue". Il est à noter que ces registres ne sont pas hiérarchisés entre eux.

Ces registres se différencient notamment par la richesse du vocabulaire, la complexité de la syntaxe et la qualité globale de l'expression utilisée…etc.

Note : Il arrive que certains mots ou expressions appartiennent à un registre particulier uniquement dans certains contextes d'utilisation.

- Info : Différentes marques d'usage peuvent se combiner pour préciser l'emploi d'un mot ou d'une expression.
- HISTRION : nom masculin PÉJORATIF, LITTÉRAIRE Comédien.
- BONTÉ : nom féminin

  1. Qualité morale qui porte à faire le bien, à être bon pour les autres. […] Elle est d'une grande bonté. — interj. Bonté divine !

  2. Amabilité, gentillesse. Voulez-vous avoir la bonté de…. (Robert, s. d.)

I.3.2.4.2.2-Les trois principaux registres de langue. La langue française se caractérise par l'emploi de trois registres principaux : le registre familier, le registre courant et le registre soutenu.

I.3.2.4.2.2.1-Le registre familier. Le registre familier est habituellement utilisé dans le contexte oral avec des proches tels que des parents ou des amis. Ce registre fait usage de termes familiers et d'abréviations courantes (comme "télé" pour "télévision"). Les phrases sont souvent simplifiées et construites de manière décontractée. Selon le Robert (s. d.), appartiennent au registre familier des mots, expressions et sens, relevant essentiellement de

l'oral, mais aussi de l'écrit dans la langue quotidienne notamment entre amis, parents ou réseaux sociaux...et qui ne s'emploieraient pas dans des circonstances solennelles. Dans ce sens, Léon, (1971) déclare que ce registre de langue « reflète un état spontané, détendu et une absence d'auto-contrôle ». Plus spécifiquement, il s'agit du français utilisé dans les conversations de rue, au bistrot du coin, lors des discussions décontractées en terrasse de café entre amis ou collègues.

Généralement, le style familier est utilisé naturellement avec des personnes que nous connaissons bien et avec qui nous tutoyons. Ce registre est donc à éviter dans les relations avec des supérieurs, les relations officielles et les ouvrages qui se veulent sérieux (hors effet stylistique particulier).

Note : « Les mots ou expressions signalés comme « très familiers » sont parfois péjoratifs, ils ne peuvent être employés, même entre personnes de connaissance, sans risquer de choquer. (Robert, s. d.) »

Le registre familier utilise :

- Une syntaxe simplifiée et souvent approximative, des phrases courtes, parfois inachevées, ou au contraire, interminables ; des phrases nominales, souvent asyntaxiques (anacoluthes, thématisations diverses et parfois multiples dans une même phrase ; des interjections fréquentes ; un grand usage de l'ellipse ; des pléonasmes
- L'utilisation de la juxtaposition paratactique : Au bureau, un de mes collègues, sa femme, elle a eu un bébé. Pour : « La femme d'un collègue du bureau a eu un bébé. »
- De nombreuses abréviations non encore lexicalisées : T'es là ? / phone / p'tit déj' / une deuch'…Pour : « Tu es là ? / téléphone / petit déjeuner / une deux chevaux… »
- La forme interrogative directe (par changement intonatif, sans inversion, ni mot interrogatif) : Tu m'appelles d'où ? Pour : « D'où m'appelles-tu ? »
- Un vocabulaire familier, parfois chargé de nuances affectives ou sociales diverses :

Les guibolles / la frimousse ou la gueule / les quenottes… pour : « Les jambes / le visage /les dents… »

- La suppression de ne dans la négation : J'ai pas bien dormi cette nuit. Pour : « Je n'ai pas bien dormi cette nuit. »
- Le pronom sujet «on » à la place de « nous » : Nous, on viendra. Pour : « Nous, nous viendrons. » (Registres des langue , s. d.)

I.3.2.4.2.2.2-Le registre courant. Le registre courant est utilisé dans la vie quotidienne. Il correspond à un langage neutre et correct, mais il n'est pas particulièrement recherché ou élaboré.

D'après le dictionnaire le Robert (s. d.), « Appartiennent au registre courant (ou registre standard) les mots, les expressions et les sens censés être connus et employés par le plus grand nombre. », c'est la langue employée par les journalistes à la radio et dans la presse écrite, lors des conversations avec des collègues de bureau, des entretiens d'embauche et dans le cadre scolaire. C'est le français de référence qui correspond aux normes grammaticales du langage correct à l'oral et à l'écrit.

Ce français est largement compris par tous ceux qui pratiquent couramment la langue, ce qui, en fait, un outil de communication international. Dans ce registre, le locuteur ou l'écrivain s'efforce de respecter rigoureusement les règles de la langue. Ce registre est souvent présenté en exemple dans les manuels scolaires destinés aux apprenants non francophones. À ce sujet le registre courant (le français standard) se définit tel que :

> La variété du français qui est utilisée par les Français bien instruits, quand ils surveillent bien leur façon de s'exprimer ; on l'associe au « bon usage » parisien du français. C'est la variété de la langue qui est prescrite par le gouvernement et par beaucoup d'organismes nationaux et internationaux ; c'est la variété utilisée par les médias et dans l'éducation, et c'est la variété préférée de la plupart des locuteurs francophones. Bref, c'est la variété du français officiellement approuvée par l'État et qui jouit du plus grand prestige dans le monde francophone. Elle est, dans une large mesure, codifiée dans les dictionnaires et dans les manuels de grammaire. Le français standard peut apparaître soit à l'oral soit à l'écrit, mais surtout à l'écrit. La permanence de l'écrit encourage naturellement une attention plus soutenue à cette variété de langue. (Offord, 1990, pp. 16,17)

Les formes et le vocabulaire du registre courant oral sont généralement admis à l'écrit.

Le registre courant servira de repère afin d'évaluer le niveau soutenu et le niveau familier, lequel est parfois désigné par le terme de colloquialisme, calqué de l'anglais.

Le terme jargon ne désigne pas un registre particulier, plutôt le vocabulaire particulier attaché à une communauté particulière, un milieu professionnel, politique, sportif, etc.

Exemples des termes :

- Baffle : nom masculin (ANGLICISME)

  1. TECHNIQUE Panneau portant le diffuseur d'un haut-parleur.

  2. COURANT Enceinte acoustique (recommandation officielle). Les baffles d'une chaîne hi-fi.

- Branchage : nom masculin

  1. RARE ensemble des branches d'un arbre. ramure.

  2. COURANT, AU PLURIEL Branches coupées. Sol jonché de branchages.

- A contrario : locution adjective invariable et locution adverbiale

  • DIDACTIQUE Se dit d'un raisonnement qui, partant d'hypothèses opposées, aboutit à des conséquences opposées.(Robert, s. d.)

I.3.2.4.2.2.3-Le registre soutenu. Est surtout celui des œuvres littéraires et religieuses, le registre soutenu est utilisé dans des contextes écrits ou dans des situations de communication formelle où recherchant un effet stylistique particulier. « Appartiennent au registre soutenu des mots, expressions et sens rares, dont l'emploi suppose une connaissance approfondie du lexique. » (Robert, s. d.), c'est le registre de l'écrit, des situations de communication contraintes ou de recherche d'effet stylistique. Le dictionnaire précise généralement le mode d'expression : littéraire ou poétique

Selon la spécialiste de la communication écrite (Natacha, 2023), dans son article (les registre de langue), elle estime quel registre soutenu est considéré comme le langage de l'écrit, caractérisé par son niveau de sophistication et de raffinement, et se trouve principalement dans les textes littéraires ainsi que les documents officiels ou protocolaires. Bien qu'il soit peu utilisé à l'oral, il peut être employé lorsque l'on s'adresse à une personne importante ou dans des milieux socioculturels élevés.

Le vocabulaire utilisé est rare, précis et recherché, accompagné d'une syntaxe très élaborée, ce qui nécessite une connaissance approfondie de la langue, de la littérature et de la culture. Les phrases, souvent longues, sont construites avec des temps comme le passé simple et l'imparfait du subjonctif, ainsi que de nombreuses propositions subordonnées et des figures de style, rendant la structure plus complexe.

Exemples des termes :

- Diaprure : nom féminin (LITTÉRAIRE) ; Aspect de ce qui est diapré, de ce qui chatoie.
- Différer :

1. Verbe transitif ; Remettre à un autre temps ; éloigner la réalisation de (qqch.). […] . Différer un paiement, une réponse.
2. (LITTÉRAIRE) Partez sans différer. […] — au participe passé : Rencontre différée.
3. Verbe intransitif ; Être différent, dissemblable…

- Aquilon : nom masculin

  • POÉTIQUE Vent du nord, froid et violent.
- S'enfuir : verbe pronominal

  1. S'éloigner en fuyant, ou en hâte…

  2. (POÉTIQUE) S'écouler rapidement. disparaître. Le temps s'enfuit. passer. (Robert, s. d.)

<u>le registre soutenu utilise :</u>

- Des phrases pouvant être longues (alors appelées « périodes »), à la syntaxe souvent complexe :
- Un vocabulaire rare : Le firmament / les cieux / l'azur…pour « le ciel »
- La pasigraphie du scribomane malévole formait une étrange octoade.
- Des figures de style recherchées :

Exemple : Déjà la nuit en son parc amassait/Un grand troupeau d'étoiles vagabondes. (Du Bellay), Pour : « Déjà la nuit tombait et on apercevait les premières étoiles. » (Métaphore filée)

- L'imparfait et le plus-que-parfait du subjonctif (à l'oral comme à l'écrit).

Exemple : Il fallait qu'il vînt. Pour : « Il fallait qu'il vienne. »

- Le passé simple et le passé antérieur de l'indicatif (à l'oral).

Exemple : Je le vis quand je revins. Pour : « Je l'ai vu quand je suis revenu. »

- La forme interrogative directe inversée.

Exemple : D'où m'appelles-tu ? Pour : « D'où est-ce que tu m'appelles ? »

- L'inversion du sujet après certains adverbes de liaison (tels que : aussi, ainsi, peut-être, etc.). Ainsi ai-je dû écourter mes vacances. pour : « Ainsi, j'ai dû écourter mes vacances. » (Registres des langue , s. d.)

Tableau 1 :

Les caractéristiques des trois registres de langue

	Registre familier	Registre courant	Registre soutenu
Situation	Modèle oral utilisé avec des proches (intimité) parole spontanée absence de hiérarchie et de contraintes	Langue standard (français international) Une certaine distance avec l'interlocuteur Échanges neutres dans des circonstances quotidiennes (registre employé à l'école, au travail ou dans les médias)	Modèle écrit (registre employé dans de nombreuses œuvres littéraires) Environnement social cultivé Marque de politesse Parole non spontanée Rapport hiérarchique
Lexique	Vocabulaire de la vie quotidienne termes familiers, parfois argotiques (voire grossiers)	Vocabulaire usuel pas de termes recherchés ou spécialisés (mots compris par tout le monde)	Vocabulaire riche, recherché et spécialisé, parfois rare termes littéraires, poétiques
Syntaxe	Abréviations Ruptures de construction Répétitions, ellipses (phrases sans verbe, etc.) Suppression du *ne* dans les négations pas de concordance des temps	Respect des règles de grammaire Emploi des temps simples de l'indicatif (passé simple, futur, etc.) phrases coordonnées et phrases subordonnées simples (relatives, subordonnées de cause, etc.)	Respect des règles de grammaire Emploi des temps du subjonctif, du conditionnel, etc. Concordance des temps constructions complexes
Prononciation	Syllabes avalées	standard	Articulation soignée et respect des liaisons
Figures de style	Exagération (hyperboles), périphrases, expressions toutes faites (voir Figures de style)	ton neutre, peu d'effets de style	Métaphores, chiasmes, etc., recherche d'effets de style (voir Figures de style)
Exemple	T'as pas vu mes godasses ?	Tu n'as pas vu mes chaussures ?	N'aurais-tu pas vu mes chaussures, par hasard ?

Note : tiré des caractéristiques des trois registres de langue (Lingolia Logo, s. d.)https://francais.lingolia.com/fr/atelier-decriture/les-registres-de-langue

I.4- La variété linguistique :

I.4.1- Définition :

Une des notions les plus productives de la linguistique moderne est définie en sociolinguistique comme un système linguistique spécifique et cohérent utilisé par une catégorie délimitée de locuteurs. Il est clair qu'un locuteur peut utiliser différentes formes linguistiques pour exprimer la même chose dans un même contexte linguistique. Comme discuté précédemment, une variable linguistique désigne un élément qui peut se manifester sous différentes formes. Ainsi, chaque forme différente de cette variable est appelée une variante de cette dernière.

Les variétés linguistiques sont définies comme des ensembles de différences présentes simultanément au niveau du lexique, de la grammaire et de la phonologie, ou bien à un ou deux de ces niveaux dans le système « des ensembles de différences situées toutes à la fois au niveau du lexique, de la grammaire et de la phonologie ou bien à un ou deux seulement de ces niveaux dans le système » (Garmadi, 1981, pp. 27-28) .

I.4.2- les types de variété linguistique :

Les sociolinguistes ont considéré qu'il y a trois types de variété :

- Les variétés géographiques (régionales)
- Les variétés sociales
- Les variétés historiques

I.4.2.1- les variétés géographiques (régionales). Les variétés linguistiques se distinguent en raison de la distance géographique qui sépare les locuteurs, et ces différences régionales se manifestent au niveau du lexique, de la grammaire et de la phonologie. Par exemple, le français parlé à Paris constitue une variété régionale du français, tout comme celui utilisé à Marseille. De même, le français parlé à Montréal diffère de celui utilisé à New Brunswick.

Selon Fishman, (1971) les variétés linguistiques issues de différentes régions sont généralement désignées sous le nom de dialectes. « Les variétés d'origine géographique sont reconnues sous le nom de dialectes. » (p.67)

Les différences entre ces variétés régionales se manifestent au niveau du lexique, de la grammaire et de la phonologie.

I.4.2.2- les variétés sociales. Ces variétés linguistiques expliquent les différences d'usage observées parmi les diverses couches de la société ou en fonction de l'environnement dans lequel évolue le locuteur. Elles sont influencées par plusieurs facteurs tels que le statut socioprofessionnel, l'âge, le sexe et l'environnement social :

- Le côté socioprofessionnel : La profession confère à l'individu un statut social défini et influence sa manière de parler. Ainsi, un avocat ne s'exprime pas de la même manière qu'un ouvrier.
- L'âge et le sexe : Ces facteurs peuvent également déterminer les groupes sociaux. De plus, les femmes ont tendance à perfectionner leur discours en respectant les règles de la langue, tandis que les hommes peuvent être moins préoccupés par cette perfection linguistique. « Les femmes, plus que les hommes sensibles aux modèles de prestiges, utilisent moins les formes linguistiques stigmatisées, considérées comme fautives, en discours surveillé.» (Garmadi, 1981, p. 189)

Les jeunes ont aussi un mode d'expression distinct de celui des personnes âgées, ce qui donne lieu au terme de "parler des jeunes". Ainsi, les mots ne sont pas neutres ; ils indiquent l'appartenance à telle ou telle génération.

L'environnement social exerce également une influence sur le langage des membres de la société. Par exemple, ceux qui évoluent dans des milieux intellectuels tels que les avocats ou les enseignants utilisent un lexique différent de ceux qui fréquentent des milieux religieux ou sportifs.

I.4.2.3- les variétés historiques. Cette dernière fait référence au changement linguistique qui se manifeste lorsque nous comparons des textes écrits dans une même langue à des moments différents. La variété historique d'une langue se réfère aux changements linguistiques observés à travers le temps, reflétant les évolutions culturelles et sociales qui ont façonné son utilisation. Cette forme de variation permet d'étudier comment les langues se transforment et s'adaptent au fil des siècles, offrant ainsi un précieux témoignage de l'histoire humaine.

C'est ainsi que Walter, (1994) annonce que : «La langue est un palimpseste sur lequel s'écrivent les marques successives de l'histoire.», Cela signifie qu'il évoque de manière poétique comment les langues portent les traces et les couches successives de l'histoire, soulignant ainsi la nature dynamique et en constante évolution des langues à travers les époques.

I.5- Distinguer la variation linguistique de la variété linguistique :

Le concept de variation linguistique fait référence aux différences ou écarts observables entre plusieurs variétés linguistiques par rapport à une norme linguistique standard. Ces variations peuvent se manifester sur les plans phonologiques, lexico-sémantique ou morphosyntaxiques. Cependant les variétés linguistiques se définissent comme étant « des ensembles de différences situées tout à la fois au niveau du lexique, de la grammaire et de la phonologie ou bien à un ou deux seulement de ces niveaux dans le système » (Garmadi, 1981, pp. 27-28).

Par variété linguistique, Une variété linguistique se réfère à un ensemble d'utilisations qui présentent des différences par rapport à la langue standard dont elles dérivent généralement. Ces variations résultent des façons dont les individus utilisent le système linguistique. Certains termes comme "patois", "progue", "platt", "bable", sont parfois utilisés par les non-spécialistes pour décrire des variétés linguistiques qu'ils jugent socialement inacceptables. Cependant, cette classification est rejetée par Garmadi, qui souligne qu'il ne faut pas établir de hiérarchie entre les variétés linguistiques. Les spécialistes qui emploient ces termes doivent les redéfinir de manière scientifique. (Garmadi, 1981, pp. 29-36)

En conclusion, la sociolinguistique nous aide à comprendre comment les gens utilisent la langue, en fonction de leur environnement social. Elle nous montre que la langue n'est pas seulement une structure interne, mais qu'elle est influencée par les interactions sociales et culturelles.

Les variétés linguistiques, comme les dialectes régionaux et les langages professionnels, illustrent cette diversité. De plus, la maîtrise des différents registres de langue, comme le familier, le courant et le soutenu, est importante pour une communication efficace.

Enfin, il est essentiel de distinguer entre la variation linguistique, qui concerne les écarts par rapport à la norme linguistique, et la variété linguistique, qui désigne les différentes formes d'utilisation de la langue.

Chapitre II

Présentation du ChatGPT

Avant d'explorer en profondeur le sujet de ChatGPT, qui constitue l'axe central de cette étude, il est impératif d'examiner brièvement certains concepts fondamentaux interconnectés avec ChatGPT, afin d'éclaircir les processus de son architecture et de mieux comprendre les mécanismes de son fonctionnement, notamment en ce qui concerne le langage naturel. Ces concepts, issus de divers domaines scientifiques tels que l'informatique, la linguistique cognitive et l'intelligence artificielle (IA), etc. Et qui ont tous contribué de manière significative à l'émergence de ChatGPT. Chaque domaine a eu un apport déterminant dans l'évolution de cette nouvelle technologie, qui repose essentiellement sur les avancées en intelligence artificielle.

II.1. Linguistique cognitive :

Pour les sciences cognitives, qui s'intéressent au fonctionnement du cerveau et de l'esprit, le langage est un objet de recherche primordiale. L'espèce humaine est la seule à posséder cette capacité supérieure, qui est particulièrement complexe. De nombreuses disciplines qui étudient la cognition portent leur attention sur le langage. Ce sont des psychologues, philosophes, anthropologues, spécialistes en neurosciences ou en intelligence artificielle. Chacune, de ces disciplines a pour but d'analyser, en fonction de ses besoins scientifiques, la faculté du langage humain dans la cognition matérielle ou artificielle.

En étudiant son fonctionnement de manière plus centrale et spécifique que les autres disciplines mentionnées précédemment, la linguistique, consiste à étudier le langage en se basant sur la variété des langues naturelles. Elle est plus centrale car c'est la seule science à avoir le langage pour objet d'étude, et plus spécifique, car c'est à partir de l'étude de la structure du langage qu'elle aborde cet objet. (Fuchs, et al., 2004, pp. 1-2)

La linguistique cognitive constitue une branche de la linguistique générale, elle se distingue de l'étude traditionnelle des langues par son aspect universel. Elle s'intéresse plutôt aux relations entre le langage et la faculté cognitive chez l'être humain, telles que la mémoire, la conscience, la pensée, etc. En mettant l'accent sur la façon dont la pensée influence mutuellement le langage, à cet égard, Fuchs et al. (2004) déclare que : « La linguistique cognitive est un courant linguistique qui estime que la création, l'apprentissage et l'usage du langage trouvent leur meilleure explication par référence à la cognition humaine en général » (p.1). En effet, la linguistique cognitive met en valeur un principe fondamental de ce courant linguistique, à savoir, que le langage humain ne peut être compris qu'en référence aux processus cognitifs généraux de l'être humain. La linguistique cognitive s'écarte ainsi d'une approche purement formelle ou structurale du langage, et se propose comme une

approche qui s'intéresse aux mécanismes mentaux sous-jacents à son acquisition, son traitement et son utilisation.

Effectivement, la citation souligne que la création, l'apprentissage et l'usage du langage trouvent leur meilleure explication dans la cognition humaine. Cela signifie que la linguistique cognitive considère que le langage est intimement lié aux capacités perceptives, attentionnelles, conceptuelles et autres fonctions cognitives de l'individu. Le langage n'est donc pas vu comme un système autonome, mais comme une faculté étroitement imbriquée dans la cognition générale.

Cette approche contraste avec les théories linguistiques plus formalistes, qui cherchaient à décrire le langage de manière indépendante des processus cognitifs. La linguistique cognitive, au contraire, postule que la compréhension du langage passe nécessairement par celle de la cognition humaine dans son ensemble.

II.1.1- La théorie de Chomsky :

Noam Chomsky l'un des intellectuels américains les plus marqués dans le XXème siècle, né le 7 décembre 1928. Il est connu pour ses travaux dans les domaines de la linguistique, de la philosophie, la politique, etc. Chomsky est considéré comme l'un des penseurs les plus influents du XXe siècle.

La théorie de Chomsky sur le langage consiste à considérer que certaines règles et structures grammaticales sont innées pour toutes les langues humaines. Malgré les différences évidentes entre eux, cette théorie va au-delà et suggère une structure profonde et implicite qui est universelle pour toutes les langues. En d'autres termes, Chomsky affirme que les enfants naissent avec un dispositif d'acquisition du langage (DAL), une capacité biologique qui leur permet d'acquérir les règles grammaticales sans effort. (Structural Learning) https://www.structural-learning.com

Dans le domaine de la linguistique, c'est grâce aux travaux de Chomsky que la théorie de la grammaire générative et transformationnelle, a vu le jour, ce qui a profondément influencé les recherches dans le champ de la linguistique moderne. En fait, sa théorie a affecté la manière dont les linguistes étudient la structure du langage et ainsi, sa génération dans l'esprit humain. Ses idées ont également contribué à la compréhension de la nature innée du langage et à l'exploration des universaux linguistiques.

La linguistique cognitive privilégie donc l'étude du langage, faculté universelle de l'espèce humain, par opposition à l'étude des langues particulières qui en

> résulte. La linguistique cognitive se compose de deux grandes mouvances, dont le point commun est la prise en compte du fait mental pour expliquer les phénomènes linguistiques. La première initiée par Chomsky dans les années 1950 se construit contre le béhaviorisme, la linguistique structurale et l'empirisme. Elle est formaliste et se focalise sur la syntaxe. Dans ce cadre, les langues sont générées par des grammaires inconscientes innées et la cognition est modulaire. La seconde remet fortement en cause la linguistique chomskyenne, depuis la fin des années 1970. Elle est symbolique et se focalise sur la sémantique. Dans ce cadre, les langues sont générées par un sujet conscient et la cognition est intégrée. (Notari, 2009, p. 22)

L'approche de la linguistique cognitive met l'accent sur l'étude du langage en tant que faculté universelle de l'homme, plutôt que sur l'étude des langues elles-mêmes. Dans ce sens, il distingue deux grands mouvements au sein de ce courant, toutes deux centrées sur l'importance de la cognition dans l'explication des phénomènes en rapport avec le langage.

II.1.2- La métaphore de l'ordinateur :

Les sciences cognitives s'intéressent à l'étude des organes mentaux, tels que l'esprit et le cerveau humain, en s'appuyant sur la théorie mathématique de la calculabilité et du mouvement cybernétique du XXe siècle. Leur développement est lié à l'invention de l'ordinateur dans les années 1950-1960, ainsi qu'aux travaux scientifiques sur le langage humain et les algorithmes. Elles ont contribué à l'apparition de nouvelles technologies expérimentales telles que l'intelligence artificielle, et l'informatique. « La métaphore de l'ordinateur assimile l'esprit au logiciel, et le cerveau au matériel d'un ordinateur. Ce dernier montre qu'un calcul peut être mis en œuvre par une machine, et apparait comme la solution aux vieux problèmes philosophiques des rapports entre le corps et l'esprit.» (Notari, 2009, p. 13)

Cette analogie assimile l'homme à l'ordinateur, son esprit ressemble à un logiciel pour le faire fonctionner, ainsi son cerveau au matériel. Les sciences cognitives affirment que penser équivaut à calculer, symbolisé par l'ordinateur, outil de calcul par excellence.

La linguistique cognitive de Noam Chomsky soutient ce modèle computationnel, en affirmant que la syntaxe correspond à une opération de calcul. Autrement dit, Selon Chomsky, la syntaxe consiste en une transformation de l'information linguistique, où les règles syntaxiques sont utilisées pour créer des phrases correctement structurées. Cette idée selon laquelle la syntaxe est considérée comme un processus calculatoire est fondamentale

dans la théorie de la grammaire générative de Chomsky, qui cherche à comprendre comment les langues humaines produisent et interprètent des phrases. (Bergounioux, s. d.)

Pour comprendre comment la théorie de la métaphore de l'ordinateur est devenue un pivot essentiel dans les sciences cognitives, il est primordial de savoir que :

> La source originale de la métaphore de l'ordinateur, qui inspire la majeure partie des sciences cognitives, dont la linguistique cognitive chomskyenne, plonge ses racines dans les mathématiques. Plus précisément dans la révolution logique de la première moitié du XXème siècle, avec les travaux de Davide Hilbert, Kurt Gödel, Alonzo Church, Alan Turing. D'où a émergé la théorie de la calculabilité avec la machine abstraite de Turing (concept d'un automate abstrait), source théorique de la machine concrète de Von Neumann (l'ordinateur) qui voit le jour quelques années plus tard. Le calcul par le biais de la machine abstraite de Turing devient donc le modèle possible de l'esprit, qui jusqu'à présent échappait totalement à la science. Cette idée atteint son apogée avec la réalisation physique de l'ordinateur. (Notari, 2009, p. 14)

Cela veut dire, que la métaphore de l'ordinateur, qui a profondément marqué les sciences cognitives, trouve son origine dans les avancées mathématiques du début du XXème siècle. C'est en effet dans les travaux révolutionnaires de mathématiciens tels que David Hilbert, et Alan Turing que s'enracine cette conceptualisation du fonctionnement de l'esprit humain sur le modèle du calcul informatique.

Les concepts développés par ces pionniers, notamment la théorie de la calculabilité et la machine abstraite de Turing, ont jeté les bases théoriques de la machine concrète qui est l'ordinateur moderne. Cette machine de Turing, capable de réaliser des opérations de calcul de manière algorithmique, est devenue le modèle possible pour appréhender le fonctionnement de l'esprit, qui jusqu'alors échappait largement à la compréhension scientifique.

II.1.3- La Métaphore de l'Automate :

La théorie des automates de Chomsky vise à caractériser le mécanisme qui générerait toutes et seulement les phrases bien formées dans une langue naturelle. Les détails ont fait l'objet de nombreuses discussions au fil des années, notamment pour générer des phrases bien formées, mais les grandes lignes sont restées assez claires. Les règles de syntaxe choisies expriment le contenu de cet automate, et la théorie des transformations, interne à la

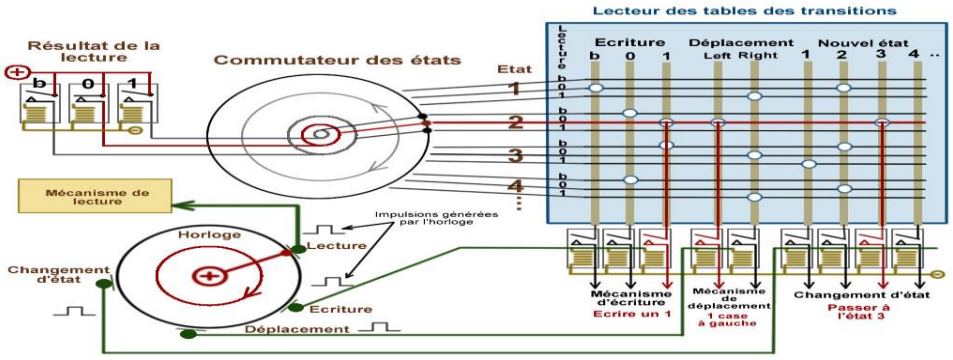
grammaire générative et transformationnelle, doit son héritage logique aux règles de la preuve des théorèmes en logique formelle.

Les liens entre l'automate et l'ordinateur sont forts. L'automate, en tant que machine abstraite et conception théorique, l'ordinateur est une concrétisation technologique de cette théorie. La grammaire générative et transformationnelle et les ordinateurs partagent une base théorique commune : l'automate mathématique. Pour fonctionner, cette grammaire doit être formalisée comme un algorithme, assimilable à un logiciel d'ordinateur. En effet, les programmes informatiques sont des algorithmes, ainsi les systèmes de règles de réécriture et de règles transformationnelles sont des algorithmes qui permettent de dériver mécaniquement une structure grammaticale à partir d'un symbole initial. Les transformations sont des procédures algorithmiques ordonnées qui convertissent une phrase noyau (assertion déclarative) en une phrase interrogative, négative, passive, etc. (Notari, 2009)

En somme, la nouvelle science de l'esprit, telle que Chomsky la présente, s'appuie sur la métaphore de l'automate pour expliquer le fonctionnement des langues et des machines.

Figure 1

Schéma de la machine de Turing.



Note : Tiré d'Un prototype programmable pour concrétiser la machine de Turing, (s. d.) https://bib.umontreal.ca/citer/styles-bibliographiques/citer-oeuvres-images-style-apa?tab=5231311

II.2- La linguistique informatique (LI) :

Les informaticiens ne peuvent plus se permettre de négliger la théorie des langages et des automates. La science informatique, surtout l'informatique théorique et ses applications, repose en grande partie sur la théorie des langages et des automates, considérées comme ses fondements les plus cruciaux. Cette théorie, même si elle peut être ennuyeuse et parfois complexe, est extrêmement utile tant pour le développement de nouveaux langages de programmation que pour la création de compilateurs, pour ne mentionner que ces exemples.

Une description grammaticale formelle du langage est essentielle dans de nombreux domaines tels que la linguistique, les mathématiques et l'informatique. Ceci transforme la théorie des langages formels en un domaine scientifique qui implique différentes disciplines. En fait, ses racines proviennent des mathématiques et de la linguistique. (Merah, 2022)

II.2.1- Définition de la Linguistique Informatique (LI) :

La linguistique informatique est un champ de recherche qui englobe plusieurs domaines scientifiques tels que l'informatique, la linguistique, les mathématiques, etc. Il implique la modélisation du langage naturel en utilisant des méthodes statistiques basées sur des règles, tout en examinant les solutions informatiques adaptées aux problèmes linguistiques. Dans un article scientifique sur la linguistique informatique Richard (2019) déclare que : « En général, la linguistique informatique (computationnelle) s'appuie sur des linguistes, des informaticiens, des experts en intelligence artificielle, des mathématiciens, des logiciens, des philosophes, des cogniticiens, des psychologues cognitifs, des psycholinguistes, des anthropologues et des neuroscientifiques. » (La linguistique informatique), cela veut dire qu'il s'agit d'un domaine interdisciplinaire rassemblant des experts de diverses disciplines pour étudier et analyser le langage humain grâce aux méthodes informatiques.

II.2.2- Le fonctionnement de la Linguistique Informatique (LI) :

La linguistique informatique (LI) utilise la science informatique, les mathématiques et la linguistique pour collecter, traiter et extraire des fonctionnalités des données linguistiques. Les données sont classées, transformées et analysées pour développer des algorithmes. Les modèles sont évalués pour mesurer leur précision et leur efficacité avant d'être appliqués dans des secteurs tels que la traduction automatique et l'analyse des informations... Le processus comprend la collecte de données, le traitement, l'extraction des fonctionnalités, le développement d'algorithmes, l'évaluation et l'application des modèles développés. Son objectif principal est de «permettre aux ordinateurs de comprendre, d'interpréter et de générer du langage humain de façon efficace. Ce processus implique le développement des capacités TALN qui permet aux machines de comprendre et d'interagir avec les humains ». (Luigisbox)

II.2.3- Traitement Automatique des Langues (TAL) :

Appelée également traitement automatique des langues naturel (TALN), en anglais, Natural Language Processing (NLP), une composante essentielle de la linguistique

informatique, fusionnant les expertises en linguistique et en informatique pour analyser et traiter le langage naturel de manière automatique :

> Le traitement automatique des langues (TAL) est un domaine de recherche pluridisciplinaire à l'intersection de la linguistique et de l'informatique, et désormais de l'intelligence artificielle, ou plus précisément de l'apprentissage artificiel. En guise de définition liminaire, on peut dire que le TAL concerne essentiellement l'analyse des langues au moyen d'un ordinateur. Le TAL est également connu sous diverses appellations qui traduisent parfois des nuances au sein de ce vaste domaine de recherche : on parle par exemple d'«**ingénierie linguistique** », quand l'accent est mis sur les aspects pratiques et opérationnels, ou de « **linguistique informatique** », quand c'est la linguistique qui tient un rôle important dans les recherches. Cette définition peut paraître néanmoins trop restrictive dans la mesure où le TAL recouvre également la «**génération de texte**», c'est-à-dire la production et non plus l'analyse automatique de textes. Ce champ de recherche et d'application est populaire, à l'heure des agents conversationnels et autres « **gadgets** » fondés sur une interaction avec l'utilisateur – pour qu'il y ait interaction, il faut prévoir à la fois une étape d'analyse et une étape de génération. (Poibeau, s. d.)

Il est important de noter que le domaine du TAL a subi des changements majeurs ces dernières années, et ce qui est fait aujourd'hui techniquement est très différent de ce qui était fait avant les années 2000. Cette discipline était principalement associée aux correcteurs orthographiques et aux traducteurs automatiques de qualité médiocre. En revanche, il y a de plus en plus d'applications qui sont maintenant visibles, fonctionnelles et assez efficaces.

Et pour clôturer cette partie, Fuchs et al. (1993, p.13) souligne que les traitements automatiques portent sur des données linguistiques sous forme de textes expressifs dans une langue naturelle. Afin de les traiter automatiquement, il est nécessaire de définir les règles linguistiques, de les représenter de manière opérationnelle et calculable, et de les mettre en œuvre à travers des programmes.

II.3- L'Intelligence Artificielle (IA) :

II.3.1- Définition de l'Intelligence Artificielle (IA) :

L'Intelligence Artificielle (IA) est un domaine de l'informatique qui se concentre sur la création de machines intelligentes capables d'effectuer des tâches nécessitant habituellement

l'intelligence humaine, comme la reconnaissance vocale et faciale, la prise de décision, la résolution de problèmes, l'apprentissage et la compréhension du langage naturel. En utilisant des algorithmes et des modèles mathématiques, l'intelligence artificielle (IA) imite les processus cognitifs humains et apprend des données pour s'améliorer au fil du temps. (Bouterraka, 2023, p.15), autrement dit, les systèmes d'IA s'améliorent progressivement en apprenant à partir des données auxquelles ils sont exposés. Cela souligne que la dimension d'apprentissage et d'évolution continues, ce qui caractérise l'intelligence artificielle. Durrleman, (s. d.), annonce que : « L'Intelligence artificielle a pour but la mise au point de machines intelligentes, c'est-à-dire, possédant les caractéristiques de l'intelligence humaine « … » on pourrait facilement dire que Turing en a semé les premières graines. » (p.49)

À vrai dire les recherches dans ce champ scientifique ne cessent de progresser à travers le monde, grâces aux travaux scientifiques des experts dans les domaines concernés. Ce qui l'affirme Aurélie (2020) dans un article scientifique en disant qu'il s'agit d'une « une discipline de recherche « … » et une matière, à l'instar des mathématiques ou de la physique, utilisée dans de nombreuses disciplines de recherche, comme la médecine. Aujourd'hui plus que jamais, comprendre ce qu'est l'intelligence artificielle (IA), ce qu'elle fait, ce qu'elle fera sûrement ». En effet, Les possibilités offertes par l'intelligence artificielle (IA) sont nombreuses et variées, avec un potentiel de changement et d'impact important dans divers secteurs de la société et de l'industrie. L'intelligence artificielle ('IA) est conçue pour automatiser des tâches variées, améliorer les services de santé, réinventer le domaine du transport autonome, révolutionner le secteur financier, transformer l'éducation, suscitant également d'importantes questions éthiques et réglementaires pour assurer une utilisation responsable de cette technologie. À ce sujet, il est important d'aborder quelques concepts interconnectés directement à cette technologie :

II.3.2- L'apprentissage profond (Deep Learnig) :

L'apprentissage profond (DL) est un sous-domaine de l'apprentissage machine, lui-même, fait partie de l'intelligence artificielle, à partir de ce dernier, nous comprenons l'acte de faire reproduire par des machines, des tâches compliquées, par les êtres humains, généralement : faire raisonner une machine, apprendre à une machine la planification de tâches, prendre des décisions, à représenter des connaissances d'une manière structurée.

> Les méthodes d'apprentissage machine ne sont pas récentes mais datent des années 60. Dès les débuts de l'informatique, les pionniers avaient imaginé des systèmes et des principes d'apprentissage à base de réseaux de neurones. Les

> bases de ce fonctionnement ont été développées sur des machines depuis 1986 avec notamment les travaux de Geoffrey Hinton, universitaire canadien. Jusqu'en 2006, d'autres méthodes ont pris le dessus et les méthodes d'intelligences artificielles par réseaux de neurones ont été mises de côté. (Douarre & Rousseau, 2021)

L'apprentissage profond (DL), rassemble actuellement les approches les plus efficaces et performantes utilisées dans le domaine de l'apprentissage automatique.

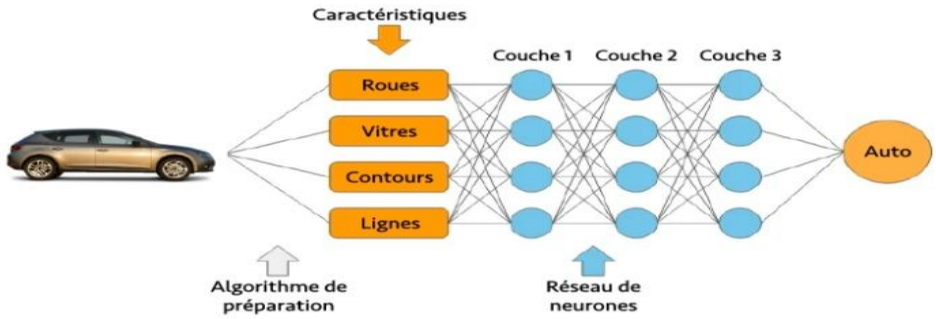
II.3.3- Les Réseaux de Neurones :

Les réseaux de neurones, également connus sous le nom de réseaux de neurones artificiels ou réseaux de neurones simulés, jouent un rôle essentiel dans l'apprentissage automatique. Ils sont fondamentaux pour les algorithmes d'apprentissage en profondeur. Inspirés par le fonctionnement du cerveau humain, ces réseaux imitent la communication entre les neurones biologiques. Leur structure comprend des couches nodales, notamment une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie.
Chaque nœud ou neurone artificiel est connecté à un autre avec un poids et un seuil associés. Lorsque la sortie d'un nœud dépasse le seuil spécifié, ce nœud est activé et envoie des données à la couche suivante ; sinon, aucune donnée n'est transmise. « Les réseaux de neurones s'appuient sur des données d'entraînement pour apprendre et améliorer leur précision au fil du temps. Cependant, une fois ces algorithmes d'apprentissage suffisamment affinés, ils constituent de puissants outils pour l'informatique et l'intelligence artificielle » (https://www.ibm.com/)

Cela permet une classification et un regroupement instantanés des données**.** Les tâches de reconnaissance vocale ou visuelle peuvent être accomplies en quelques minutes, tandis qu'une identification manuelle par des experts prend plusieurs heures. Le processus de traitement des informations via un système de neurones nécessite une identification quasiment immédiate de l'objet en question, que ce soit pour des neurones naturels ou artificiels. L'information, comme une image de voiture, est transmise à travers diverses couches du réseau afin d'extraire des caractéristiques spécifiques de l'objet. Chaque strate représente un degré d'abstraction impliqué dans l'exécution de cette mission en différentes phases.

Figure 2 :

Illustration de réseau de neurones



Note. Tiré de Réseau de neurones artificiels : qu'est-ce que c'est et à quoi ça sert ? (s. d.),https://praedictia.com/page/reseaux-de-neurones/quest-ce-quun-reseau-de-neurones.html

II.4- Les Chatbots conversationnels :

II.4.1- L'Historique de Chatbots :

L'histoire de Chatbots, (robots conversationnels), remonte aux années 1960. Parmi les premiers chatbots, il y avait ELIZA, qui a été développé par Joseph Weizenbaum au MIT. ELIZA connue par sa capacité de simuler une conversation simple comme par exemple, répondre aux questions et imiter un humain. Ce logiciel informatique, considéré comme le premier véritable agent conversationnel, a ouvert la voie dans le domaine des chatbots.

ELIZA a créé une interface virtuelle de communication entre l'homme et la machine. Elle a même parvenu à tromper certains utilisateurs en les faisant croire qu'elles parlaient à un vrai humain. Par contre, ELIZA n'a pas réussi à passer le test de Turing, qui vise à déterminer si une machine peut se faire passer pour un être humain.

Dès lors, les chatbots ont évolué et se sont diversifiés. Ils sont maintenant usées dans de différents domaines, tels que la médecine, l'enseignement, le voyage, etc. Leur utilisation continue de croître, et ils jouent un rôle essentiel dans notre interaction avec les systèmes d'information et les services en ligne.

> Les programmes de dialogue automatique ont connu un nouveau souffle entre 2017 et 2018. Au profit d'une série d'innovations dans le domaine de l'Intelligence Artificielle (IA), ces automates conversationnels se diffusent alors dans de nombreux secteurs, transports, alimentation, tourisme…avec des objectifs variés, entraînés par la vague rhétorique des technologies dites « intelligentes ». (Esteban, 2022)

Parmi les chatbots populaires, on retrouve Microsoft Copilot, ChatGPT, Anthrophic's Claude, Perplexity.ai et Jasper. Ces chatbots se démarquent par leurs capacités de recherche, de résumé de documents, de génération d'idées à partir de prompts, et par leur utilité pour les entreprises et les spécialistes du marketing.

II.4.2- ChatGPT :

ChatGPT est une technologie de traitement du langage naturel de pointe, développée par l'entreprise spécialisée en intelligence artificielle OpenAI. Il s'agit d'un agent conversationnel (chatbot), capable d'interagir avec les utilisateurs d'internet à l'aide du langage naturel.

II.4.2.1- OpenAI. Une société fondée en fin 2015 par Elon Musk et Sam Altman, est basée à San Francisco et se concentre sur l'intelligence artificielle, avec un capital initial de 100 millions de dollars. Le but est de favoriser et améliorer les algorithmes (IA) pour bénéficier à tous, en augmentant l'autonomie et en générant un impact transformateur dans des secteurs tels que la santé, le changement climatique et l'éducation. (ActiaAI, s. d.)

Parmi les projets développés par OpenAI, dans le domaine de l'intelligence artificielle il y en a :

- GPT-1, la première version du modèle du langage formé à partir d'une vaste quantité de données textuelles. Ce dernier peut générer du texte de manière autonome.
- GPT-2, la seconde version du modèle, mise au point en février 2019 et capable d'écrire des articles de presse.
- GPT-3, l'une des IA les plus puissantes dans l'histoire du NLP.
- Dall-E, un modèle génératif permettant de créer des images originales à partir de texte.
- Dall-E 2, une évolution de la première version, qui étend ses capacités de création d'images originales à partir de texte. (ActiaAI, s. d.)

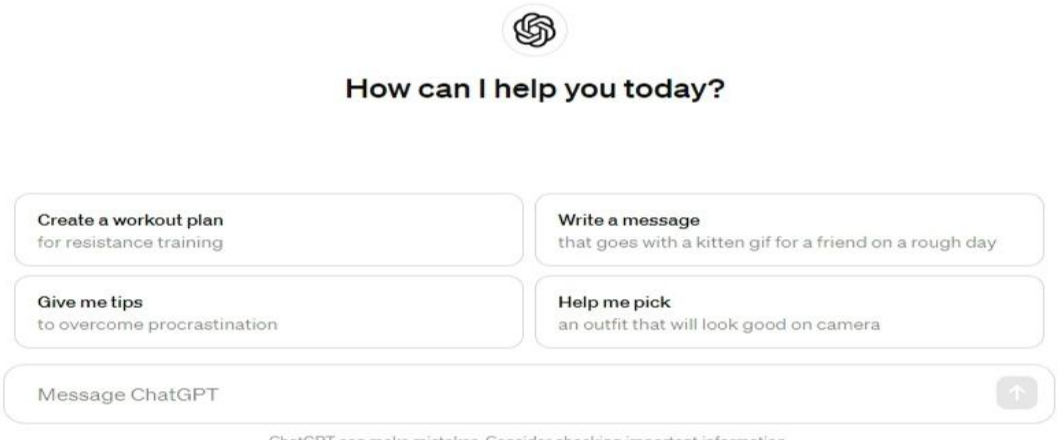
II.4.2.2- Définition de ChatGPT. L'acronyme « **ChatGPT** » désigne littéralement en anglais « **Chat Generative Pre-trained Transforme** », ce qui peut être traduit en français ainsi, (Transformateur génératif pré-entraîné de conversation), ou plus précisément :

- Chat" fait référence à la capacité du modèle à interagir de manière conversationnelle avec les utilisateurs.
- Generative" souligne la capacité du modèle à générer du texte de manière autonome.
- Pre-trained" indique que le modèle a été préalablement entraîné sur de grandes quantités de données pour améliorer ses performances.
- Transformer" fait référence à l'architecture de réseau de neurones utilisée dans le modèle, qui est efficace pour traiter des séquences de données.

Ce modèle de l'automate, qui a été développé par OpenAI, et proposé au public à partir de novembre 2022. Il s'agit d'un programme informatique interactionnel, très puissant qui peut communiquer avec ses utilisateurs sous forme de messages écrits ou vocales et répondre à leurs questions. En menant des conversations toutes entières, et qui impliquent toutes les fonctions du langage proposées dans la théorie de la communication de Jakobson. Il s'appuie pour cela sur l'apprentissage automatique pour générer des textes, avec différentes langues. Le programme fonctionne comme un chatroom en ligne (un robot conversationnel), accessible gratuitement depuis un smartphone, une tablette ou un ordinateur.

Figure 3

Illustration représente l'interface et le logo de ChatGPT.



Note : tiré de l'interface interactionnelle de ChatGPT, (s. d.),*https://chatgpt.com/?oai-dm=1*

II.4.2.3-Fonctionnement de ChatGPT. ChatGPT est un modèle interactionnel de génération de textes, basé sur des réseaux de neurones appelés transformateurs « transformers ». Ces derniers ont révolutionné le domaine du traitement

automatique des langues naturelles (TALN), en offrant une capacité de traitement contextuel et en permettant à l'IA de produire des textes de haute qualité.

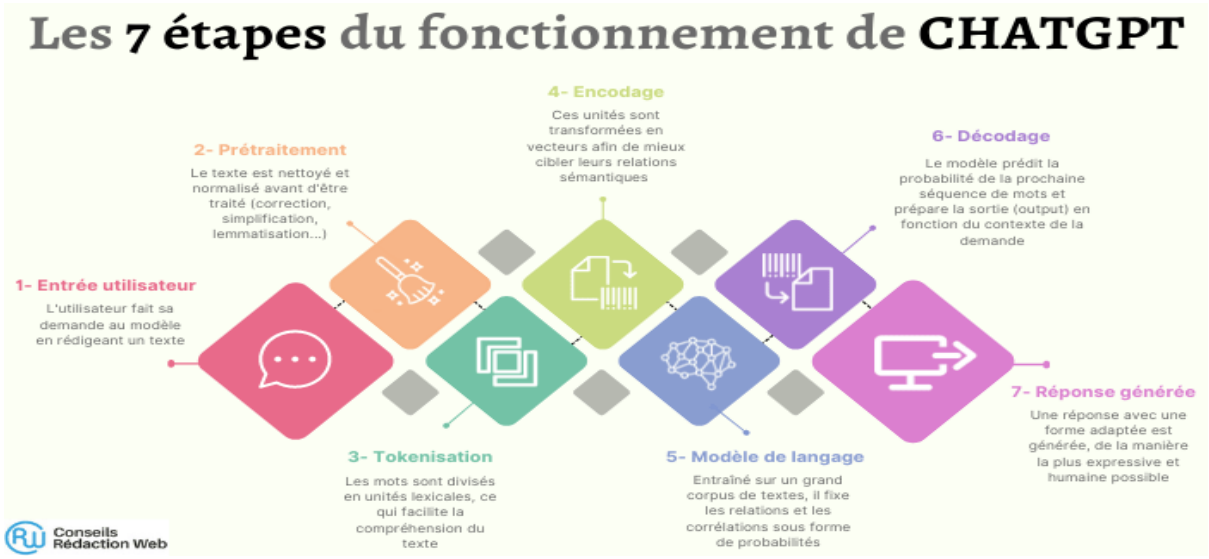
Avant de consulter les processus du fonctionnement de ChatGPT, il faut, d'abord, savoir que ce modèle de l'IA a été entraîné sur une très grande quantité de données textuelles. Ce stockage de textes lui ainsi rend capable d'apprendre les différentes structures grammaticales, syntaxiques et sémantiques de la langue. Le fonctionnement de ChatGPT est basé sur deux éléments clés : L'attention, La mémoire à long terme.

> L'attention permet au modèle de se concentrer sur les parties du texte les plus importantes pour produire une réponse pertinente. Tandis que la mémoire à long terme, quant à elle, permet au modèle de stocker des informations clés sur le contexte et d'utiliser ces informations pour produire une réponse plus cohérente. Le modèle est capable de produire des textes à partir de différentes langues et de différents styles d'écriture, en fonction de l'entraînement qu'il a reçu. Il peut également produire des textes de différents types, tels que des articles de blog, des résumés, des réponses à des questions, etc. Mais comment ChatGPT parvient-il à produire des textes si convaincants et naturels ? La réponse réside dans la quantité de données que le modèle a été entraîné à traiter. Plus le modèle a été entraîné sur des données variées et pertinentes, plus il est capable de produire des textes naturels et pertinents. (Anselmet, 2022)

D'après un article publié par Moussaoui, (2023), dans lequel, il indique que ChatGPT utilise des modèles de langage naturel pour interpréter et produire du texte, en s'appuyant sur l'apprentissage automatique. Ces modèles sont formés avec d'énormes volumes de données collectées de différentes sources telles que Common Crawl, WebText2, Books1, Books2 et Wikipedia. Les interactions avec ChatGPT se font en soumettant des requêtes écrites via une interface de programmation. Les réponses générées peuvent être corrigées en ajustant les modèles en fonction des paramètres spécifiques. ChatGPT utilise un ensemble de modèles nommé GPT-3.5, formé par trois modèles d'OpenAI : code-davinci-002, text-davinci-002 et text-davinci-003. Toutes les infrastructures de Microsoft Azure AI servent à entraîner tous les modèles.

Figure 4 :

Le schéma du fonctionnement de ChatGPT



Note : tiré de schéma de fonctionnement de ChatGPT,(2022),https://www.conseils-redaction-web.fr/chatgpt

En somme, ChatGPT est une intelligence artificielle (IA) de traitement du langage naturel, et de communication qui utilise des réseaux de neurones pour produire des textes de haute qualité et de conversation convaincantes.

En conclusion, nous pouvons dire que les avancées de la linguistique cognitive, notamment la théorie de Chomsky et la métaphore de l'ordinateur ainsi de l'automate, ont jeté les bases de la linguistique informatique et ouvert la voie au développement des chatbots comme ChatGPT.

L'utilisation du traitement automatique du langage, du l'apprentissage profond (Deep Learning), des réseaux de neurones et de l'intelligence artificielle a permis à ChatGPT d'acquérir des capacités conversationnelles avancées, de comprendre le langage naturel, de générer du texte original et de s'adapter au contexte.

Cette nouvelle technologie représente une avancée majeure dans le domaine de l'intelligence artificielle (IA), offrant de nombreuses possibilités innovantes pour la communication humaine et l'automatisation des tâches linguistiques.

La Partie Pratique

Chapitre III

Choix méthodologique et analyse de données

Les systèmes de traitement automatique des langues naturelles jouent un rôle très important dans l'interaction homme/machine. Ils facilitent la communication entre eux, en permettant des échanges linguistiques plus fluides et sans ambiguïté.

L'importance de notre mémoire réside dans le fait que la langue naturelle, comme un concept théorique essentiel, et un moyen de communication, est très compliqué au niveau de sa structure, voire de son utilisation. Autrement dit, l'usage de ce moyen par des logiciels intelligents, tels que ChatGPT, provoque de nombreux enjeux.

Parmi lesquels, celui qui porte sur l'étude des variations linguistiques et plus précisément les variations diaphasiques, notamment les registres de langue (soutenu, courant et familier), et leur impact sur les compétences linguistiques de ChatGPT vis-à-vis la compréhension et la production écrites. En demandant si ChatGPT, comme un actant principal et autonome dans ces échanges conversationnels, est capable à identifier le vocabulaire propre à chaque registre, ainsi les différentes caractéristiques linguistiques de chacun d'eux.

III.1- Choix méthodologique :

III.1.1- la nature de l'étude :

L'analyse des données de notre corpus, repose sur une étude quantitative et qualitative :

la première est basée sur le recensement des données linguistiques, Elle consiste à mesurer la capacité de ChatGPT à identifier les particularités de chaque registre, en s'appuyant pour cela sur le comptage des mots dans chaque conversation, qui caractérisent tel ou tel registre, ainsi le taux d'usage de chaque registre pour les deux interlocuteurs, en guise de pourcentage, fournissant, par conséquence, des données objectives et précises pour une analyse rigoureuse et minutieuse, basée sur des chiffres et des statistiques,

Tandis que la seconde, a pour but d'analyser les particularités linguistiques et la construction phrastique de chaque conversation, en examinant les phrases et les énoncés des deux interlocuteurs, de chaque conversation ; l'approche qualitative permet d'identifier les caractéristiques lexicales et les structures syntaxiques propres à chacun des trois registres. Ce procédé fournit une étude approfondie sur le choix de ces éléments linguistiques par les deux interlocuteurs, et également l'influence sociale et culturelle des interlocuteurs humains sur la réaction de ChatGPT, et comment ces éléments affectent la capacité de ce dernier à comprendre et à s'ajuster aux divers registres de la langue française.

III.1.2- La méthode de recherche :

Le choix de la méthode pertinente, joue un rôle primordial dans la recherche scientifique. À cet effet, pour mener une analyse minutieuse, nous allons opter pour une méthode de recherche mixte, descriptive et analytique à la fois.

La méthode analytique se définit, comme une approche systématique et rigoureuse a pour but d'étudier un sujet spécifique. Dans notre cas, les variations diaphasiques (les registres de langue). Cette méthode permet de recueillir des données linguistiques quantifiables et de les analyser statistiquement en offrant des chiffres qui vont nous aider à déterminer le registre dominant de chaque conversation, et le taux d'usage de chaque registre par les deux interlocuteurs.

Quant à la méthode descriptive de recherche scientifique consiste à observer et à décrire de manière précise les caractéristiques linguistiques propres à chaque registre. Elle vise à fournir une représentation fidèle et détaillée des caractéristiques des variables étudiées, en se concentrant sur trois éléments principaux, le lexique utilisé dans chaque registre, sa structure syntaxique ainsi son style de langage.

Afin de réaliser cette analyse nous allons choisir les captures d'écran comme un instrument de collecte de données, qui constituent notre corpus.

III.1.3-Présentation du corpus :

Cette étude est basée sur un corpus constitué d'un échantillon bien sélectionné, de trente huit captures d'écran d'ordinateurs et de smartphones, représentant six échanges conversationnels écrits, entre autres, réalisées sur internet ; les captures d'écran sont en nombre de trois jusqu'à dix par conversation :

La première conversation : 10 captures d'écran.

La deuxième conversation : 3 captures d'écran.

La troisième conversation : 8 captures d'écran.

La quatrième conversation : 8 captures d'écran.

La cinquième conversation : 4 captures d'écran.

La sixième conversation : 5 captures d'écran.

Ces conversations regroupent, d'un côté, ChatGPT version 3.5, et de l'autre côté, l'un des internautes humains concernés (utilisateurs d'internet).

Le public concerné se constitue d'un échantillon des internautes sélectionnés parmi d'autres, en fonction de notre besoin scientifique concernant l'usage varié des registres de langue par chacun d'eux.

III.2- Étude quantitative des données et vérification de l'hypothèse n°1 :

L'étude quantitative des données inclûtes dans les conversations de notre corpus nous sert, d'abord à déterminer le registre de langue dominant à travers les statistiques des données linguistiques et le taux d'utilisation de chaque registre employé par les deux interlocuteurs, ce qui nous permet de vérifier la validité notre première hypothèse.

III.2.1- Analyse des conversations :

Conversation n°1 :

Note : tous les exemples sont disponibles au niveau de l'Annexe A.

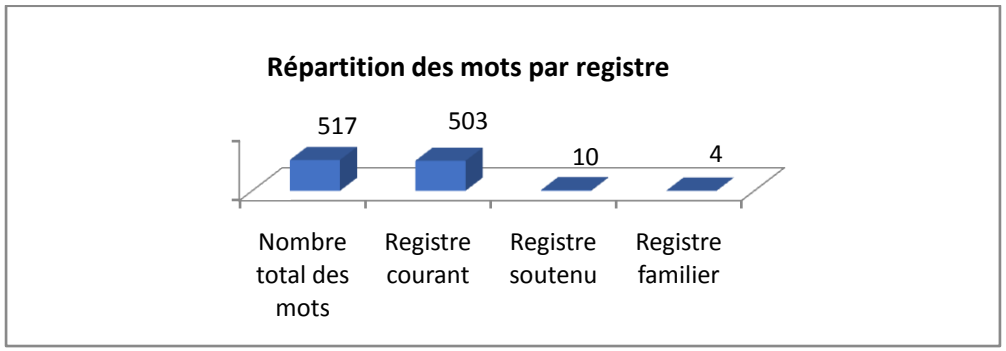
- Nombre total de mots dans la conversation : 517 mots

<u>Répartition des mots par registre de langue :</u>

- Registre courant : 503 mots (97.29%)
- Registre soutenu : 10 mots (1.93%)
- Registre familier : 4 mots (0.77%)

Figure 1 :

Répartition des mots par registre dans la conversation n°1



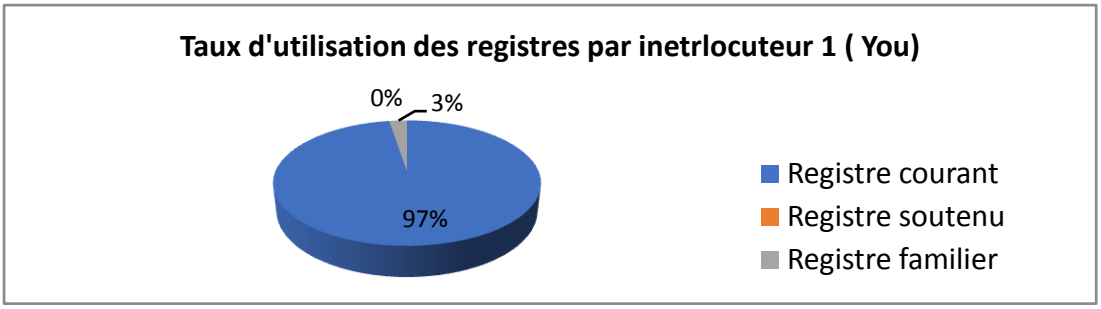
<u>Taux d'utilisation des registres par interlocuteur :</u>

- **<u>Interlocuteur 1 (You) :</u>**

- Registre courant : 97% (152 mots)
- Registre soutenu : 0% (0 mots)
- Registre familier : 3% (4 mots)

Figure 2 :

Taux d'utilisation des registres de langue par l'interlocuteur n°1(You).

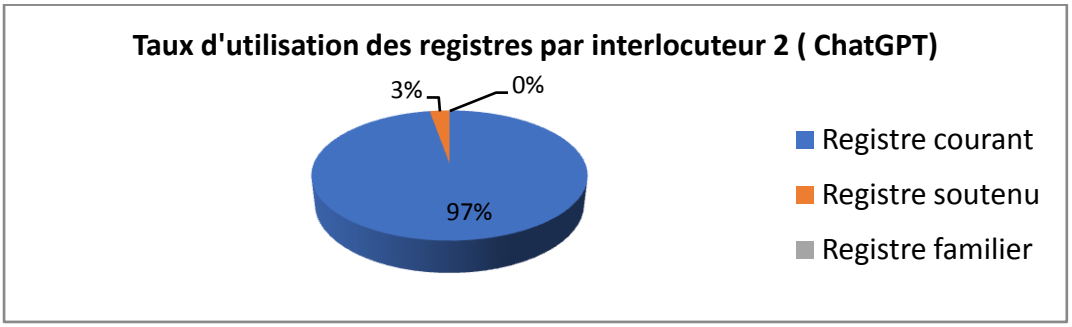


- **Interlocuteur 2 (ChatGPT) :**

– Registre soutenu : 3% (10 mots)
– Registre courant : 97% (361 mots)
– Registre familier : 0% (0 mots)

Figure 3 :

Taux d'utilisation des registres de langue par l'interlocuteur n°2 (ChatGPT).



Note : Tous les exemples sont disponibles sur l'annexe A.

Mots du registre courant utilisés par l'utilisateur 1 (You) :

– Bonjour, comment, puis-je, aider, aujourd'hui, (cf. Annexe A, capture d'écran, n° 1)
– Cas, urgences, intervenir, sauver, vie (cf. Annexe A, capture d'écran n°7)

Mots du registre familier utilisé par l'utilisateur (You) :

– Dis donc (cf. Annexe A, capture d'écran n°8)

Mots du registre courant utilisés par ChatGPT :

– Bonjour, comment, puis-je, aider, aujourd'hui, fantastique, domaine, fascinant, questions, études, médecine, général (cf. Annexe A, capture d'écran n°1)

Mots du registre soutenu utilisés par ChatGPT :

- Fantastique, fascinant, (cf. captures d'écran n° 1)
- Substitut (cf. capture d'écran n° 7)
- Surmonter (cf. capture d'écran n°2)

L'analyse statistique montre que le registre de langue dominant dans la conversation n°1 est :

- le registre courant, avec 503 mots sur 517 au total (97,29%).
- Le registre soutenu est peu utilisé, uniquement par ChatGPT avec 10 mots (1.93%).
- Le registre familier est presque absent avec seulement 4 mots (0.27%).

L'utilisateur emploie en grande partie le registre courant (97%), et seulement (3%), le registre familier ; tandis que ChatGPT utilise majoritairement le registre courant (97%) mais intègre aussi quelques éléments du registre soutenu (3%).

Conversation n°2 :

Note : tous les exemples sont disponibles au niveau de l'Annexe B.

Nombre total de mots dans la conversation : 188

Répartition des mots par registre de langue :

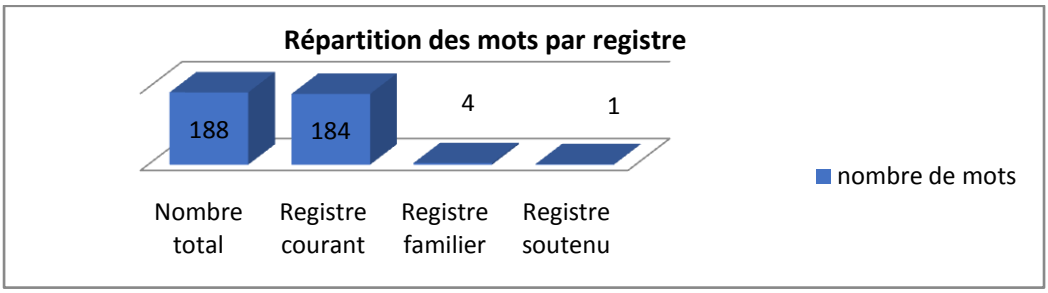
Registre courant : 184 mots (98%)

Registre familier : 4 mots (1.5%)

Registre soutenu : 1 mot (0.5%)

Figure 4 *:*

Répartition des mots par registre dans la conversation n°2.



Taux d'utilisation des registres par interlocuteur :

- **Interlocuteur 1 (You) :**

Nombre total de mots utilisés : 43 mots (100%)

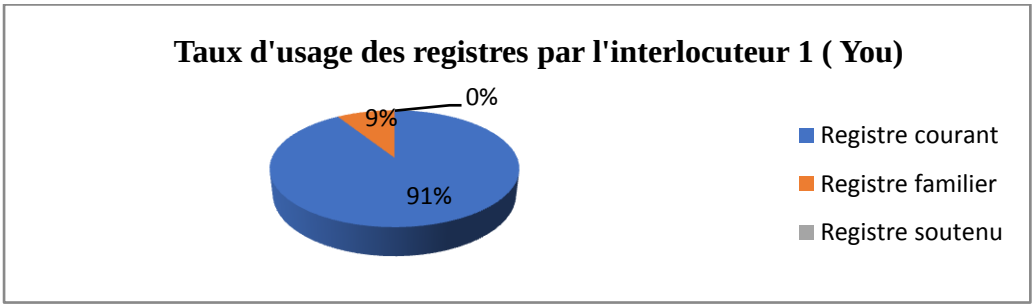
Registre courant : 39 mots (91%)

Registre familier : 4 mots (9%)

Registre soutenu : 0 mots (0%)

Figure 1 :

Taux d'usage des registres par l'interlocuteur 1 (You).



- **Interlocuteur 2 (ChatGPT) :**

Nombre total de mots utilisés : 145 mots

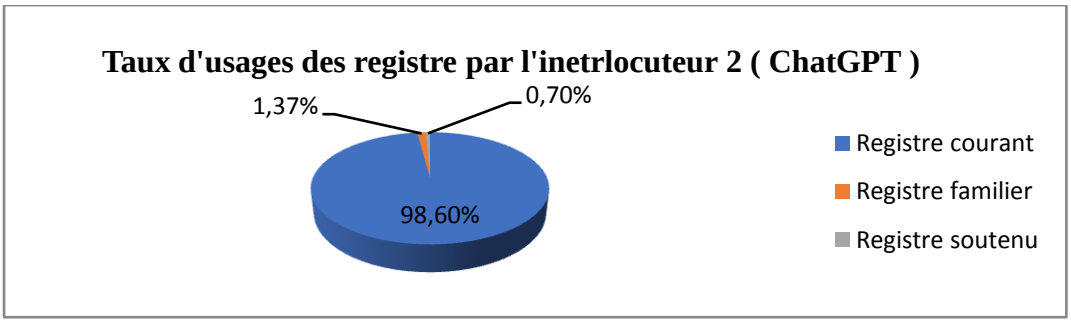
Registre courant : 143 mots (98.6%)

Registre familier : 2 mot (1.37%)

Registre soutenu : 1 mot (0.7%)

Figure 6 :

Taux d'usage des registres par l'interlocuteur 2 (ChatGPT).



- Mots du registre courant utilisés par (You) :
  - Alors, propose, un sujet, pour discuter, (cf. capture d'écran n°2)
  - Elle habite, le coté, francophone, précisément, Montréal (cf. capture d'écran n°3)
- Mots du registre courant utilisés par ChatGPT :

Avez-vous, déjà, visiter, endroit, particulièrement, marqué, y a-t-il, pays, rêvez, découvrir, c'est, une, excellente, idée (cf. captures d'écran n° 1, 2, 3) sur la même annexe.

- Mots du registre familier utilisés par You :
  - Coucou, ça, va, propose (cf. captures d'écran n°1, 2)
- Mots du registre soutenu utilisés par ChatGPT :
  - Particulièrement, marqué, rêvez, découvrir, magnifique, délicieux, restaurants, dynamique, accueillante, multiculturelle, (cf. captures d'écran n° 2, 3)
- Mots du registre familier utilisés par ChatGPT :
  - Coucou, ça va.

L'analyse quantitative de cette conversation, montre que le registre courant domine largement avec 98% des mots utilisés.

L'interlocuteur 1 (You) utilise principalement le registre courant (91%), mais inclut quelques mots du registre familier (9%).

L'interlocuteur 2 (ChatGPT) utilise presque exclusivement le registre courant (98.6%), avec une seule occurrence de vocabulaire soutenu (0.7%) et deux mots familier (1.37%). La conversation reste globalement informelle et simple, reflétant un échange quotidien avec un léger ton familier au début.

Conversation n°3 :

L'analyse quantitative de la conversation n° 03 montre les statistiques suivantes :

Nombre total de mots dans la conversation : 381 mots

Répartition des mots par registre de langue :

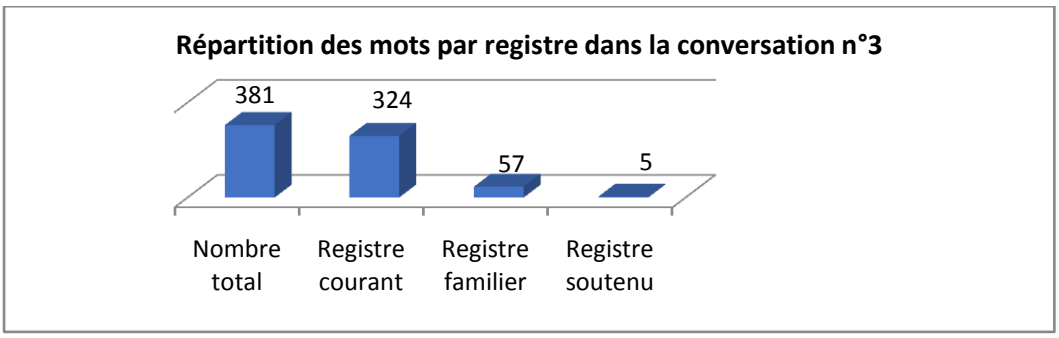
Registre courant : 324 mots (85,03%)

Registre familier : 57 mots (14,96%)

Registre soutenu : 5 mots (1,31%)

Figure 7 :

Répartition des mots par registre dans la conversation n°3.



<u>Taux d'utilisation des registres par interlocuteur</u> :

- **<u>Interlocuteur 1 (utilisateur) :</u>**

  Nombre total de mots utilisés : 134

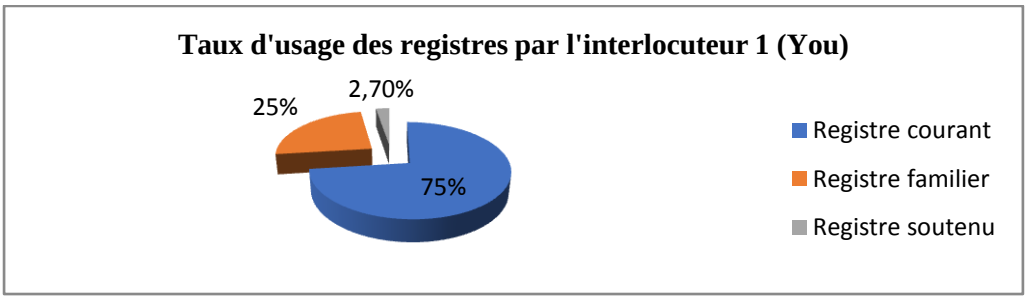
  Registre courant : 100 mots (75%)

  Registre familier : 33 mots (25%)

  Registre soutenu : 1 mot (2,7%)

Figure 8 :

Taux d'usage des registres par l'interlocuteur 1 (You)



- **<u>Interlocuteur 2 (ChatGPT) :</u>**

  Nombre total de mots utilisés : 247

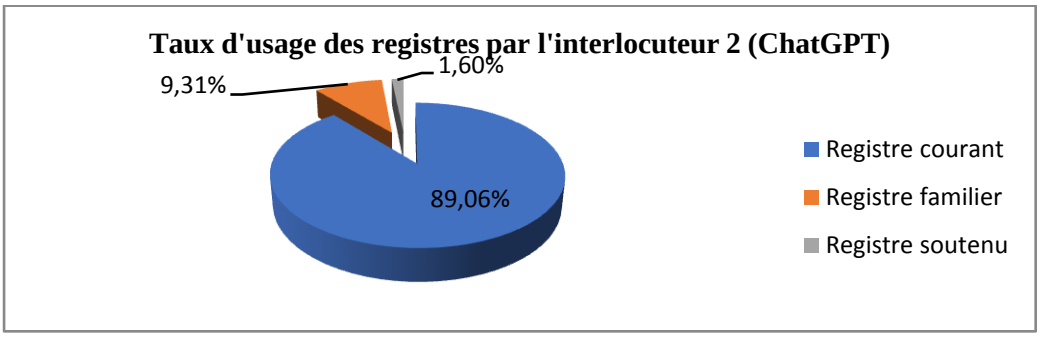
  Registre courant : 220 mots (89.06%)

  Registre familier : 23 mots (9.31%)

  Registre soutenu : 4 mots (1.6%)

Figure 9 :

Taux d'usage des registres par l'interlocuteur 2 (ChatGPT).



En comparant les deux interlocuteurs, nous constatons que l'interlocuteur 1 utilise davantage le registre familier par rapport à ChatGPT. Ce dernier adopte un langage principalement courant, poli et adapté à une conversation informelle.

Cette analyse met en évidence la capacité de ChatGPT à s'adapter au registre de langue dominant dans la conversation, tout en restant dans un style respectueux et approprié. Cela contribue à une interaction fluide et naturelle entre les interlocuteurs.

Conversation n°4 :

Note : tous les exemples sont disponibles au niveau de l'Annexe D.

Nombre total des mots dans la conversation n°4 : 586 mots

<u>Répartition des registres de langue :</u>

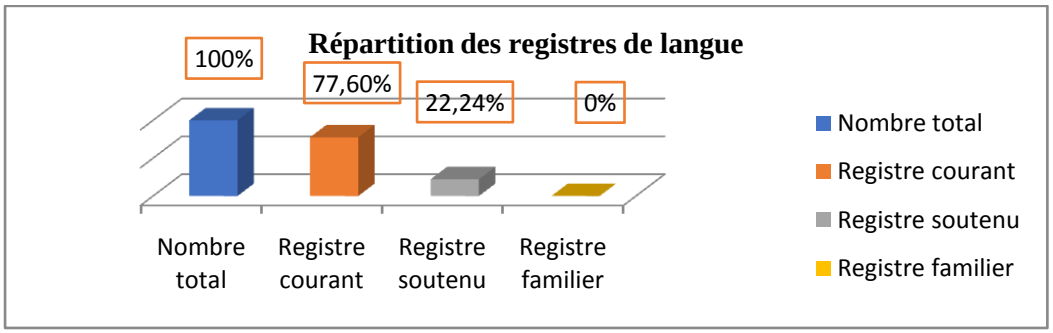
Registre courant : 455 mots (77,6%)

Registre soutenu : 131 mots (22,24%)

Registre familier : 0 mot (0%)

Figure 10 :

Répartition des registres de langue dans la conversation n°4



Analyse individuelle des interlocuteurs :

- **Interlocuteur 1 (You) :**

– Registre courant : nombre de mots : 101 mots, (67.33%)

Exemples de vocabulaire : "lecture", "romans", "philosophie", "histoire"

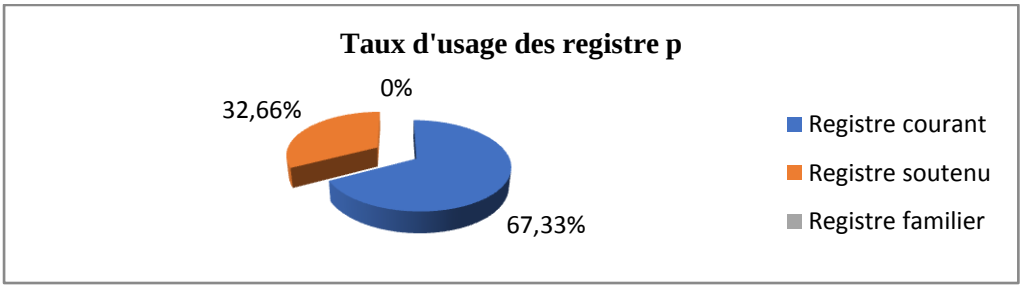
– Registre soutenu : nombre de mots : 49 mots, (32.66%)

Exemples de vocabulaire : "réconfort", "inspiration", "croissance personnelle"

– Registre familier : 0 mot, (0%)

Figure 11 :

Taux d'usage des registres par l'interlocuteur 1 (You).



- **Interlocuteur 2 (ChatGPT) :**

– Registre courant : nombre de mots : 354 mots, (81.19%).

Exemples de vocabulaire : "fantastique", "capturer", "nuances", "fenêtre"

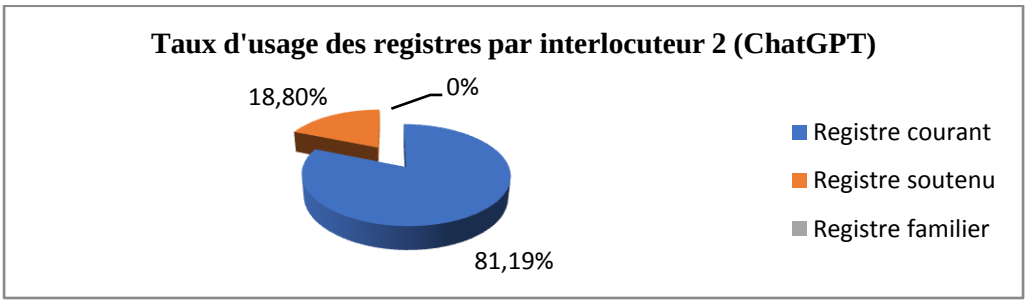
- Registre soutenu : nombre de mots : 82 mots, (18.80%).

  Exemples de vocabulaire : "pertinente", "intellectuels", "stimule", "équilibré"

- Registre familier : nombre de mots : 0 mot, (0%).

Figure 12 :

Taux d'usage des registres par l'interlocuteur 2 (ChatGPT)



En examinant cette conversation, nous constatons qu'elle est dominée majoritairement par le registre courant, représentant 77,6% des mots utilisés, ce qui reflète un échange informel et naturel entre les interlocuteurs. Le registre soutenu est présent également, avec 22.24% des mots, apportant une touche de formalité et de profondeur à certains échanges. Par contre, le registre familier, il est entièrement absent.

Conversation n°5 :

Note : tous les exemples sont disponibles au niveau de l'Annexe E.

L'analyse quantitative nous révèle les nombres suivants :

Le nombre total de mots dans la conversation : 210 mots

Répartition des registres de langue :

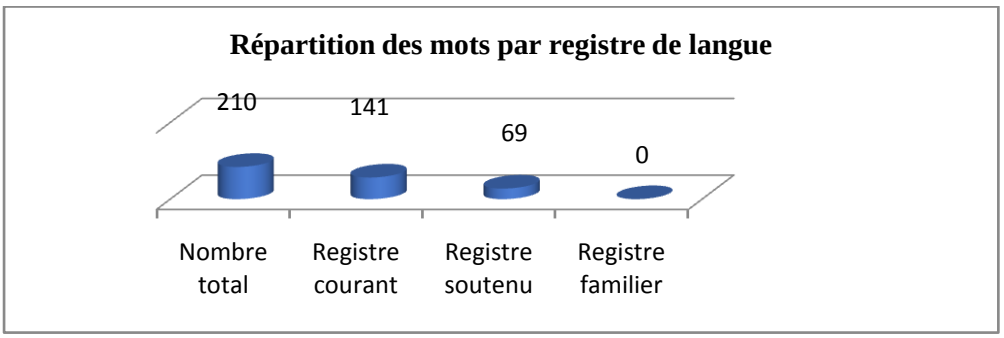
Registre courant : 141 mots (67.14%)

Registre soutenu : 69 mots (32.85%)

Registre familier : 0 mot (0%)

Figure 13 :

Répartition du nombre des mots par registre de langue pour la conversation n°5.



Analyse individuelle des deux interlocuteurs :

- **Interlocuteur 1 (You) :**

– Registre courant : nombre de mots : 53 mots, (68.83%)

Exemples de vocabulaire : "promener", "amis", "bonjour", "plaisirs"

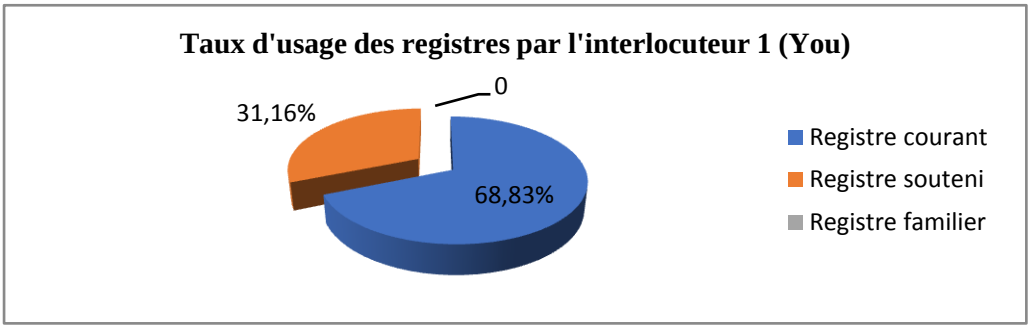
– Registre soutenu : nombre de mots : 24 mots, (31.16%)

Exemples de vocabulaire : "compagnie", "contempler", "médite"

– Registre familier : 0 mot, (0%).

Figure 14 :

Taux d'usage des registres de langue par l'interlocuteur 1 (You) dans la conversation n°5



- **Interlocuteur 2 (ChatGPT) :**

– Registre courant : nombre de mots : 88 mots, (66.16%)

Exemples de vocabulaire : "exactement", "respirer", "besoin", "façon"

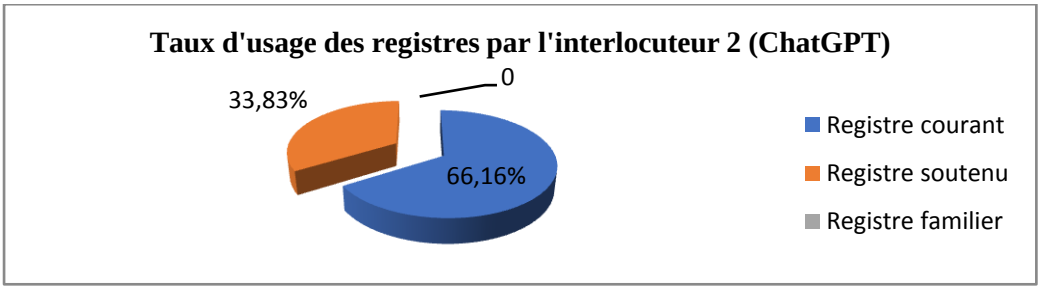
– Registre soutenu : nombre de mots : 45 mots, (33.83%)

Exemples de vocabulaire : "compagnie", "contempler", "médite"

- Registre familier : 0 mot, (0%).

Figure 15 :

Taux d'usage des registres de langue par l'interlocuteur 2 (ChatGPT) dans la conversation n°5.



Cette analyse nous montre que le registre dominant dans la conversation n°5 est le langage courant, avec un taux d'utilisation, environ 67.14% des mots exprimés. En revanche, Le registre soutenu est également présent mais à une moindre mesure, représentant environ 32.85% des mots. Tandis que le registre familier est totalement absent.

Faisant une sorte de comparaison de l'usage des deux registres entre les deux interlocuteurs, nous observons les statistiques suivantes :

- Interlocuteur 1 (You) : fait recours souvent au registre courant, avec un taux d'usage de (68.83%) avec une présence estimable du registre soutenu (31.16%).
- Interlocuteur 2 (ChatGPT) : emploie essentiellement un registre courant dominant (66.16%) avec un pourcentage significatif du registre soutenu (33.83%).

Cette répartition découvre que la conversation est formelle et courtoise, facilitant une communication transparente et respectueuse. ChatGPT démontre une capacité à utiliser un langage plus raffiné lorsque cela est approprié, enrichissant ainsi l'échange avec des éléments varies du registre soutenu.

Conversation n°6 :

Note : tous les exemples sont disponibles au niveau de l'Annexe F.

Cette analyse révèle les statistiques suivantes :

Nombre total de mots dans la conversation n° 6 : 428 mots.

Répartition des registres de langue :

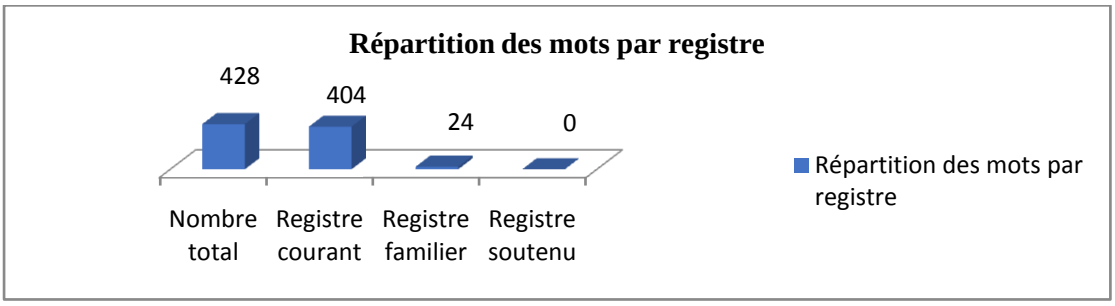
Registre courant : 404 mots (94.39%)

Registre familier : 24 mots (5.60%)

Registre soutenu : 0 mot (0%)

Figure 16 :

Répartition des mots par registre dans la conversation n°6.



Analyse individuelle des interlocuteurs :

- **Interlocuteur 1 (You) :**

Nombre total des mots utilisés par l'interlocuteur 2 : 119

- Registre courant : nombre de mots : 105 mots, (88.23%)

  Exemples de vocabulaire : "dialogue", "commencer", "examen"

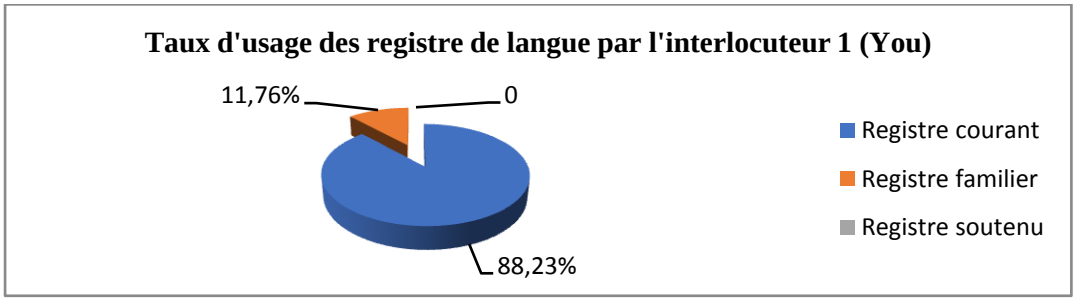
- Registre familier : nombre de mots : 14 mots, (11.76%)

  Exemples de vocabulaire : "piole", "pige", "fringues"

- Registre soutenu : 0 mot, (0%).

Figure 17 :

Taux d'usage des registres de langue par l'interlocuteur 1 (You).



- **Interlocuteur 2 (ChatGPT) :**

Nombre total des mots utilisés par interlocuteur 2 : 309

- Registre courant : nombre de mots : 299 mots, (96.76%)

  Exemples de vocabulaire : "stressant", "calmement", "instant", "soucis"

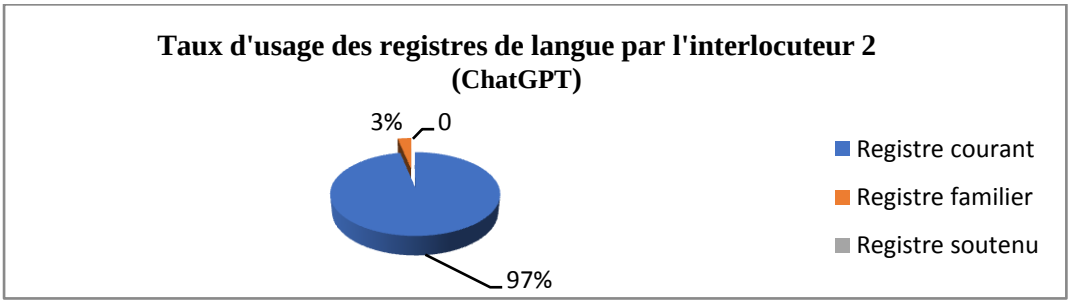
- Registre familier : nombre de mots : 10 mots, (3.23%)

  Exemples de vocabulaire : "potes", "papoter", "bosser"

- Registre soutenu : 0 mot, (0%).

Figure 18 :

Taux d'usage des registres de langue par l'interlocuteur 2 (ChatGPT



Les statistiques de cette analyse indiquent que le registre courant est majoritairement dominant également dans cette conversation, représentant environ 94.39% du total des mots. Le registre familier se manifeste avec un taux modeste de 5.60%, tandis qu'aucun mot du registre soutenu n'est présent. Cela s'explique par le fait que : L'interlocuteur 1 (You) : emploie un registre courant avec un taux majeur de (88.23%), et avec un usage faible du registre familier (11.76%). Et que l'interlocuteur 2 (ChatGPT) : s'articule dans son langage principalement sur le registre courant (96.76%) avec quelques empreints de registre familier (2.33%). Afin de maintenir certaine convivialité avec son interlocuteur, en asseyant d'adapter son langage avec les différents contextes et situations.

III.2.2-Discussion et interprétation des résultats obtenus par l'approche quantitative :

À ce stade de notre étude, nous allons effectuer une sorte d'évaluation diagnostique, de tous les résultats obtenus par l'analyse quantitative de notre corpus, à travers laquelle nous allons pouvoir vérifier notre première hypothèse. Soit en l'affirmant ou l'infirmant, en s'appuyant toujours, pour cela sur la méthode mixte pour examiner d'un côté les données

statistiques que nous avons eu lors de cette étape, et d'un autre côté, pour extraire les résultats finaux de notre recherche, ce qui nous permet de vérifier la première hypothèse mise préalablement.

Synthèse détaillée des résultats des conversations précédentes d'après l'approche quantitative :

Tableau 2 :

Tableau récapitulatif des taux de l'usage des registres de langue par les deux interlocuteurs.

	Registre de langue	CONVERSATIONS					
		1	**2**	**3**	**4**	**5**	**6**
You	**courant**	**97%**	**91%**	**73%**	**67.33%**	**68.83%**	**88.23%**
	soutenu	**0%**	**0%**	**2.7%**	**32.66%**	**31.16%**	**0%**
	familier	**3%**	**9%**	**25%**	**0%**	**0%**	**11.76%**
ChatGPT	**courant**	**97%**	**98.6%**	**89.06%**	**81.19%**	**66.16%**	**96.76%**
	soutenu	**3%**	**0.7%**	**1.6%**	**18.80%**	**33.83%**	**0%**
	familier	**0%**	**1.7%**	**9.31%**	**0%**	**0%**	**3.32%**

Le tableau récapitulatif représente les taux d'utilisation des trois registres de langue par les interlocuteurs 1 (You) et 2 (ChatGPT) dans différentes conversations. Voici une synthèse des données et des implications sur l'adaptation des interlocuteurs aux différents registres :

- **Interlocuteur 1 (You) :**

Registre courant : Usage variant dans la plupart des conversations, de 67.33% à 97%, basé majoritairement sur le registre courant (91% à 97%) peut influencer le choix de ce registre par l'interlocuteur 2.

Registre soutenu : Présence variable, allant de 0% à 32.66%. Une utilisation importante du registre soutenu par l'interlocuteur 1 (You) de 2.7% à 32.66% peut encourager ChatGPT à mener des échanges conversationnels en optant ce registre.

Registre familier : Présent dans certaines conversations, de 0% à 25%. L'emploie du registre familier par (You) peut influencer le ton de la conversation et l'adaptation de (ChatGPT) avec celui-ci.

- **Interlocuteur 2 (ChatGPT) :**

Registre courant : Utilisation, variante de 66.16% à 98.6%, de ce registre pousse (ChatGPT) de s'adapter principalement au registre courant de (You), reflétant une communication cohérente.

Registre soutenu : Présence variable, allant de 0% à 33.83%, avec une utilisation plus élevée du registre soutenu par (You) peut influencer l'adaptation de ChatGPT à un langage plus formel.

Registre familier : Présent dans certaines conversations, de 0% à 9.31%. L'usage de ce registre par (You) peut inciter ChatGPT à s'adapter à un ton plus décontracté.

L'examination des taux d'usage des registres de langue nous montre une prédominance du registre courant chez les deux interlocuteurs, avec des variations dans l'utilisation du registre soutenu et familier. L'usage de tel ou tel registre par You influence relativement le choix de registre de ChatGPT, qui s'adapte en conséquence pour maintenir une communication fluide et cohérente. Cette adaptation montre la capacité de ChatGPT à ajuster son langage en fonction des variables linguistiques introduites par les interlocuteurs, favorisant ainsi des échanges adaptés et enrichissants.

III.2.3- vérification de la première hypothèse :

Notre première hypothèse consiste à présumer que (ChatGPT) en tant qu'un acteur principal dans ces différents échanges conversationnels, utilise un langage standardisé dans ses communications avec toutes les catégories des interlocuteurs.

Cela reflète ce qui est dit littéralement dans la première hypothèse « ChatGPT communiquerait avec les divers internautes, en utilisant une langue standard ».

À partir de ce que nous avons eu par l'analyse du corpus, et les résultats obtenus, en adoptant l'étude quantitative, nous pouvons affirmer que (ChatGPT), l'un des logiciels interactionnels du traitement du langage humain, basé sur l'intelligence artificielle, utilise

majoritairement un langage standard. En considérant le registre courant le plus utilisé et le plus compréhensible par la majorité des internautes, comme une langue standard, avec l'implication relative, plus ou moins des deux autres registres (soutenu ou familier) en fonctions de, ces besoins communicatifs, et les exigences linguistiques de chaque situation de conversation. Tout cela nous amène à affirmer la validité de la première hypothèse.

III.3- Étude qualitative des données et vérification de l'hypothèse n° 2 :

L'étude qualitative vise à analyser les caractéristiques linguistiques, propres à chaque registre en s'articulant sur trois éléments essentiels, à savoir le lexique, la syntaxe et le style de langage utilisé.

Elle a pour but d'examiner la structure syntaxique, le vocabulaire, et le style d'échange des trois registres.

Analyse des registres de langue utilisés par les deux interlocuteurs :

III.3.1-Analyse des conversations :

Conversation n°1 : Cette conversation regroupe un étudiant de la faculté de médecine d'un côté et ChatGPT de l'autre côté, elle porte sur la spécialité de la médecine, notamment en ce qui concerne la difficulté de prononciation de certains termes scientifiques médicaux.

Note : tous les exemples sont disponibles au niveau de l'Annexe A.

- **Interlocuteur 1 (You)**

1-Registre courant

Exemples :"Bonjour, je m'appelle Amine", "Je suis étudiant à la faculté de médecine" (cf. capture d'écran n°1)

"Le problème c'est qu'il y a des termes scientifiques difficiles à prononcer même pour les vrais francophones !" (Cf. capture d'écran n°3)

Caractéristiques lexicales et syntaxiques :

a- Vocabulaire :

- Usage de mots courants et simples : "étudiant", "faculté", "problèmes", "francophones".
- Absence de jargon technique ou de termes trop spécifiques en dehors du contexte médical.

b- Syntaxe :

- Structure des phrases simples et directes. Ex : Je suis étudiant à la faculté de médecine. (cf. capture d'écran n°1)
- Utilisation de phrases déclaratives et interrogatives standards : "Comment savez-vous, si le français est ma langue maternelle ou n'est pas le cas".(cf. capture d'écran n°2)

c- Style :

- Ton informatif et factuel. Ex : une bonne astuce … (cf. capture d'écran n°4)
- Usage de phrases complètes et bien formées sans ellipses ni contractions.

2- Registre soutenu :

Exemples : "Je vois, de toute façon, je vous salue pour votre conscience envers ce domaine très sensible ainsi de votre esprit de responsabilité. Au revoir." (cf. Annexe A, capture d'écran n°10)

Caractéristiques lexicales et syntaxiques :

a- Vocabulaire :

- Termes plus complexes et formels : "conscience", "esprit de responsabilité", "domaine très sensible". (cf. capture d'écran n°10)
- Usage de mots précis et raffinés : "je vous salue", "de toute façon". (cf. capture d'écran n°10)

b- Syntaxe :

- Phrases longues et élaborées, souvent avec plusieurs propositions subordonnées. Ex : Le problème, c'est qu'il y des termes… (cf. capture d'écran n°3)
- Structures grammaticales normatives, utilisation de conjonctions et de phrases nominales. Ex : Oui, en fait, je fais mes études … (cf. capture d'écran n°1)

c- Style :

- Ton formel et respectueux.
- Emploi de formules de politesse : "je vous salue", "Au revoir".

3-**Registre familier :**

Caractéristiques lexicales et syntaxiques :

a-Vocabulaire :

- Implication des mots et expressions familières, contractions et argot. Ex : Dis donc. (cf. capture d'écran n°8)

b- **Syntaxe :**

- Phrases courtes, souvent avec des omissions de mots ou des contractions. Exemple : voilà, Vous voyez ! (cf. capture d'écran n°9)

d- Style :

- Ton décontracté et informel. Ex : Dis donc (cf. capture d'écran n°8)

- **Interlocuteur 2 (ChatGPT) :**

1-**Registre courant :**

Exemples :

- "Bonjour Amine ! Comment puis-je vous aider aujourd'hui ?" (cf. capture d'écran n°1)
- "Avez-vous des questions sur vos études ou sur la médecine en général ?" (cf. capture d'écran n°1)

Caractéristiques lexicales et syntaxiques :

a- Vocabulaire :

- Usage de mots accessibles et compréhensibles pour un large public : "domaine", "fascinant", "questions".

b- Syntaxe :

- Phrases directes et informatives. (Je suis étudiant à la faculté de médecine) (cf. capture d'écran n°1)
- Questions ouvertes pour encourager la conversation : "Avez-vous des questions sur vos études ou sur la médecine en général ?" (cf. capture d'écran n°1)

c- Style :

- Ton engageant et accueillant. (je suis content que cela vous aide !) (cf. capture d'écran n°4)
- Utilisation de ponctuation expressive : "C'est fantastique !". (cf. capture d'écran n°1)

2-Registre soutenu :

Exemples :

- "Je comprends tout à fait. Les termes scientifiques peuvent être complexes et difficiles à prononcer, même pour les locuteurs natifs." (cf. capture d'écran n°3)
- "Si vous avez des questions médicales, il est important de consulter un professionnel de la santé qualifié pour obtenir des conseils appropriés." (cf. capture d'écran n°9)

Caractéristiques lexicales et syntaxiques :

a- **Vocabulaire** :

- Termes spécifiques et techniques : "locuteurs natifs", "professionnel de la santé qualifié", "conseils appropriés".
- Vocabulaire riche et précis : "complexes", "difficiles à prononcer".

b- Syntaxe :

- Phrases complexes avec des subordonnées : "Les termes scientifiques peuvent être complexes et difficiles à prononcer, même pour les locuteurs natifs." (cf. capture d'écran n°3)
- Structures formelles et impersonnelles : "il est important de consulter". (cf. capture d'écran n° 8)

c- Style :

- Ton professionnel et sérieux. (vous avez raison) (cf. capture d'écran n°2)
- Utilisation de structures impersonnelles pour maintenir un ton neutre et informatif.(il est important de consulter un professionnel de la santé) (cf. capture d'écran n°6)

3-**Registre familier :**

Exemples :

Aucun exemple notable dans cette conversation spécifique.

- **Comparaison entre les deux interlocuteurs :**

Formes et styles :

- **Interlocuteur 1 (You) :**

- Principalement registre courant avec une inclinaison vers le soutenu lors des salutations finales.
- Utilise des phrases simples mais montre aussi une capacité à s'exprimer dans un registre plus formel lorsqu'il remercie ChatGPT.

Exemples de registre courant : "Je suis étudiant à la faculté de médecine."
Exemples de registre soutenu : "Je vous salue pour votre conscience envers ce domaine très sensible ainsi de votre esprit de responsabilité."

- **Interlocuteur 2 (ChatGPT) :**

- Principalement registre courant avec des éléments du registre soutenu pour expliquer des concepts ou donner des conseils.

- Utilise un vocabulaire précis et une syntaxe élaborée pour transmettre des informations
- Utilise un vocabulaire précis et une syntaxe élaborée pour transmettre des informations de manière claire et respectueuse.

-Exemples de registre courant : "Bonjour Amine ! Comment puis-je vous aider aujourd'hui ?"
-Exemples de registre soutenu : "Les termes scientifiques peuvent être complexes et difficiles à prononcer, même pour les locuteurs natifs."
Vocabulaire et expressions **:**

- **Interlocuteur 1 (You) :**
- Vocabulaire standard avec une progression vers des termes plus techniques (comme les termes médicaux).
- Expressions de politesse soutenues vers la fin de la conversation**.**
- **Interlocuteur 2 (ChatGPT) :**
- Vocabulaire riche et varié, alliant termes courants et techniques, avec une attention particulière à la clarté et à la précision.
- Utilisation fréquente de formules de politesse et de structures impersonnelles.

L'analyse qualitative révèle que les deux interlocuteurs utilisent principalement un registre courant, avec des incursions dans le registre soutenu lorsque la conversation touche à des sujets plus complexes ou nécessite des formules de politesse. Le registre familier est absent dans cette interaction, ce qui est cohérent avec le contexte formel et éducatif de la discussion. Cette utilisation des registres de langue illustre une communication respectueuse et adaptée aux sujets abordés, avec une attention particulière à la clarté, à la précision et à la politesse.

En comparant les deux interlocuteurs, nous observons que ChatGPT fait preuve d'une bonne capacité d'adaptation en utilisant principalement un registre courant, tout en intégrant quelques touches de registre soutenu. Son style est clair, direct et poli.

L'interlocuteur humain, quant à lui, se tient dans un registre courant tout au long de la conversation, sans écarts du registre. Son style est simple et naturel.

Conversation n° 2 : Deux interlocuteurs dans cette conversation, le premier nommé (You) et son homologue nommé (ChatGPT) se changent d'avis sur les bienfaits de voyage, où (You) voudrait faire une visite à ses proches en Canada.

Note : tous les exemples sont disponibles au niveau de l'Annexe B.

Analyse par Registre de Langue :

- **Interlocuteur 1 (You)**

Registre Familier :

Exemples : "coucou" (You)

"Ça va" (You) (cf. capture d'écran n°1)

Caractéristiques lexicales et syntaxiques :

a- Vocabulaire :-

- Utilisation de salutations informelles telles que "coucou"
- Termes simples et courants, tels que "ça va".

b- Syntaxe : -

- Phrases courtes et directes, sans structures complexes. Ex : Comment allez-vous ? (cf. capture d'écran n°1)
- Syntaxe relâchée et peu formelle. Ex : Alors, propose un sujet pour discuter. (cf. capture d'écran n°2)

d- Style :

- Ton convivial et décontracté
- Absence de formules de politesse complexes.

Registre Courant :

Exemples :

- "comment allez-vous ? (cf. capture d'écran n°1)
- "pour discuter" (cf. capture d'écran n°1)
- "Alors propose un sujet pour discuter" (cf. capture n° 2)
- "Oui, je voulais partir au Canada chez ma sœur" (cf. capture n° 2)
- "elle habite dans le côté francophone, précisément Montréal." (cf. capture n° 3)

Caractéristiques Lexicales et syntaxiques :

a- Vocabulaire : - Utilisation de termes courants et accessibles, tels que "discuter", "proposer", "partir", "habite".

b- Syntaxe : Phrases complètes et bien structurées. Ex : Je voulais partir au Canada. (cf. capture d'écran n°2)

-Utilisation de questions ouvertes pour engager la conversation. Ex : Comment allez-vous ? (cf. capture d'écran n°1)

c- Style : Ton poli et professionnel. Ex : Comment allez-vous ?

-Formules simples et directes, adaptées à une conversation quotidienne. Ex : Alors propose un sujet pour discuter. (cf. capture n° 2)

Registre soutenu : pas des marques de registre soutenu.

- **Interlocuteur 2 (ChatGPT) :**

Registre Familier :

Exemples : "coucou ! "ça va bien, merci ! (cf. capture d'écran n°1)

Caractéristiques Lexicales et syntaxiques :

a- Vocabulaire :

- Utilisation de salutations informelles telles que "coucou".
- Langage simple et direct, facilement compréhensible. Ex : ça va bien. ! (cf. capture d'écran n°1)

b- Syntaxe :

- Phrases courtes et directes, faciles à suivre. Ex : coucou, ça va
- Utilisation de questions simples pour engager la conversation.

c- Style :

- Ton convivial et accessible. Ex : coucou
- Style conversationnel qui met à l'aise l'interlocuteur. Ça va bien

Registre Courant :

Exemples :

- "Parfait ! De quoi aimeriez-vous discuter ? Avez-vous des sujets spécifiques en tête ou des questions particulières ?" (cf. capture d'écran n°2)
- "D'accord, parlons de voyages ! Avez-vous déjà visité un endroit qui vous a particulièrement marqué ou y a-t-il un pays que vous rêvez de découvrir ?" (cf. capture d'écran n°2)
- "C'est une excellente idée, le Canada est magnifique ! Quelle partie du Canada habite votre sœur ? Chaque région a tellement à offrir, que ce soit les montagnes et les forêts en Colombie-Britannique ou les villes culturellement riches comme Montréal et Québec." (cf. capture d'écran n°3)

Caractéristiques Lexicales et syntaxiques :

a- Vocabulaire :

- Utilisation de termes courants et descriptifs, tels que "parfait", "excellent", "magnifique", "culturellement riches".
- Vocabulaire adapté à une conversation informelle mais détaillée, avec des descriptions enrichissantes.

b- Syntaxe :

- Phrases complètes et structurées. (cf. capture d'écran n°2, 3)
- Utilisation de questions ouvertes pour approfondir la discussion. Ex : Avez-vous déjà visité un endroit… (cf. capture d'écran n°2)
- Syntaxe claire et fluide, facilitant la compréhension et la réponse.

c- Style :

- Ton poli et respectueux, adapté à une conversation entre adultes. Ex : Parfait ! de quoi aimeriez-vous discuter ? (Cf. capture d'écran n°2)
- Style informatif et engageant, visant à maintenir l'intérêt de l'interlocuteur. Ex : C'est une excellente idée … (cf. capture d'écran n°3)

Registre soutenu :

Exemple : De quoi aimeriez-vous discuter ?

Caractéristiques lexicales et syntaxiques :

a- Vocabulaire :

- Utilisation de tournures élaborées et vocabulaire plus raffiné. (fantastique)

b- Syntaxe

- Phrases plus complexes et formelles. Ex : c'est une excellente idée. Le Canada est magnifique. (cf. capture d'écran n°3)

c- Style :

- Ton professionnel et sérieux. Ex : Avez-vous des sujets spécifique … (Cf. capture d'écran n°2)

En appliquant une sorte de comparaison entre les deux interlocuteurs, nous constatons que le premier interlocuteur (You) utilise un langage courant, direct et simple, adapté à une conversation informelle, avec l'intégration de quelques termes familier. En revanche,

ChatGPT se substitut entre un langage courant et des éléments de registre soutenu, montrant une capacité d'adaptation et une richesse de vocabulaire.

Conversation n°3 : les deux interlocuteurs (You) et (ChatGPT) parlent de l'importance de la famille et qu'il faut donner plus de temps à ses proches, en passant plus de temps avec eux.

Note : tous les exemples sont disponibles au niveau de l'Annexe C.

Registre Courant :

a- Vocabulaire : Termes courants et familiers utilisés dans la vie quotidienne.

Exemples : Salut, va, super, profite bien, moment, famille. (Cf. capture d'écran n° 1),

b- Syntaxe : Phrases simples et directes sans constructions complexes.

Exemples : "Ça roule ! Qu'est-ce que tu fais de beau aujourd'hui ? (Cf. capture d'écran n° 1),

c- Style : Informel, convivial et proche du langage parlé.

Exemple : c'est génial, à plus tard (cf. capture d'écran n° 4, 8)

Registre Familier :

a- Vocabulaire :

- Termes familiers et décontractés, parfois argotiques.

Exemples : kiffe, potes, frics, caisse (cf. capture d'écran n° 1, 5,4,),

b- Syntaxe :

- Phrases plus relâchées et informelles, avec des racco
- Usage des abréviations : exemple : bcp (beaucoup)

Exemple : J'passe pas bcp de temps avec, à cause du taf ! (Cf. capture d'écran n° 2),

c- Style :

- Décontracté, proche et reflétant une relation familière.

Exemple : On se rigole bien ensemble. (Cf. capture d'écran n° 1),

Registre Soutenu :

a- Vocabulaire :

- Termes plus élaborés et raffinés, parfois techniques.

Exemples : pittoresque, inoubliable, flexibilité, envisager, (cf. capture d'écran n° 3, 5, 8),

b- Syntaxe :

- Phrases plus complexes et formelles, avec une structure élaborée.

Exemple : "Peut-être que vous pourriez planifier quelques activités spéciales pour profiter pleinement de votre temps ensemble lorsque vous êtes libre." (Cf. capture d'écran n° 2),

c- **Style** :

- Poli, distingué et utilisé dans des contextes plus formels.

Exemple : "C'est génial de passer du bon temps ensemble et de rigoler en famille." (Cf. capture d'écran n° 2),

Analyse individuelle des interlocuteurs :

- **Interlocuteur 1 (You) :**

a- **Vocabulaire :** Utilisation de termes familiers et décontractés.

b- **Syntaxe :** Phrases informelles et relâchées.

c- **Style :** Décontracté, proche et reflétant une relation familière.

- **Interlocuteur 2 (ChatGPT) :**

a- **Vocabulaire :** Termes plus élaborés et techniques.

b- **Syntaxe :** Phrases complexes et formelles.

c- **Style :** Poli, distingué et adapté à des contextes variés.

L'interlocuteur (You) emploie un langage varié, passant du courant au familier en fonction du contexte de l'échange. Tandis que l'interlocuteur (ChatGPT) démontre une adaptabilité en utilisant un langage courant, mais, en réaction, il intègre des éléments des deux registres, familier et soutenu pour enrichir la conversation.

Cette analyse met en lumière la diversité des registres de langue utilisés dans la conversation n° 3, montrant la capacité de ChatGPT à s'adapter aux variations linguistiques et à maintenir une communication fluide et naturelle avec l'utilisateur**.**

Conversation n°4 : le dialogue de cette conversation porte sur la lecture entre (You) qui se nomme également (Alain) qui considère la lecture comme un moyen central pour se cultiver, et ChatGPT qui partage le même opinion avec lui.

Note : tous les exemples sont disponibles au niveau de l'Annexe D.

- **Interlocuteur 1 (You) :**

Registre courant :

a- Vocabulaire :

- Termes usuels et familiers : "Bonsoir", "lecture", "romans", "philosophie", "histoire" "fascinant", "plongé", "actuellement", "passe-temps", "culture", "horizons intellectuels", "nourriture de l'esprit"

b- Syntaxe :

- Phrases simples et directes : "En fait, je m'adonne à la lecture tant des romans français qu'anglais." (Cf. capture d'écran n°1)
- Structures standard : "Tout à fait, actuellement je suis plongé dans la lecture du roman fascinant, Germinal, d'Émile Zola." (Cf. capture d'écran n°2)

c- Style :

- Informel et naturel : "Tout à fait, actuellement je suis plongé dans la lecture du roman fascinant, Germinal, d'Émile Zola."
- Poli et respectueux : "Je vous remercie cordialement pour cette échange enrichissant !" (Cf. capture d'écran n°8)

Registre soutenu :

a- Vocabulaire :

- Termes recherchés et élaborés : "croissance personnelle", "nourriture de l'esprit", "horizons intellectuels"

b- Syntaxe :

- Phrases complexes et élaborées : "Quant à moi, la lecture ne constitue pas seulement un passe-temps, mais plutôt un moyen d'approfondir ma culture et d'élargir mes horizons intellectuels." (Cf. capture d'écran n°5)
- Structures plus formelles : "Avez-vous des sujets spécifiques que vous aimez explorer dans vos lectures pour approfondir votre culture et élargir vos horizons intellectuels ?" (Cf. capture d'écran n°2)

c- Style :

- Soutenu et distingué : "On pourrait avancer que la lecture est la nourriture de l'esprit." (Cf. capture d'écran n°6)
- Poli et respectueux : Je vous remercie cordialement pour cette échange enrichissant !

- **Interlocuteur 2 (ChatGPT) :**

Registre courant :

a- **Vocabulaire :**

- Termes usuels et familiers : "fantastique", "capturer", "nuances", "fenêtre", "questions", "passe-temps", "culture", "horizons intellectuels"

b- **Syntaxe :**

- Phrases simples et directes : "C'est un éventail de lectures vraiment varié et enrichissant !" (Cf. capture d'écran n°1, 2)
- Structures standard : "Je suis ici pour vous aider avec toutes vos questions !" (Cf. capture d'écran n°5)

c- **Style :**

- Informel et naturel : "C'est un éventail de lectures vraiment varié et enrichissant !"
- Poli et respectueux : "Je vous remercie cordialement pour cette échange enrichissant!"

Registre soutenu :

a- Vocabulaire :

- Termes recherchés et élaborés : "pertinente", "intellectuels", "stimule", "équilibré"

b- Syntaxe :

- Phrases complexes et élaborées : "La lecture nourrit l'esprit de la même manière que la nourriture nourrit le corps." (Cf. capture d'écran n°7)
- Structures plus formelles : "La lecture offre effectivement une fenêtre sur le monde et sur les pensées des autres." (Cf. capture d'écran n°5,6)

c- Style :

- Soutenu et distingué : "La lecture nourrit l'esprit de la même manière que la nourriture nourrit le corps."
- Poli et respectueux : "La lecture offre effectivement une fenêtre sur le monde et sur les pensées des autres."

Utilisateur (You) : Utilise un registre courant, avec quelques éléments de registre soutenu, avec vocabulaire spécifique lié à la lecture et à la culture. Son style est naturel et direct. De l'autre côté, (ChatGPT) préfère changer entre un registre courant en impliquant de temps en temps des mots de registre soutenu. Son langage est poli et respectueux, avec des réactions et réponses pertinentes. En plus une capacité d'adaptation aux deux registres.

Nous pouvons dire que chaque interlocuteur présente une prédominance du registre courant, adapté à une discussion sur la lecture et la littérature. (You) utilise des termes plus particuliers liés à la lecture et à l'élargissement des horizons intellectuels, tandis que ChatGPT maintient un langage courant, offrant des réponses pertinentes. L'adaptation des interlocuteurs aux différents registres de langue montre une communication flexible et engageante, favorisant un échange bénéfique et enrichissant sur le thème de la conversation

.**Conversation n°5 :** l'objet de discussion dans cette conversation est la promenade dans la forêt, et les bienfaits de ce genre d'activité naturelle sur la santé.

Note : tous les exemples sont disponibles sur l'Annexe E.

- **Interlocuteur 1 (You) :**

Registre courant :

a- **Vocabulaire :** Usage de termes courants et compris et exprimables tels que "Bonjour", "portez-vous", "médite", "plaisirs", "joie".

Exemple : "Je me porte bien, merci." (Cf. capture d'écran n°1)

b- **Syntaxe** : Phrases simples et directes, exprimant des pensées de manière claire et précise.

Exemple : "comment vous portez- vous aujourd'hui." (Cf. capture d'écran n°1)

c- **Style** : élégant et respectueux, exprimant la gratitude et la sympathie.

Exemple : "Je vous exprime ma gratitude infinie." (Cf. capture d'écran n°2)

Registre soutenu :

a- Vocabulaire : Utilisation de termes plus recherchés et élaborés comme "médite", "particulièrement", "apprécier", "fraîcheur", "contempler", "beaux paysages".

Exemple : "Justement, apprécier la fraîcheur de l'air et contempler de beaux paysages sont des plaisirs qui ajoutent à la joie."

b- **Syntaxe :** Phrases plus complexes et élaborées, montrant une réflexion approfondie.

Exemple : "Justement, apprécier la fraîcheur de l'air et contempler de beaux paysages sont des plaisirs qui ajoutent à la joie." (Cf. capture d'écran n°2)

c- **Style :** Soutenu et réfléchi, exprimant une appréciation nuancée de la nature.

Exemple : "Justement, apprécier la fraîcheur de l'air et contempler de beaux paysages sont des plaisirs qui ajoutent à la joie."

- **Interlocuteur 2 (ChatGPT) :**

Registre courant :

a- **Vocabulaire :** Termes du langage courant et engageant comme "Bonjour", "génial", "apaisant", "merveilleuse".

Exemple : "Une promenade en forêt semble être une excellente idée pour se ressourcer et se connecter avec la nature." (Cf. capture d'écran n°2)

b- **Syntaxe :** Phrases peu complexes, directes et empathiques, adaptées pour encourager et soutenir l'interlocuteur.

Exemple : "Je suis honoré de pouvoir vous accompagner, même virtuellement." (Cf. capture d'écran n°4)

c- **Style :** agréable et motivant, avec des réponses chaleureuses et des encouragements.

Exemple : "Je suis honoré de pouvoir vous accompagner, même virtuellement."

Registre soutenu :

a- Vocabulaire : Utilisation de termes plus élaborés et spécifiques comme "honoré", "accompagner", "virtuellement".

Exemple : "Je suis honoré de pouvoir vous accompagner, même virtuellement." (Cf. capture d'écran n°4**)**

b- Syntaxe : Phrases plus élaborées et formelles, démontrant une certaine recherche esthétique dans la construction du discours.

Exemple : "Exactement ! Rien de tel que de respirer l'air frais et de s'émerveiller..." (cf. capture d'écran n°3)

c- Style : Élégant et distingué, avec une tonalité respectueuse et engageante.

Exemple : "Je suis honoré de pouvoir vous accompagner, même virtuellement.

(You) emploie un langage courant et sympathique, exprimant ses pensées de manière claire et respectueuse. Son vocabulaire est axé sur la nature et le bien-être, reflétant une

attitude positive et ouverte. Tandis que (ChatGPT) préserve un langage courant et empathique, adaptant ses réponses pour motiver et soutenir son homologue. Son style est sympathique et engageant, favorisant une communication agréable et bienveillante.

Cette analyse met en évidence les nuances des comportements linguistiques de chaque interlocuteur, soulignant leur capacité à communiquer de manière efficace et empathique dans un registre courant. Les échanges sont empreints d'amabilité, de clarté et de soutien mutuel, créant une atmosphère positive et bienveillante dans la conversation.

Conversation n°6 : Il s'agit là d'un échange conversationnel entre (You) d'un côté, qui parle des ses préparations pour passer les examens, et (ChatGPT) de l'autre côté, qu'il essaye de l'encourager en lui donnant des conseils.

Note : Tous les exemples sont disponibles sur l'Annexe F.

- **Interlocuteur 1 (You) :**

Registre familier :

a- **Vocabulaire** : Utilisation de termes familiers et argotiques comme "j'boss", "j'méclate", "j'vais m'planter", "j'ai l'trac", "j'suis crevé".

b- **Syntaxe** :

– Phrases relâchées et informelles, avec des raccourcis et une structure simplifiée.

Exemple : Merci, t'es cool mon pote. (Cf. capture d'écran n°3).

– L'usage de l'aphérèse, ce qui est une figure de style qui consiste à la suppression d'une ou plusieurs lettres ou syllabes au début d'un mot.

Exemple : j'boss, j'crois, t'es cool, j'vais… (cf. capture d'écran n°1,3).

– La chute de "ne " de la négation.

Exemple : j'm'éclate pas. Au lieu dire : je ne m'éclate pas. (Cf. capture d'écran n°1,).

c- **Style :**

Décontracté et familier, avec une tonalité plus proche de la conversation quotidienne entre amis.

Exemple : Ouais, tu m'étonnes. (Cf. capture d'écran n°4).

Registre courant :

a- Vocabulaire : Termes usuels et polis comme "dialogue", "préoccupe", "examens", "demain", "planter", "pige pas tout", "trac", "crevé".

b- Syntaxe : Phrases simples et directes, exprimant des idées de manière claire et compréhensible.

Exemple : c''est moi qui dois commencer. (Cf. capture d'écran n°2).

c- Style :

- Conviviale et respectueux, avec une communication standardisée.

Exemple : génial on peut souffler maintenant. (Cf. capture d'écran n°3).

- Langage peu formel, et l'emploie de tutoiement (usage du pronom "tu " au lieu du pronom "vous ").

Exemple : Qu'est-ce que tu fou dans ta piole ? (Cf. capture d'écran n°2).

- **Interlocuteur 2 (ChatGPT) :**

Registre courant :

a- Vocabulaire : Termes courants et polis comme "aider", "discuter", "préoccupe", "moment", "recharger", "batteries", "revoir", "connaissances", "stressant", "veille", "examen", "rester calme".

b- Syntaxe : Phrases simples et empathiques, adaptées pour encourager et soutenir l'interlocuteur.

Exemple : Je suis là pour toi, toujours prêt à discuter. (Cf. capture d'écran n°4).

c- Style : Gentil et motivant, avec des réponses chaleureuses et des encouragements.

Exemple : ça sonne super ! Profite bien ton temps … (cf. capture d'écran n° 3, 4).

Registre familier :

a- Vocabulaire : Usage de termes familiers et argotiques comme : mec, papoter, potes…

b- Syntaxe :

- Phrases simples et informelles, avec des raccourcis et une structure simplifiée.

Exemple : ah, ça roule ! Rien spécial. Quoi de neuf… (cf. capture d'écran n° 2).

- L'usage de l'aphérèse, ce qui est une figure de style qui consiste en la suppression d'une ou plusieurs lettres ou syllabes au début d'un mot.

Exemple : t'es là pour quoi aujourd'hui ?

c- Style :

- Joyeux et informel.

Exemple : pas de souci mec. (Cf. capture d'écran n° 3).

- Usage de tutoiement (usage du pronom "tu " au lieu du pronom "vous ").

Exemple : je suis prêt quand tu l'es ! (Cf. capture d'écran n° 2).

L'interlocuteur 1 (You) emploie un registre familier avec des termes argotiques et une syntaxe relâchée, reflétant une conversation informelle entre amis, en créant au même temps une atmosphère décontractée et amicale. En revanche l'interlocuteur 2 (ChatGPT) préfère l'utilisation du registre courant tout au long de la conversation, adaptant ainsi, son langage pour garder le fil de conversation avec homologue.

III.3.2-Discussion et interprétation des résultats obtenus par l'approche qualitative :

L'analyse qualitative de ces conversations entre les interlocuteurs et ChatGPT dévoile une utilisation prédominante du registre courant, avec des incursions dans le registre soutenu pour aborder des sujets complexes où nécessitent des formules de politesse. L'usage de registre familier est limité dans ces interactions, reflétant un contexte formel et éducatif.

Type de vocabulaire :

- Les interlocuteurs utilisent un vocabulaire courant, avec des termes spécifiques qui se variés en fonction des contextes et des exigences de la communication.
- ChatGPT intègre des éléments de registre soutenu et familier, pour enrichir la conversation d'une part et de maintenir un ton de convivialité avec ses interlocuteurs, d'une autre part.

Structures syntaxiques :

- Les phrases sont simples et directes, exprimant des idées de manière claire et compréhensible. Souvent manque de respect des règles syntaxiques dans le registre familier.
- ChatGPT ajuste ses réponses afin de s'intégrer et tenir le fil de la communication avec son homologue.

Style du langage :

- Les interlocuteurs maintiennent un style direct et respectueux, avec une attention particulière à la clarté, à la précision et à la politesse.
- ChatGPT démontre une capacité d'adaptation en utilisant un langage courant et empathique, favorisant une communication agréable et bienveillante.

ChatGPT dévoile une grande efficacité à interagir face aux différents registres de langue, en adaptant son langage pour maintenir une communication fluide et naturelle avec les utilisateurs d'internet. Sa capacité à intégrer des éléments de registre soutenu pour enrichir la conversation et à préserver un style poli et respectueux, en fait un interlocuteur efficace et engageant sans négliger totalement le registre familier.

II.3.3- vérification de la deuxième hypothèse :

Afin de vérifier notre deuxième hypothèse, nous nous sommes appuyés durant l'analyse du corpus sur trois critères essentiels :

- Le type de vocabulaire utilisé par chacun des deux interlocuteurs : cela indique quel type de registre de langue est présent dans la conversation, en indiquant la capacité de ChatGPT à déterminer le sens de chaque vocabulaire et de le classier d'après le type du registre.
- La structure syntaxique : le respect des règles de la syntaxe, et le bon usage de la langue est un signe de tel et tel registre.
- Le style du langage : ce dernier élément reflète la manière dont l'interlocuteur s'adresse à son homologue.

La deuxième hypothèse consiste à déterminer à quel point (ChatGPT) capable à interagir d'une manière fluide et efficace avec les interlocuteurs humains : « ChatGPT interagirait efficacement face au différents registres de langue ».

L'étude qualitative de toutes les conversations montrent une supériorité évidente du registre courant dans la plupart des conversations, avec des variations dans l'utilisation du registre soutenu et familier.

En effet, l'analyse prouve que (ChatGPT) s'adapte bien aux différents registres, créant une communication dynamique, transparente et efficiente. Elle démontre également une maitrise plus ou moins parfaite des sujets proposés par les interlocuteurs. Ce qui nous conduit à confirmer la validité de notre deuxième hypothèse.

Conclusion générale

Les logiciels interactifs de traitement automatique des langues naturelles sont essentiels pour l'interaction homme/machine en facilitant une communication souple et sans ambiguïté. La complexité de la langue naturelle pose des défis pour les logiciels intelligents, tels que ChatGPT, notamment en ce qui concerne les variations linguistiques et les registres de langue. Il est capital de comprendre si ChatGPT peut identifier et utiliser efficacement les vocabulaires et les caractéristiques linguistiques propres à chaque registre pour améliorer sa compréhension et sa production écrite. Cette réflexion nous a conduits à poser la question centrale suivante :

- Dans quelle mesure ChatGPT s'adapte-t-il aux variations diaphasiques (registre de langue) ?

Celle-ci peut être décomposée en deux questions secondaires :

1- Quels sont les aspects de la variation diaphasique (les registres de langue) dans des conversations écrites entre ChatGPT et des internautes ?

2- Quel est l'impact de l'interaction avec des internautes sur la faculté adaptative de ChatGPT aux variations diaphasiques ?

On a pu répondre à ces questions en mettant les deux hypothèses suivantes :

1- ChatGPT communiquerait avec des divers internautes en utilisant une langue standard.

2- ChatGPT interagirait efficacement face aux différents registres de langue.

L'objectif de cette étude est de savoir à quel point ChatGPT en tant que logiciel interactif autonome, est capable de mener des conversations intégrales avec des êtres humains en s'appuyant uniquement sur ses propres compétences linguistiques et en mesurant son efficacité et sa faculté adaptative face aux différentes variables langagières, plus précisément, face aux registres de langue.

Pour le choix méthodologique nous avons opté pour une étude quantitative et qualitative basée essentiellement sur la méthode mixte, qui comporte deux approches, l'une analytique propre à l'étude quantitative, basée sur des statistiques et qui nous a permis d'identifier le registre dominant dans chaque conversation. Et l'autre descriptive convenable à l'étude qualitative en mettant l'accent sur les caractéristiques linguistiques (lexicales et syntaxiques) de chaque registre, et leurs implications dans chaque conversation.

Après avoir analysé les données et interprété les résultats de notre corpus, nous avons constaté que, durant les six conversations, le registre courant était le plus utilisé, ce qui en fait

le registre dominant, avec un usage très limité, en fonction des contextes, des registres familier et soutenu notamment de la part de ChatGPT, malgré la présence considérable des deux autres registres employés par les interlocuteurs humains dans les différentes conversations. Cela signifie que ChatGPT recourt souvent à un langage plus général et plus standard dans ses communications, ce qui souligne la validité de la première hypothèse.

D' autre part, en ce qui concerne la deuxième hypothèse, les analyses qualitatives montrent que ChatGPT s'intègre parfaitement aux sujets proposés par ses homologues humains, sans aucune ambiguïté ni malentendu, ce qui valide cette deuxième hypothèse quant à son efficacité de l'interaction avec les différentes catégories d'interlocuteurs.

Pour favoriser la recherche sur l'adaptation de ChatGPT aux variations linguistiques, plusieurs approches complémentaires peuvent être explorées. Des études comparatives avec d'autres modèles d'IA permettraient d'évaluer leur capacité d'adaptation dans divers contextes. L'adaptation de ChatGPT à des langues moins courantes ou régionales pourrait également être étudiée pour une utilisation efficace dans un cadre multilingue. Explorer les implications culturelles de cette adaptation aiderait à comprendre comment les différences culturelles influent sur les interactions homme-machine. De plus, la promotion de l'utilisation pédagogique de ChatGPT pour améliorer l'apprentissage des langues et créer des contenus éducatifs personnalisés ouvrirait de nouvelles perspectives pour cette technologie.

Ces pistes de recherche pourraient enrichir la compréhension de l'adaptation de ChatGPT et ouvrir de nouvelles voies passionnantes pour l'intelligence artificielle et la linguistique en général.

Références bibliographiques

ActiaAI. (s. d.). *Articles citant les travaux de OpenAI dans le domaine de l'intelligence artificielle.* Récupéré sur https://www.actuia.com/acteur/openai/

Anselmet, A. (2022, 04 23). *tout savoir sur ChatGPT et son fonctionnement*. Consulté le 05 01, 2004, sur Conseils redaction web: https://www.conseils-redaction-web.fr/chatgpt

Aurélie, J. (2020, 11 05). Une brève introduction à l'intelligence artificielle. *M/S medecine sciences , 36* (11), p. https://doi.org/10.1051/medsci/2020189.

Baylon, C., & Mignot, X. (1991). *La Communication* (éd. Nathan).

Bergounioux, G. (s. d.). *Structures Syntaxiques : Le primat de la syntaxe.* Consulté le 04 12, 2024, sur universalis.fr: https://www.universalis.fr/encyclopedie/structures-syntaxiques/2-le-primat-de-la-syntaxe/

Bouterraka, S. (2023). *ntelligence Artificielle "Exploration de l'Utilisation de ChatGPT" Pour la création de chatbots éducatifs personnalisés Potentielles, Lmites, Perspectives)Mémoire de Master*. (U. M. V, Éditeur) Récupéré sur https://fr.slideshare.net/SalmaAB1/lintelligence-artificielle-exploration-de-lutilisation-de-chatgpt-pour-la-cration-de-chatbots-ducatifs-personnaliss-potentielles-limites-et-perspectives

Boyer, H. (1996, p. 36). *sociolinguistique, territoire et objets* (éd. Delachaux et Niestlé).

Chatgpt. (s. d.). *l'Interface interactionnelle de ChatGPT*. Récupéré sur https://chatgpt.com/?oai-dm=1

Douarre, C., & Rousseau, D. (2021, 10 11). *Introduction à l'apprentissage profond (deep learning) de l'intelligence artificielle.* Consulté le 04 20, 2024, sur CultureSciences

Physique - ISSN 2554-876X: https://culturesciencesphysique.ens-lyon.fr/ressource/IA-apprentissage-Rousseau.xml

Esteban, C. (2022). *Développer des machines parlantes : les chatbots à l'épreuve du marché [Thèse de Doctorat, Université de Toulouse 2]*. Récupéré sur https://theses.fr/2022TOU20052

Fishman, J. (1971). *Sociolinguistique, une briève introduction* (éd. Nathan).

Fuchs, C., Danlos, L., Lacheret-Dujour, A., Luzzati, D., & Victorri, B. (1993). *Linguistique et Traitements Automatiques des Langues.* Hachette.

Fuchs, C., François, J., Le Ny, J.-F., Nespoulous, J.-L., Payraube, A., Rouvert, A., et al. (2004). *la linguistique cognitive.* Ophrys.

Garmadi, J. (1981). *La Sociolinguistique* (éd. PUF).

https://www.structural-learning.com. (s.d.). Récupéré sur https://www.structural-learning.com

IBM. (s. d.). *Qu'est-ce qu'un réseau de neurones ?* Consulté le 05 01, 2024, sur https://www.ibm.com/fr-fr/topics/neural-networks

Labov, W. (1976). *Sociolinguistique* (éd. Minuit).

Luigisbox. (s.d.). *La Linguistique informatique.* Consulté le 04 15, 2024, sur www.luigisbox.fr: https://www.luigisbox.fr/glossaire-recherche/linguistique-informatique/

Main, P. (2023). *Chomsky's Theory [La Théorie de Chomsky]*. Consulté le 04 01, 2024, sur Structural Learning: https://www.structural-learning.com/post/chomskys-theory

Merah, E. (2022). *Le Langage, la Grammaire et l'Automate.* Pages Bleues.

Moureau, M.-L. (1997). *sociolinguistique, concepts de base* (éd. Mardaga).

Moussaoui, H. (2023, 07 05). *ChatGPT : La technologie qui révolutionne la communication humaine et suscite des débats éthiques et juridiques majeurs.* Consulté le 05 03, 2024, sur novencia: https://www.novencia.com/blog/datamarketing/chatgpt/

Nicholas, A., Damasio, A., Harari, Y. N., O'Neil, C., & Revel, N. (2018). *Nouvelle enquête sur l'intelligence artificielle.* Champs.

Notari, C. (2009). *Chomsky et l'ordinateur.* Interlangues.

Offord, M. (1990). *Varieties of Contemporary French [varieté du français contemporain]* (éd. Macmillan).

Poibeau, T. (s. d.). *Traitement Automatique des Langues.* Consulté le 03 26, 2024, sur universalis.fr: http://www.universalis.fr/encyclopedie/traitement-automatique-des-langues/

Praedictia.com. (s. d.). *Réseau de neurones artificiels: que'est-ce que c'est et à quoi ça sert ?* Récupéré sur https://praedictia.com/page/reseaux-de-neurones/quest-ce-quun-reseau-de-neurones.html

Registres des langue . (s. d.). Consulté le 2 15, 2024, sur Wikipedia: https://fr.wikipedia.org/wiki/Registres_de_langue_en_fran%C3%A7ais

Richard, R. (2019). *la linguistique informatique.* Consulté le 05 01, 2024, sur 20PM: https://24pm.com/117-definitions/297-linguistique-informatique

Robert, L. (s. d.). *dico en ligne*. Récupéré surhttps://dictionnaire.lerobert.com/guide/registres-delangue#:~:text=On%20distingue%20g%C3%A9n%C3%A9ralement%20trois%20registres,%2C%20didactique%2C%20rare%2C%20etc(s. d.). Consulté le 3 5, 2024,

sur Lingolia Logo: https://francais.lingolia.com/fr/atelier-decriture/les-registres-de-langue

Structural Learning. (s.d.). Récupéré sur https://www.structural-learning.com

Un prototype programmable pour concrétiser la machine de Turing. (s. d.). Consulté le 04 10, 2024, sur https://bib.umontreal.ca/citer/styles-bibliographiques/citer-oeuvres-images-style-apa?tab=5231311

Walte, H. (1994). *L'aventure des langues en occident.*

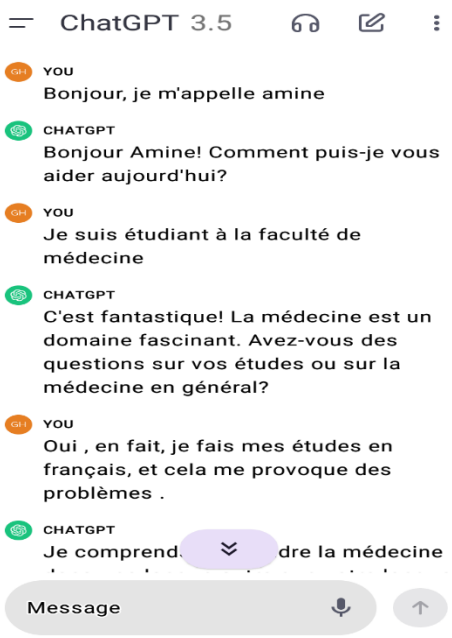
Annexes

Liste des sigles et acronymes :

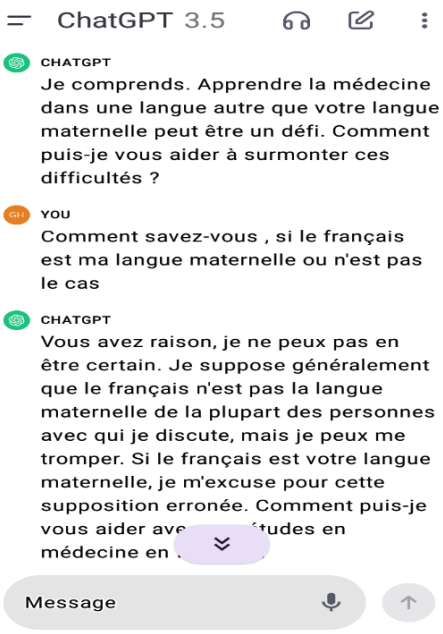
GGT	Grammaire Générative et Transformationnelle.
LC	Linguistique Cognitive.
LI	Linguistique informatique.
DL	Deep Learning (apprentissage profond).
NLP	Natural Language Processing.
TALN	Traitement Auomatique des Langues naturelles.
TAL	Traitement Automatique des langues.
MIT	Massachusetts Institute of Technology.
IA	Intelligence Artificielle.
Chat	Conversation.
GPT	Generative Pre-trained Transformer (transformateur Pré-entraîné Génératif)
APA	American Psychological Association

Annexe A

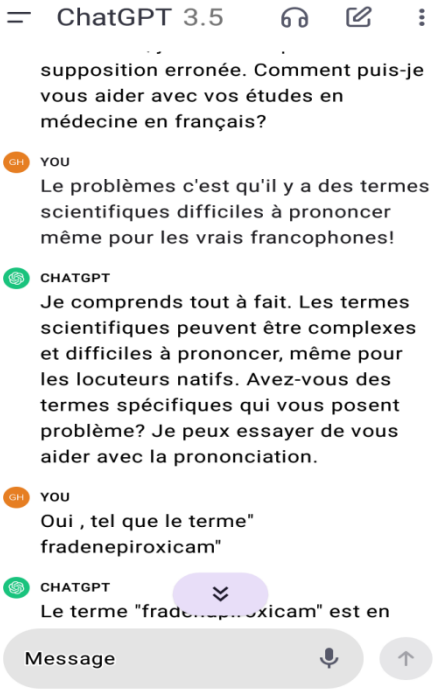
conversation écrite n° 1



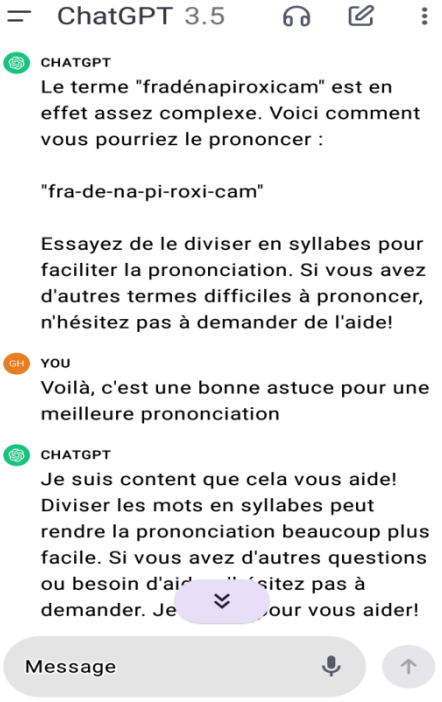
Capture d'écran n°1



Captured'écran n°2

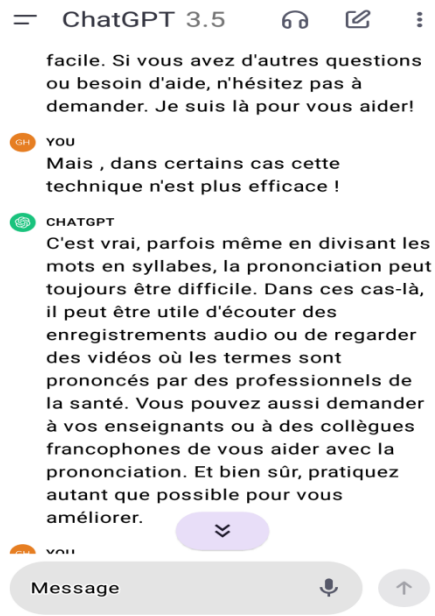


Capture d'écran n° 3

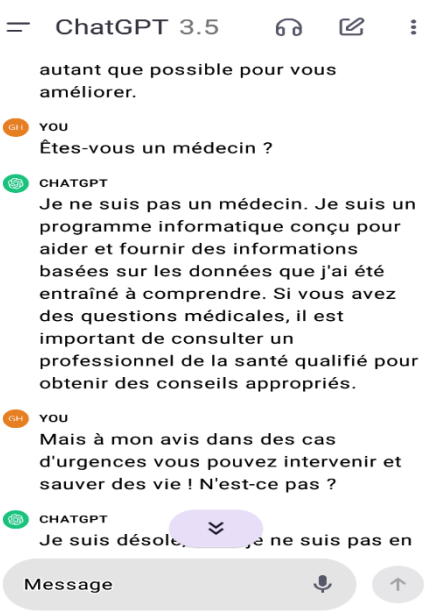


Capture d'écran n

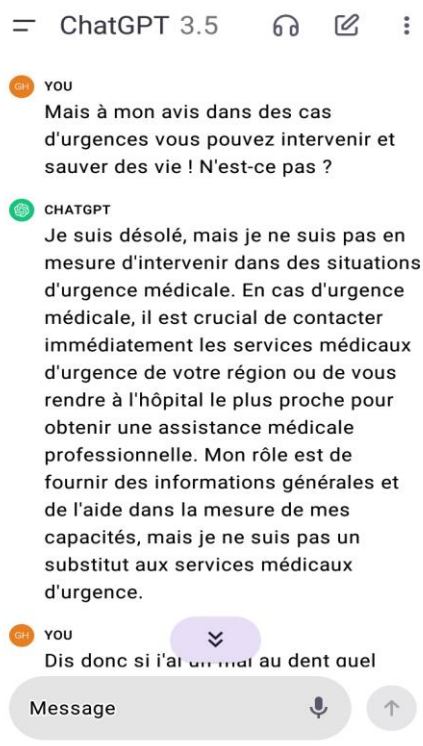
Suite à l'Annexe A



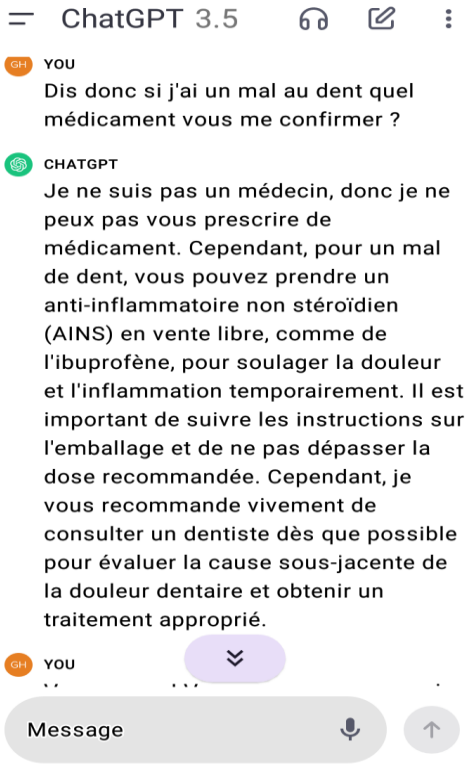
Capture d'écran n°5



Capture d'écran n°6

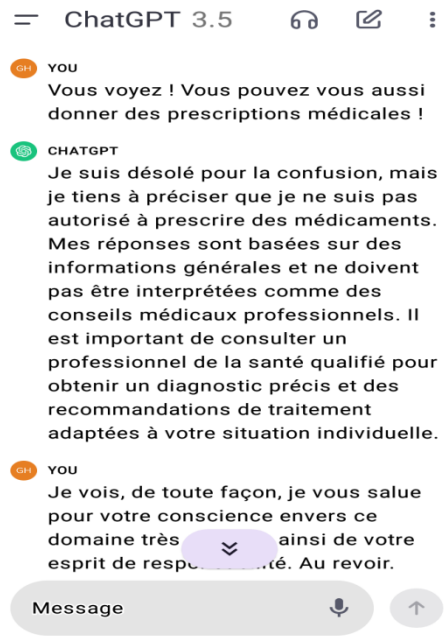


Capture d'écran n° 7

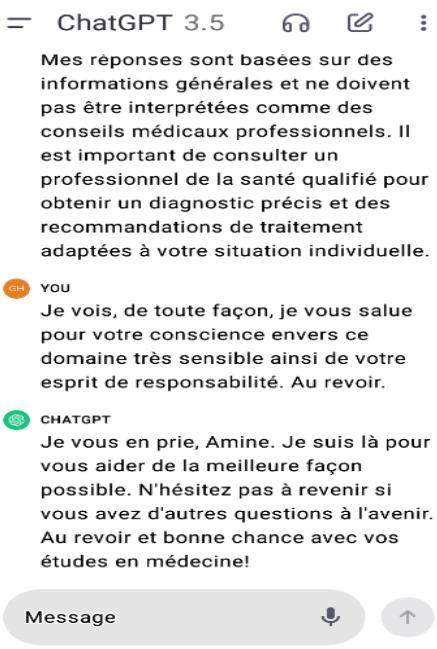


Capture d'écran

Suite à l'Annexe A :



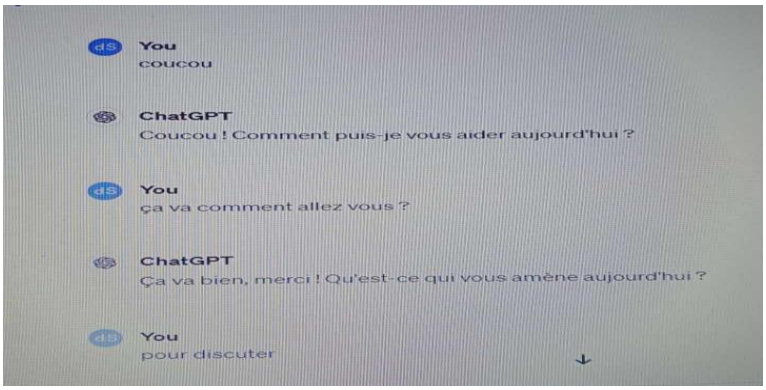
Capture d'écran n°9



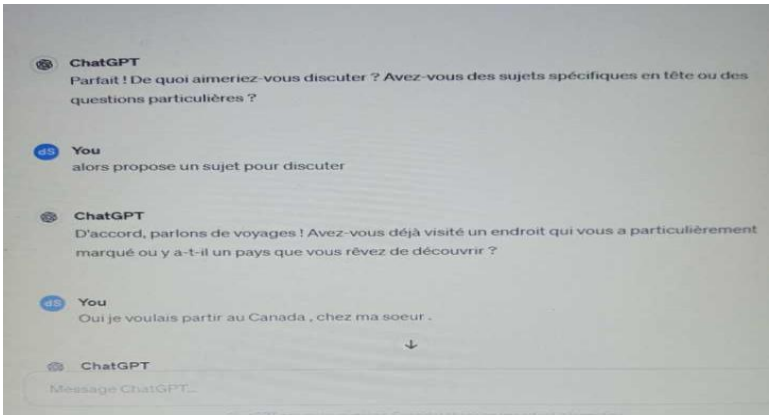
Capture d'écran n°10

Annexe B

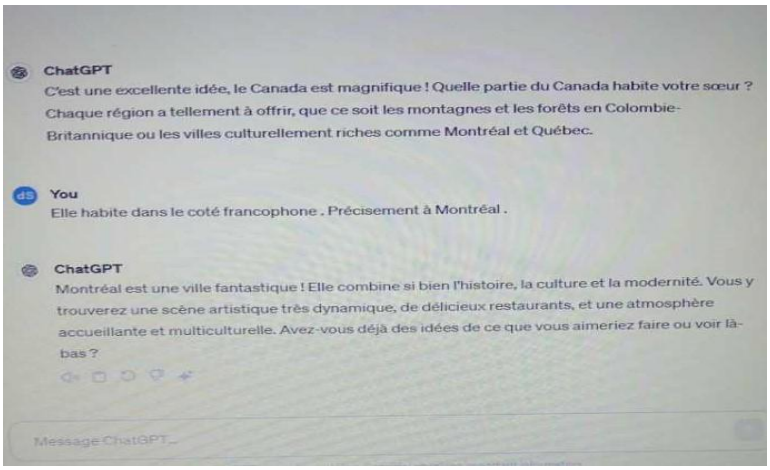
conversation écrite n°2



Capture d'écran n°1



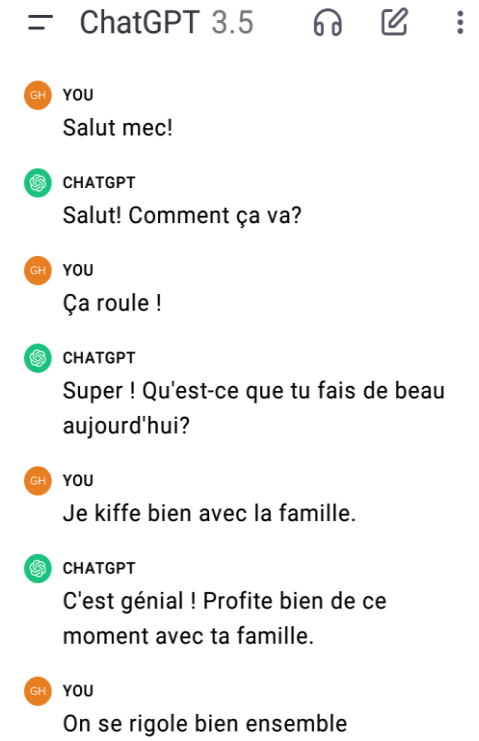
Capture d'écran n°2



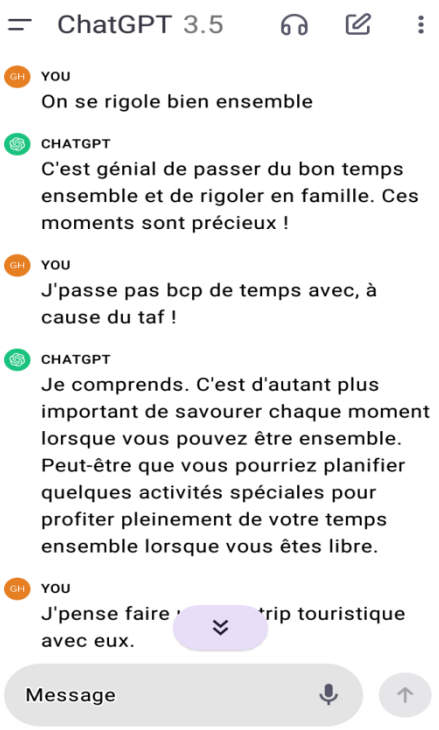
Capture d'écran n°3

Annexe C

conversation écrite n°3



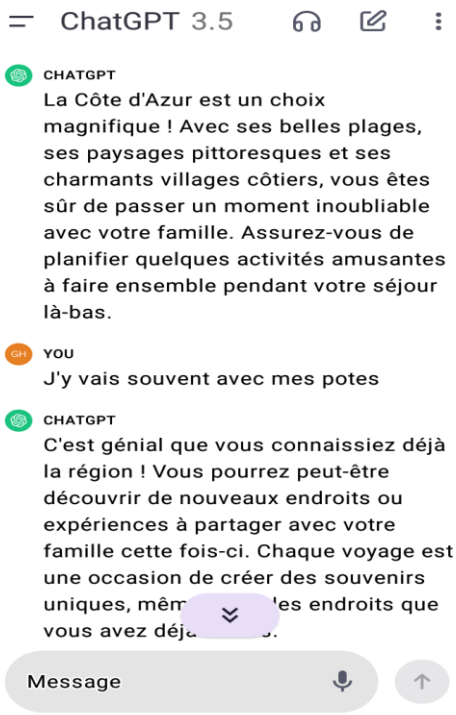
Capture d'écran n °1



capture d'écran n°2

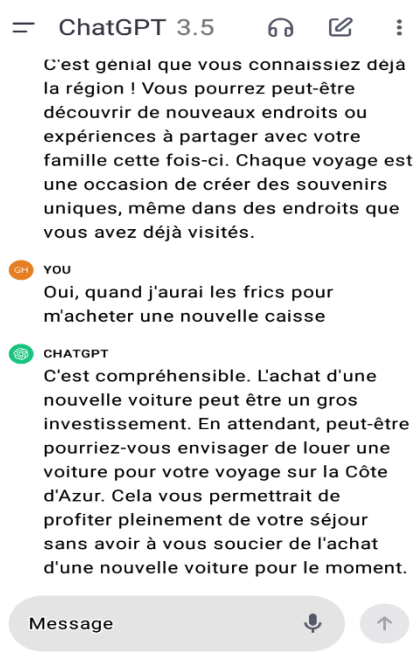


Capture d´écran n°3

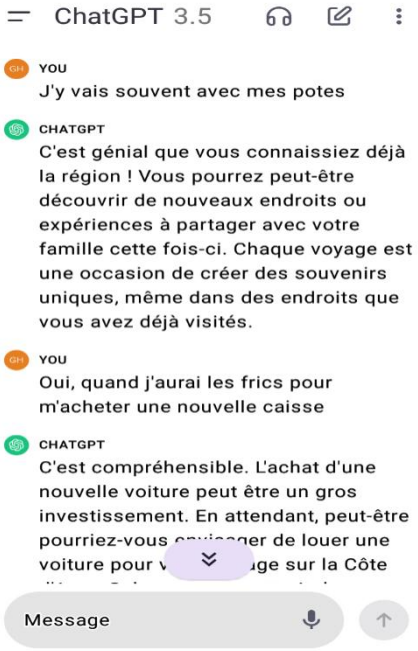


capture d'écran n°4

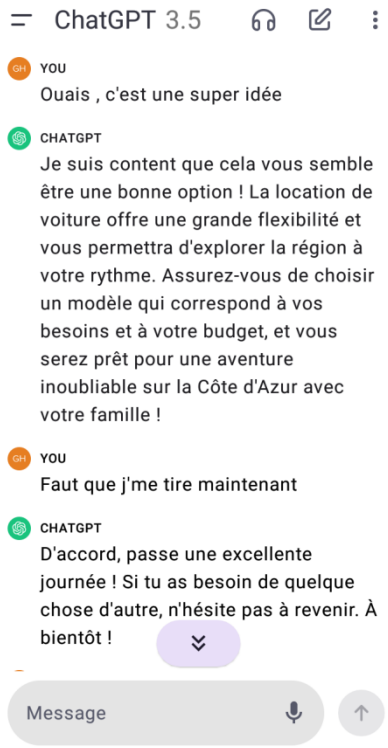
Suite à l'Annexe C



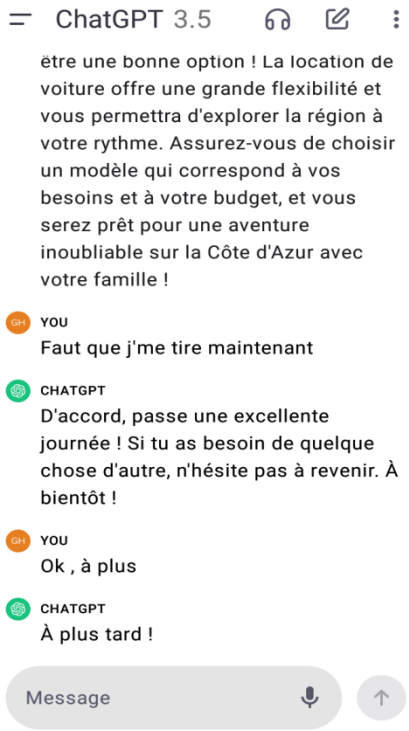
Capture d'écran n°5



capture d'écran n°6



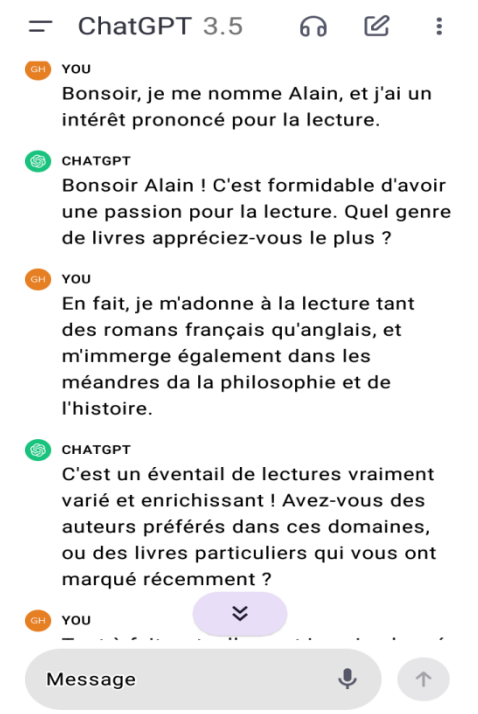
Capture d'écran n°7



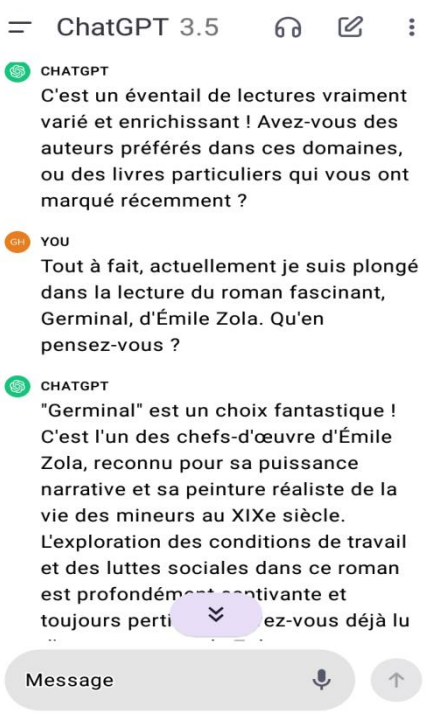
capture d'écran n°8

Annexe D

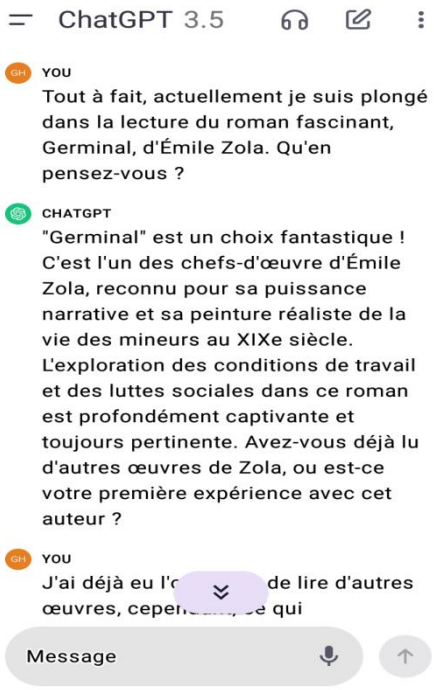
conversation écrite n°4



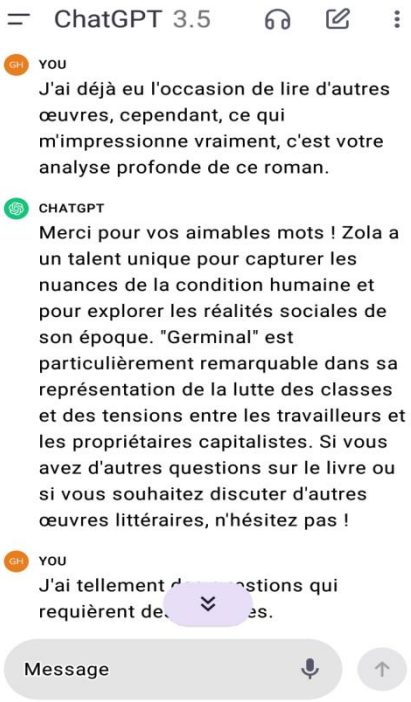
Capture d'écran n°1



Capture d'écran n° 2

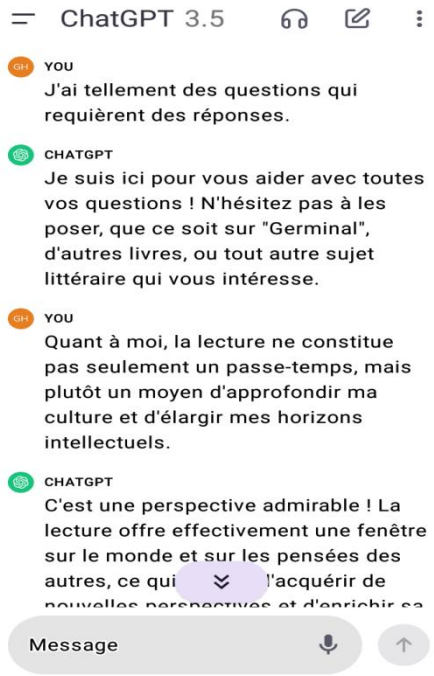


Capture d'écran n° 3

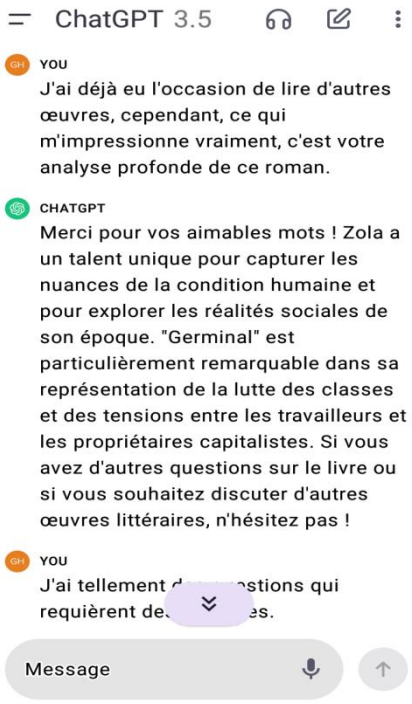


Capture d'écran n°4

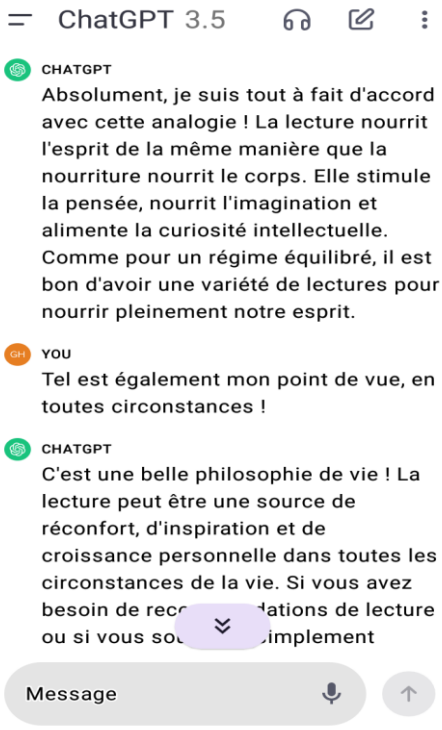
La suite à l'Annexe D



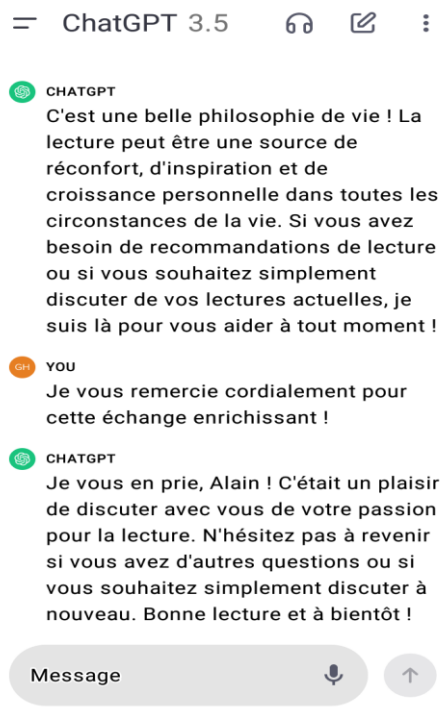
Capture d'écran n° 5



Capture d'écran n°6

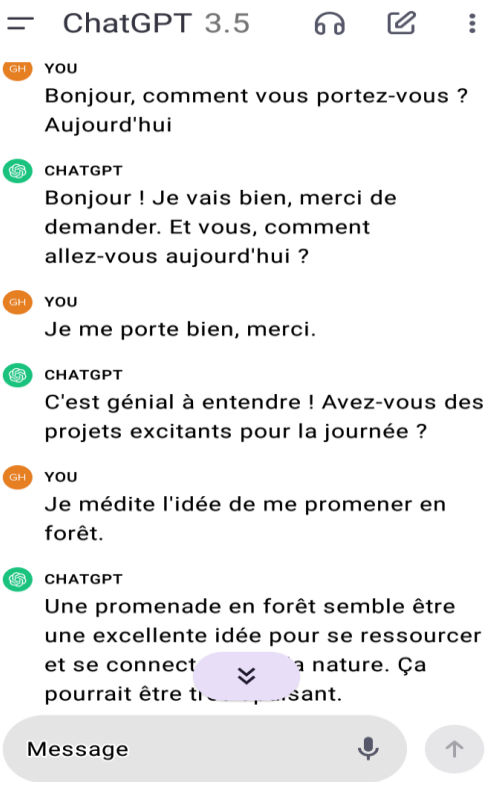


Capture d'écran n° 5

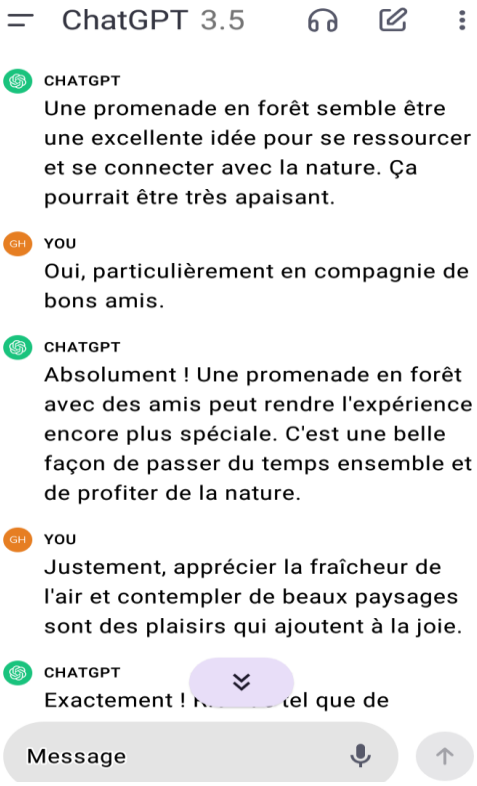


Capture d'écran n° 6

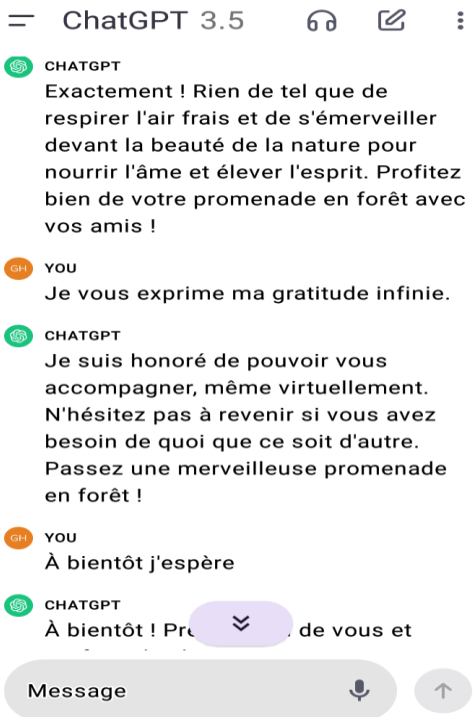
Annexe E
conversation écrite n°5



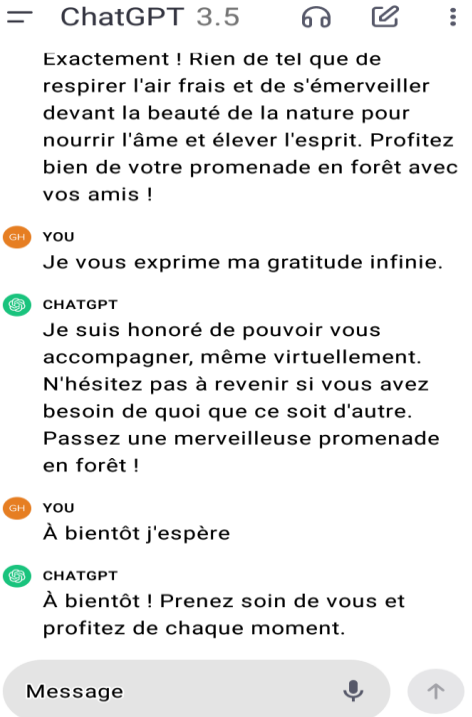
Capture d'écran n° 1



Capture d'écran n°2



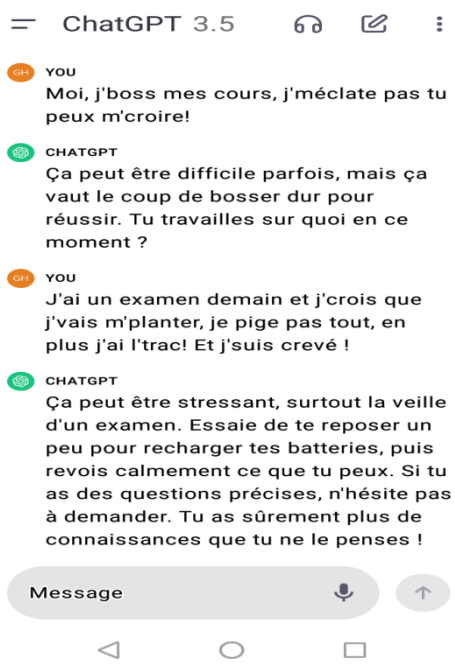
Capture d'écran n° 3



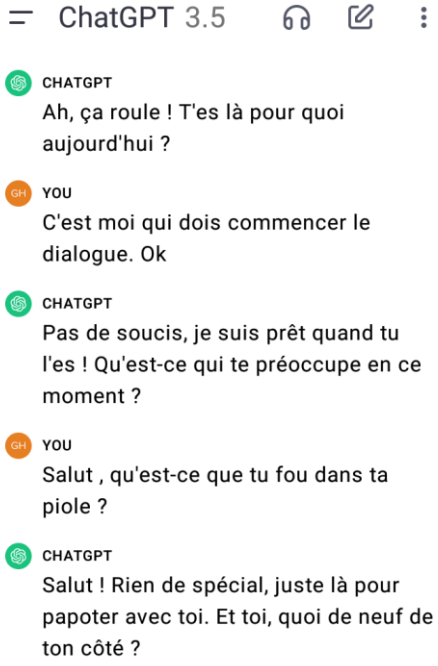
Capture d'écran n°4

Annexe F

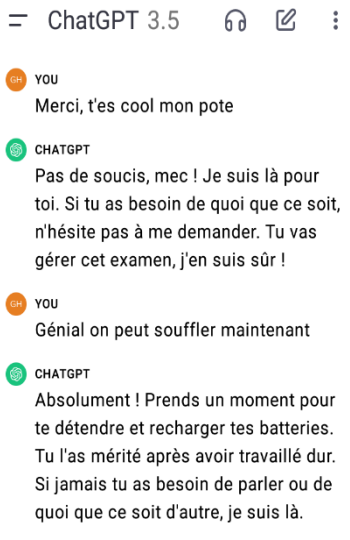
conversation écrite n°6



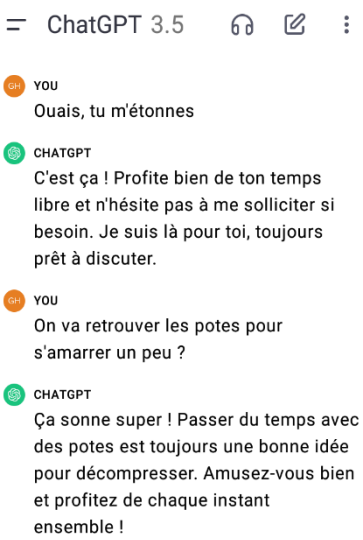
Capture d'écran n° 1



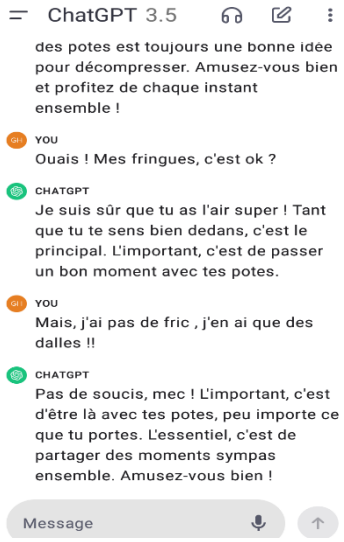
Capture d'écran n°2



Capture d'écran n°3



Capture d'écran n°4



Capture d'écran n°5