

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ MOHAMED BOUDIAF - M'SILA

FACULTÉ DES LETTRES ET DES LANGUES
DÉPARTEMENT DES LETTRES ET
LANGUE FRANÇAISE



DOMAINE : LETTRES ET LANGUES
ÉTRANGÈRES
FILIERE : LANGUE FRANÇAISE
SPÉCIALITÉ : DLE

N° :

Mémoire présenté pour l'obtention
Du diplôme de Master Académique
Par : Ilhem Aya Ould Chakmakdji

Intitulé :

Entre subjectivité pédagogique et objectivité algorithmique :
Vers une évaluation hybride des productions écrites à l'ère
de l'IA (ChatGPT).

- Cas des élèves de 1^{ère} AM, au collège
" Belouaer Belgassem, d'Al Anasser-BBA.

Soutenu devant le jury composé de :

Nom et Prénom	Grade	Qualité	Établissement
KEFSI Nadia	MAB	Président	Université Mohamed Boudiaf - M'sila
ROUABAH Fatima Zohra	MCA	Rapporteur	Université Mohamed Boudiaf - M'sila
BENSFA Nabil	MAB	Examineur	Université Mohamed Boudiaf - M'sila

Année universitaire : 2024/2025

Remerciements

*Je remercie **Dieu** le tout-puissant, de m'avoir inspiré le courage, la force et la volonté indispensables à la réalisation de ce modeste projet de recherche.*

*Je souhaite exprimer ma profonde gratitude à ma directrice de recherche, **Dre ROUABAH Fatima Zohra**, pour sa disponibilité, ses conseils éclairés et son accompagnement rigoureux, qui ont considérablement enrichi cette recherche.*

Merci...

*Je tiens à remercier **les membres du jury** pour avoir accepté d'examiner ce travail, pour le temps qu'ils y consacrent ainsi que pour l'intérêt qu'ils lui accordent.*

*J'adresse également mes sincères remerciements à **Monsieur Chaker**, le responsable des laboratoires de langue, pour ses aides précieuses, sa disponibilité et le temps généreusement consacré à l'accompagnement de ce travail.*

Dédicace

Je dédie ce mémoire, avec une profonde gratitude, à toutes les personnes qui ont soutenu et accompagné mon parcours, à ceux qui, quels que soient les termes embrassés, je n'arriverais jamais à leur exprimer mon amour sincère.

*À la femme qui a souffert sans me laisser souffrir, qui n'a jamais dit non à mes exigences et qui n'a épargné aucun effort pour me rendre heureuse, **ma chère mère Nadia**, pour son amour, ses prières et son soutien inébranlable.*

À l'homme qui représente pour moi le plus précieux des dons de Dieu, à celui à qui je dois ma vie, ma réussite et tout mon respect : mon père Kamel, pour ses encouragements constants et sa confiance indéfectible.

*À **ma sœur Donia** et **mon frère Noureddine**, pour leur affection, leur patience et leur présence constante.*

À ma grand-mère, mes oncles et mes tantes, pour leurs mots d'encouragement et leur générosité de cœur. Que Dieu leur accorde une vie longue et remplie de bonheur.

*À Mon oncle **Louail Ishek** qui mérite ma plus grande reconnaissance pour son aide généreuse, ses nombreuses contributions, sa disponibilité sans faille et sa présence authentique et continue.*

*À **mes amies sincères**, Yasmin, Aya, Maroua, qui m'ont apporté réconfort, motivation et sourires, même dans les moments difficiles.*

À chacun d'entre vous, merci du fond du cœur. Ce mémoire vous est dédié.

Table des matières

Introduction Générale	1
-----------------------------	---

Chapitre I

L'évolution de la production écrite : Approches et limites

Introduction partielle	5
I.1. Cadre théorique de la production écrite en FLE	7
I.1.1. Définition et spécificités de la production écrite en langue étrangère	7
I.1.2. Distinction entre production écrite en langue maternelle et en langue étrangère	7
I.1.3. Compétences linguistiques et cognitives mobilisées lors de la production écrite	8
I.1.4. Place de la production écrite dans le CECR.....	9
I.1.5. Processus cognitifs impliqués dans la production écrite	9
I.1.6. Modèles théoriques du processus rédactionnel.....	10
I.1.6.1. Modèle de Hayes et Flower (1980).....	10
I.1.6.2. Modèle de Bereiter et Scardamalia (1987).....	10
I.1.7. Étapes de la production écrite : planification, mise en texte, révision.....	10
I.8. Rôle de la mémoire dans le traitement de l'information.....	11
I.9. Difficultés spécifiques des apprenants de FLE en production écrite	12
I.9.1. Interférences avec la langue maternelle.....	12
I.9.2. Défis linguistiques (lexique, grammaire, syntaxe)	13
I.9.3. Défis culturels et pragmatiques	13
I.2. Approches didactiques pour l'enseignement de la production écrite en FLE	14
I.2.1. Evolution des approches méthodologiques.....	14
I.2.1.1. De la méthode traditionnelle à l'approche communicative	14
I.2.1.2. L'approche actionnelle	15
I.2.2. Influence des TIC sur les pratiques d'enseignement de l'écrit.....	15
I.2.3. Approche par compétences et production écrite.....	16
I.2.4. Typologie des activités de production écrite	17
I.2.4.1. Activités de production écrite selon les niveaux du CECR	17
I.2.4.2. Stratégies d'enseignement efficaces de la production écrite.....	18
I.2.4.3. Traitement de l'erreur en production écrite.....	19
I.2.5. Techniques d'évaluation et de remédiation	22
I.2.5.1. Évaluation de la production écrite	23
Conclusion partielle.....	34

Chapitre II

L'évaluation automatisée des productions en classe de FLE

Introduction partielle	34
II.1. Fondements technologiques de l'évaluation automatisée des productions écrites	35
II.1.1. Evolution des technologies d'analyse automatique du langage	35
II.1.2. Intelligence artificielle et apprentissage automatique	36
II.1.3. Principes de fonctionnement des modèles génératifs.....	38
II.1.4. Applications de l'évaluation automatisée dans le domaine du FLE.....	39
II.1.4.1. Évaluation des compétences linguistiques	39
II.1.4.2. Évaluation des aspects discursifs	41
II.1.4.3. Études de cas de systèmes d'évaluation automatisée en FLE	43
II.1.4.4. Applications pédagogiques de ChatGPT dans l'évaluation formative	46
II.1.4.5. Avantages et limites de l'évaluation automatisée	48
II.1.5. Comparaison entre évaluation humaine et évaluation automatisée des.....	56
II.1.5.1. Analyse comparative des processus d'évaluation.....	56
II.1.5.1.2. Fiabilité et validité des évaluations.....	58
II.1.5.2. Complémentarité des approches.....	60
II.1.5.3. Impact sur les pratiques pédagogiques	64
Conclusion partielle.....	68

Chapitre III

L'expérimentation

Introduction partielle	69
III.1. Démarche méthodologique	69
III.1.1. Objectifs de la recherche.....	69
III.1.2. La méthode utilisée	69
III.1.3. Le public visé	70
III.1.4. Présentation de l'établissement	70
III.1.5. Le cadre spatio-temporel	70
III.1.6. Le matériel utilisé.....	70
III.1.7. Le choix des critères de réussite	70
III.2. Déroulement de l'expérimentation	71
III.2.1. Corpus	71
III.2.1.1. Evaluation de ChatGPT des copies des élèves	71
III.2.1.2. Evaluation de l'enseignante des mêmes copies.....	94
III.2.2. Analyse Comparative Détaillée : Evaluation Humaine/Evaluation ChatGPT	106

III.2.3. Synthèse de l'Analyse Comparative : Évaluation Humaine/Evaluation ChatGPT	113
Conclusion Partielle.....	116

Conclusion Générale

Conclusion Générale	118
Références bibliographiques.....	123
Annexe N°1 : Photos des élèves lors de l'activité d'écriture.....	130
Annexe N°2 : Corpus	131
Résumé.....	133

Introduction Générale

L'évaluation des productions écrites en langue étrangère est une activité incontournable dans l'enseignement-apprentissage du français langue étrangère (*FLE*).

Elle ne se limite pas à mesurer les compétences des apprenants ; elle vise aussi à leur offrir un retour constructif pour les aider à progresser dans leurs expressions écrites. Jusqu'à présent, cette évaluation était l'apanage de l'expertise des enseignants, basée sur une analyse des textes selon différents critères, tels que linguistique, discursif et pragmatique. Toutefois, récemment, avec l'avènement des nouvelles technologies et le développement de l'intelligence artificielle, divers outils automatisés ont vu le jour. Néanmoins, ceci soulève des questions quant à l'efficacité, l'éventualité d'un remplacement du type d'évaluation humaine.

L'évaluation menée par un enseignant repose sur une analyse qualitative prenant en compte plusieurs paramètres : la cohérence du texte, la richesse du vocabulaire, la maîtrise des structures syntaxiques, ainsi que l'originalité et la pertinence des idées développées. Toutefois, bien qu'elle soit assez précise, elle est sujette à de nombreux aléas. En effet, Elle présente l'inconvénient d'être corrompue par le critère de subjectivité du correcteur, d'où des divergences relevées d'un évaluateur à un autre. En tant que notation humaine, la subjectivité de l'évaluation est selon Cizek, G. J, « *le jugement humain demeure une composante essentielle de l'évaluation, mais il peut parfois entraîner des divergences dans l'appréciation des productions écrites* » (1996, p.67).

Il convient de noter les plus grandes difficultés humaines de l'évaluation des écrits, qui sont connues sous les effets parasites. Tout d'abord, la perception subjective dépend de l'humeur, de la manière de dormir et de nombreux autres aspects. Chaque enseignant interprétera et évaluera individuellement toutes les parties d'un énoncé, donnant un sens particulier. Les distorsions peuvent être typiques et communes à tous les évaluateurs ou être basées uniquement sur des préférences. Le type d'expression grammatical, l'idiosyncrasique et même la fatigue après un long épuisement peuvent être affectés. De même, il y a un immense volume de déclaration à évaluer. Cela est particulièrement vrai dans la vie lors d'examens ou d'évaluations massives, où il est difficile de s'assurer de l'absence d'erreurs humaines, de différences flagrantes dans les scores, et même d'une fatigue générale qui affecte l'ensemble de la qualité de l'évaluation. De nombreux enseignants, restreints dans des délais et des moyens, ont du mal à offrir à chaque apprenant une rétroaction

individualisée, bien que le souhait de fournir des commentaires aussi détaillés et constructifs soit présent. Ainsi, il est évident que cette question est également discutée.

Cela peut être retracé dans le livre de Daniel Kahneman, "*Thinking, Fast and Slow*". Il avance que les biais cognitifs affectent notre capacité à juger de manière objective. Nous faisons confiance à des heuristiques rapides qui sont souvent fausses, mais qui nous semblent intuitivement justes, ce qui nous mène à des évaluations erronées.

Face à ces difficultés, l'intelligence artificielle (IA) apparaît comme une solution potentielle pour améliorer l'évaluation des productions écrites en classe de FLE.

En effet, L'intelligence artificielle a considérablement élargi la portée de l'évaluation, en introduisant des systèmes intelligents qui évaluent le texte selon des paramètres prédéfinis. En outre, grâce à des algorithmes spécialisés de NLP, cette technologie est capable de trouver la faute, d'évaluer la structure et de corriger l'auteur. Ce point est important sur les plans professionnel et universitaire. Certains chercheurs le soutiennent car un tel support est précieux pour les enseignants pendant le processus de notation et permet de gagner du temps.

Ainsi, Attali et Burstein soulignent que « *l'évaluation automatisée garantit une cohérence dans la notation et offre aux apprenants un retour immédiat, venant en complément du jugement humain* » (2006, p. 28).

Toutefois, malgré tous les avantages, ces technologies ont également des défauts significatifs. Ils ne sont pas en mesure d'évaluer de manière adéquate la créativité, l'intentions de l'auteur et les subtilités stylistiques qui restent des aspects clés dans l'évaluation de l'œuvre littéraire. En outre, certaines des études les plus récentes concluent que ces systèmes risquent présenter un certain biais en fonction des bases de données sur lesquelles ils ont été formés. (Perelman, 2014).

Ce mémoire vise, donc, à étudier plus largement les particularités et les défis de l'évaluation des productions en comparant l'évaluation humaine et l'évaluation automatisée.

Ce sujet a été choisi en raison de la pression liée à la correction des copies de production écrite dans un délai restreint. De ce fait, quelques professeurs font appel à l'IA pour optimiser et accélérer leur travail.

La question centrale qui guidera cette recherche est la suivante : Dans quelle mesure l'intelligence artificielle peut-elle compléter et améliorer l'évaluation humaine des productions écrites en FLE sans altérer la qualité du jugement porté sur ces textes ?

Tout au long de ce mémoire, nous essayerons d'apporter des réponses aux sous-questions suivantes :

Le recours aux outils d'IA, tels qu'Antidote, Scribens, ou encore ChatGPT, dans la correction des productions écrites permet-il de réduire les erreurs grammaticales et stylistiques, tout en stimulant l'autonomie des apprenants ? Quelle place ces outils doivent-ils occuper dans une évaluation équilibrée, en complément ou en remplacement de l'évaluation humaine ? Existe-t-il une complémentarité pédagogique entre les retours automatisés et ceux fournis par l'enseignant ?

Afin d'explorer cette problématique, nous formulerons les hypothèses suivantes :

L'évaluation humaine serait encore requise pour évaluer la cohérence, l'originalité et l'intention communicative des œuvres. L'intelligence artificielle pourrait offrir un support précieux en garantissant plus d'objectivité et permettrait d'accélérer l'évaluation, mais elle ne pourrait pas remplacer totalement le rôle de l'enseignant.

À l'issue de ce qui précède, le principal objectif de ce mémoire est d'interroger les implications pédagogiques de l'utilisation des outils d'IA dans l'évaluation des productions écrites en français, en explorant leurs effets sur l'amélioration des compétences rédactionnelles des apprenants et sur la perception des évaluations par l'IA comparées aux évaluations traditionnelles.

Ce travail vise à :

- Etudier les fonctionnalités et les performances des outils d'IA (*surtout ChatGPT*) dans l'évaluation des écrits.
- Comparer les évaluations réalisées par des correcteurs humains et celles assistées par IA (*ChatGPT*).
- Analyser l'impact de l'utilisation de ces outils sur les apprentissages en FLE.
- Proposer des recommandations pour une intégration optimale des outils d'IA dans les pratiques pédagogiques.

Dans cette perspective la première partie de ce mémoire sera consacrée à la définition des principes et des critères pour évaluer les productions écrites en FLE. Dans la seconde partie, les forces et les faiblesses des systèmes d'évaluation automatisée seront explorées, à l'aide de nombreux exemples d'outils AI dans l'apprentissage. Enfin, la dernière partie proposera une expérimentation comparative entre l'évaluation humaine et l'évaluation automatisée.

Afin de mieux cerner les apports et les limites de ces deux approches dans l'enseignement-apprentissage du FLE, nous adopterons une approche expérimentale et qualitative, qui privilégie l'expérience. Pour ce faire, notre travail se divisera en trois chapitres :

- Dans Le premier chapitre intitulé "*L'évaluation des productions écrites en classe de FLE*", nous proposerons la définition de la Production écrite, ses types, son importance en classe de FLE, les difficultés rencontrées par les enseignants dans l'évaluation : subjectivité, charge de travail, incohérences et enfin les Critères d'évaluation des productions écrites (*précision grammaticale, richesse lexicale, cohérence textuelle, etc.*). Ce chapitre sera clôturé par une brève introduction de l'évaluation automatisée pour introduire le chapitre suivant.
- Dans le chapitre deux, nous aborderons les notions d'outil d'IA, son incorporation dans l'éducation, l'évaluation connexe, l'histoire de l'IA et son développement, et enfin l'avantage et la limite de l'évaluation IA.
- Dans le troisième chapitre, destiné à l'expérimentation, nous présenterons, dans un premier temps, le cadre général et le déroulement de l'expérimentation : démarche méthodologique, description du lieu de l'expérimentation, description de l'échantillon, description de l'outil d'analyse IA. En deuxième lieu, nous présenterons les résultats de notre recherche et l'analyse des données collectées, en relevant la comparaison entre les évaluations faites l'enseignant et celles de l'IA. Enfin, nous clôturerons notre étude par une conclusion générale où nous ressortirons le bilan de la recherche et verrons si nos hypothèses de départ ont été confirmées ou infirmées.

Chapitre N°I

L'évaluation de la production écrite :

Approches et limites

Introduction partielle

Dans le contexte actuel de mondialisation et d'échanges internationaux croissants, l'apprentissage des langues étrangères occupe une place prépondérante dans les systèmes éducatifs du monde entier. Le français, en tant que langue internationale parlée sur les cinq continents, attire de nombreux apprenants désireux de maîtriser cette langue pour des raisons professionnelles, académiques, culturelles ou personnelles. L'enseignement du français langue étrangère (*FLE*) s'est ainsi, considérablement développé ces dernières décennies, s'enrichissant de nouvelles approches méthodologiques et didactiques visant à optimiser l'acquisition des compétences langagières.

Parmi les quatre compétences fondamentales que sont la compréhension orale, la compréhension écrite, la production orale et la production écrite, cette dernière occupe une position particulière dans l'apprentissage d'une langue étrangère. En effet, la production écrite représente souvent un défi majeur pour les apprenants, car elle mobilise simultanément de nombreuses connaissances et compétences : maîtrise du code linguistique (lexique, grammaire, orthographe, syntaxe), organisation textuelle, adaptation au contexte de communication, prise en compte du destinataire, entre autres. Comme le souligne l'ouvrage de référence *"La production écrite - Didactique des langues étrangères"* (CLE International), *"rédiger est un processus complexe et amener les apprenants à acquérir une compétence en expression écrite n'est pas une tâche aisée"*.

Cette complexité intrinsèque à l'acte d'écriture se trouve amplifiée dans le contexte de l'apprentissage d'une langue étrangère, où l'apprenant doit non seulement composer avec les difficultés inhérentes à la production écrite, mais également avec les spécificités linguistiques et culturelles de la langue cible. Les interférences avec la langue maternelle, les différences dans les conventions d'écriture, les particularités stylistiques et rhétoriques propres au français constituent autant d'obstacles supplémentaires que l'apprenant de FLE doit surmonter pour développer une compétence scripturale satisfaisante.

Ce chapitre se propose d'adopter une perspective à la fois théorique et pratique. Il s'agira d'abord d'examiner les fondements théoriques de la production écrite en langue étrangère, en s'appuyant sur les recherches en psycholinguistique, en didactique des langues et en sciences cognitives. Nous nous intéresserons ensuite aux différentes approches didactiques qui ont marqué l'enseignement de la production écrite en FLE, en mettant en lumière leurs principes, leurs apports et leurs limites. Enfin, nous nous pencherons sur les pratiques pédagogiques concrètes et les modalités d'évaluation de la production écrite, en proposant des pistes d'action pour les enseignants de FLE.

La production écrite en classe de FLE soulève des enjeux multiples qui se situent à l'intersection de plusieurs champs disciplinaires : la linguistique, la psychologie cognitive, la didactique des langues, les sciences de l'éducation.

En effet, nous nous appuierons, dans cette partie, sur une analyse critique de la littérature scientifique et didactique, ainsi que sur des exemples concrets de pratiques pédagogiques. Notre objectif est de proposer une réflexion approfondie sur les enjeux de la production écrite en classe de FLE, tout en offrant des pistes d'action concrètes pour les enseignants désireux d'optimiser leur enseignement dans ce domaine.

Cette partie nous invite à explorer les fondements conceptuels de la production écrite en langue étrangère, contribuant ainsi à une meilleure compréhension des processus d'apprentissage. D'autre part, elle nous conduit à examiner les implications pédagogiques de ces connaissances théoriques, dans une perspective d'amélioration des pratiques d'enseignement. En ce sens, ce chapitre s'inscrit dans une démarche de recherche-action, visant à établir un dialogue fécond entre théorie et pratique dans le domaine de la didactique du FLE.

I.1. Cadre théorique de la production écrite en FLE

I.1.1. Définition et spécificités de la production écrite en langue étrangère

La production écrite constitue l'une des quatre compétences fondamentales dans l'apprentissage d'une langue étrangère, aux côtés de la compréhension écrite, de la compréhension orale et de la production orale. Elle se définit comme la capacité à produire des textes écrits de différentes natures, adaptés à des situations de communication variées. Cependant, la production écrite en langue étrangère présente des spécificités qui la distinguent nettement de celle réalisée en langue maternelle.

I.1.2. Distinction entre production écrite en langue maternelle et en langue étrangère

En langue maternelle, l'apprenant dispose déjà d'un système linguistique intériorisé et d'une compétence communicative développée naturellement depuis l'enfance. L'apprentissage de l'écrit s'appuie alors sur ces acquis oraux préexistants. Comme le souligne Vygotski (1997), l'enfant qui apprend à écrire dans sa langue maternelle transpose un système symbolique qu'il maîtrise déjà à l'oral vers un nouveau système de représentation.

En revanche, en langue étrangère, l'apprenant se trouve confronté à un double défi : il doit simultanément acquérir le système linguistique (*phonologie, lexicale, morphosyntaxe*) et développer des compétences scripturales dans cette nouvelle langue. Cette situation crée ce que Cummins (1979) appelle une "*interdépendance linguistique*" : les compétences développées en langue maternelle peuvent faciliter ou, au contraire, interférer avec l'apprentissage de l'écrit en langue étrangère.

Par ailleurs, comme le note Cornaire (1999), « *l'écriture en langue seconde est à la fois un problème linguistique, cognitif et socioculturel* » (P.54). Cette dimension socioculturelle est particulièrement importante, car les conventions d'écriture, les genres textuels et les attentes rhétoriques varient considérablement d'une culture à l'autre. Un apprenant de FLE doit donc non seulement maîtriser le code linguistique français, mais également s'approprier les normes culturelles qui régissent la production écrite dans l'espace francophone.

I.1.3. Compétences linguistiques et cognitives mobilisées lors de la production écrite

La production écrite en FLE mobilise un ensemble complexe de compétences linguistiques et cognitives. Sur le plan linguistique, l'apprenant doit maîtriser :

- Le système graphique du français, qui présente des particularités comme les accents, les lettres muettes ou les correspondances graphèmes-phonèmes complexes
- Le lexique, avec ses nuances sémantiques, ses collocations et ses expressions idiomatiques
- La morphosyntaxe, incluant la conjugaison verbale, les accords, la structure des phrases
- La cohésion textuelle, assurée par les connecteurs logiques, les anaphores et autres procédés de reprise
- Les registres de langue, adaptés aux différentes situations de communication

Sur le plan cognitif, la production écrite sollicite des processus de haut niveau tels que :

- La planification du contenu et de l'organisation du texte
- La génération d'idées et leur mise en relation
- La résolution de problèmes linguistiques et textuels
- La révision et l'autoévaluation
- La gestion de la charge cognitive, particulièrement lourde en langue étrangère

Comme l'explique Barbier (2003), « *la surcharge cognitive est l'un des obstacles majeurs à la production écrite en langue étrangère* ». (P. 32). En effet, l'attention de l'apprenant étant largement mobilisée par les aspects formels de la langue (*orthographe, grammaire*), il dispose de moins de ressources cognitives pour se concentrer sur le contenu et l'organisation de son texte.

I.1.4. Place de la production écrite dans le CECR

Le Cadre Européen Commun de Référence pour les langues (*CECR*) accorde une place importante à la production écrite, qu'il intègre dans une perspective actionnelle considérant l'apprenant comme un "acteur social" ayant à accomplir des tâches dans un contexte donné. Le CECR définit des niveaux de compétence (*de A1 à C2*) pour la production écrite, avec des descripteurs précis pour chaque niveau.

Au niveau A1 (*débutant*), l'apprenant peut "écrire des expressions et phrases simples isolées" comme des informations personnelles ou des messages courts. Au niveau C2 (*maîtrise*), il est capable de « produire un texte clair, fluide et bien structuré, démontrant un usage contrôlé de moyens linguistiques de structuration et d'articulation ».

Le CECR distingue également différentes activités de production écrite :

- L'écriture créative (*récits, poèmes, etc.*)
- Les essais et rapports
- La correspondance (*lettres, courriels, etc.*)
- Les notes, messages et formulaires

Cette approche par compétences et par tâches a considérablement influencé la didactique de la production écrite en FLE, en orientant l'enseignement vers des objectifs communicatifs concrets et en favorisant une progression cohérente dans l'acquisition des compétences scripturales.

I.1.5. Processus cognitifs impliqués dans la production écrite

La compréhension des processus cognitifs qui sous-tendent l'acte d'écriture est essentielle pour concevoir un enseignement efficace de la production écrite en FLE. Les recherches en psycholinguistique et en sciences cognitives ont permis d'élaborer plusieurs modèles théoriques qui éclairent ces processus.

I.1.6. Modèles théoriques du processus rédactionnel

I.1.6.1. Modèle de Hayes et Flower (1980)

Bien que développé initialement pour la langue maternelle, reste une référence incontournable pour comprendre les processus d'écriture en langue étrangère. Ce modèle identifie trois processus majeurs :

1. La planification : génération d'idées, organisation du contenu, définition des objectifs
2. La mise en texte : transformation des idées en langage écrit
3. La révision : évaluation et modification du texte produit

Ces processus ne sont pas linéaires mais récursifs, c'est-à-dire qu'ils peuvent intervenir à tout moment de la production et s'entremêler. En langue étrangère, comme le souligne Barbier (2003), « *la récursivité des processus est encore plus marquée, l'apprenant revenant constamment sur son texte pour vérifier la correction linguistique* ». (P. 22)

I.1.6.2. Modèle de Bereiter et Scardamalia (1987)

Il distingue deux approches de la rédaction :

- La "*knowledge telling*" (*dire les connaissances*) : approche linéaire où le scripteur transcrit ses idées au fur et à mesure qu'elles lui viennent à l'esprit
- La "*knowledge transforming*" (*transformer les connaissances*) : approche plus élaborée où l'écriture devient un outil de réflexion et de construction de la pensée.

En FLE, les apprenants adoptent souvent la première approche, se concentrant sur la production d'un texte grammaticalement correct plutôt que sur la transformation de leurs connaissances. L'enjeu pédagogique consiste à les faire évoluer vers la seconde approche, plus sophistiquée.

I.1.7. Étapes de la production écrite : planification, mise en texte, révision

La planification constitue une étape cruciale dans le processus d'écriture, particulièrement en langue étrangère. Elle permet à l'apprenant de réduire la charge cognitive en préparant le contenu avant de se confronter aux difficultés linguistiques. Comme le note Moirand (1979), « *la planification en langue étrangère nécessite souvent un travail préalable plus important qu'en langue maternelle* ». (P.44). Cette phase peut inclure des activités comme le remue-méninge, l'élaboration de cartes mentales ou la recherche documentaire.

La mise en texte représente le cœur du processus d'écriture, où l'apprenant transforme ses idées en un texte cohérent. En FLE, cette phase est particulièrement exigeante car elle mobilise simultanément des connaissances linguistiques encore fragiles et des compétences discursives en développement. Selon Cornaire et Raymond (1999), « *la mise en texte en langue étrangère est souvent ralentie par la recherche lexicale et la vérification grammaticale constantes* » (P. 64)

La révision, enfin, joue un rôle essentiel dans l'amélioration de la qualité du texte. Elle comprend la relecture, la détection des erreurs et la réécriture. En FLE, la révision se concentre souvent sur les aspects formels (orthographe, grammaire) au détriment des aspects textuels (cohérence, cohésion). Comme le souligne Manchón (2009), « *il est nécessaire d'encourager les apprenants à réviser non seulement la forme mais aussi le contenu de leurs productions* ». (P. 77)

I.8. Rôle de la mémoire dans le traitement de l'information

La mémoire joue un rôle fondamental dans la production écrite en langue étrangère. Le modèle de la mémoire de travail de Baddeley (2000) permet de comprendre comment les différentes composantes de la mémoire interagissent pendant l'acte d'écriture :

- La boucle phonologique traite l'information verbale et joue un rôle dans la formulation des phrases
- Le calepin visuo-spatial intervient dans la représentation graphique et la mise en page
- L'administrateur central coordonne ces systèmes et gère l'attention
- Le buffer épisodique intègre les informations provenant de différentes sources

En langue étrangère, la mémoire de travail est particulièrement sollicitée car l'apprenant doit gérer simultanément de nombreuses contraintes : recherche lexicale, application des règles grammaticales, organisation textuelle, etc. Cette surcharge peut entraîner ce que Kellogg (1996) appelle une "*compétition pour les ressources attentionnelles*", limitant la qualité de la production.

La mémoire à long terme intervient également dans le processus d'écriture, en fournissant les connaissances linguistiques, textuelles et thématiques nécessaires. En FLE, ces connaissances étant moins automatisées qu'en langue maternelle, leur récupération demande un effort cognitif plus important. Comme l'explique Barbier (2003), « *l'automatisation progressive des processus de bas niveau (orthographe, syntaxe) permet de libérer des ressources cognitives pour les processus de haut niveau (planification, organisation)* ». (P. 71)

I.9. Difficultés spécifiques des apprenants de FLE en production écrite

Les apprenants de français langue étrangère rencontrent des difficultés particulières en production écrite, qui varient selon leur langue maternelle, leur niveau de compétence et leur familiarité avec l'écrit en général. Identifier ces difficultés permet d'adapter l'enseignement et de proposer des activités ciblées.

I.9.1. Interférences avec la langue maternelle

L'influence de la langue maternelle (L1) sur la production écrite en français (L2) constitue l'une des principales sources de difficultés. Ces interférences peuvent se manifester à différents niveaux :

- Phonologique : la perception des sons du français influence l'orthographe (*par exemple, la difficulté à distinguer /p/ et /u/ pour les arabophones*)
- Lexical : les faux amis, les calques lexicaux ou les emprunts inappropriés
- Syntaxique : le transfert de structures grammaticales de la L1 vers le français
- Discursif : l'organisation textuelle selon les conventions de la L1

Comme le souligne Odlin (1989), « *le transfert linguistique peut être positif lorsque les structures sont similaires dans les deux langues, mais négatif lorsqu'elles diffèrent* ». (P.55). Ainsi, les apprenants dont la langue maternelle est proche du français (italien, espagnol) bénéficient d'un transfert positif pour certains aspects, mais peuvent être victimes d'interférences plus subtiles et persistantes.

L'étude de Lukogho Vagheni (2019) sur la production écrite par mail chez des apprenants congolais de niveau A2 montre que « *les interférences avec les langues locales affectent principalement la syntaxe et l'organisation discursive des productions* ». (P. 19). Ces interférences peuvent persister même à des niveaux avancés, nécessitant une attention particulière de la part des enseignants.

I.9.2. Défis linguistiques (lexique, grammaire, syntaxe)

Les spécificités du système linguistique français posent de nombreux défis aux apprenants de FLE :

- Le système verbal complexe, avec ses nombreux temps et modes
- Les accords en genre et en nombre
- Les articles et les prépositions, souvent absents ou différents dans d'autres langues
- La négation à deux éléments
- L'ordre des mots relativement fixe mais avec des variations significatives

Le lexique français, avec ses nuances, ses expressions idiomatiques et ses registres variés, représente également un obstacle majeur. Comme le note Tréville (2000), « *la compétence lexicale en langue étrangère ne se limite pas à la connaissance de mots isolés, mais implique la maîtrise des collocations, des expressions figées et des connotations culturelles* ». (P. 60)

La syntaxe française, caractérisée par des phrases souvent longues et complexes, avec des propositions subordonnées et des constructions particulières (*mise en relief, voix passive, etc.*), constitue un autre défi de taille. Selon l'étude de Guernine et Kacem (2024) sur l'exploitation de l'erreur dans la production écrite, « *les erreurs syntaxiques sont parmi les plus fréquentes et les plus persistantes chez les apprenants de FLE, même à des niveaux avancés* ». (P. 84)

I.9.3. Défis culturels et pragmatiques

Au-delà des aspects purement linguistiques, la production écrite en FLE implique une dimension culturelle et pragmatique souvent sous-estimée. Les conventions d'écriture, les genres textuels et les attentes rhétoriques varient considérablement d'une culture à l'autre. Par exemple, la dissertation à la française, avec sa structure tripartite (*thèse, antithèse, synthèse*) et son style argumentatif spécifique, peut dérouter des apprenants habitués à d'autres traditions rhétoriques. De même, les conventions épistolaires françaises, avec leurs formules d'ouverture et de clôture codifiées, nécessitent un apprentissage explicite.

La dimension pragmatique concerne l'adéquation du texte à la situation de communication : choix du registre, degré de formalité, prise en compte du destinataire, etc. Comme le souligne Moirand (1979), « *écrire en langue étrangère, c'est aussi s'approprier*

les règles socioculturelles qui régissent la communication écrite dans la communauté cible ». (P. 20)

Les travaux de Beacco (2007) sur la dimension culturelle de l'enseignement des langues montrent que la maîtrise des genres discursifs propres à une culture est indissociable de la compétence linguistique. Cette dimension culturelle de la production écrite doit donc être explicitement enseignée, au même titre que les aspects linguistiques.

I.2. Approches didactiques pour l'enseignement de la production écrite en FLE

I.2.1. Evolution des approches méthodologiques

L'enseignement de la production écrite en français langue étrangère a connu une évolution significative au fil des décennies, reflétant les changements de paradigmes dans la didactique des langues. Cette évolution témoigne d'une compréhension de plus en plus fine des processus d'apprentissage et d'une adaptation aux besoins communicatifs des apprenants.

I.2.1.1. De la méthode traditionnelle à l'approche communicative

L'approche traditionnelle, dominante jusqu'aux années 1950, accordait une place prépondérante à l'écrit, mais dans une perspective essentiellement normative. La production écrite était principalement envisagée comme un exercice de traduction ou de rédaction sur des thèmes littéraires, visant la maîtrise de la grammaire et du style. Comme le souligne Puren (1988), l'écrit était considéré comme le vecteur privilégié de la culture savante, et la production écrite comme l'aboutissement de l'apprentissage linguistique.

La méthode audio-orale et la méthode SGAV (*structuro-globale audio-visuelle*), qui ont dominé les années 1950-1970, ont paradoxalement marginalisé la production écrite au profit de l'oral. L'écrit était essentiellement conçu comme une transcription de l'oral, et les activités de production écrite se limitaient souvent à des exercices structuraux écrits. Selon Germain (1993), ces méthodes considéraient l'écrit comme un système secondaire, dont l'apprentissage devait être différé après une solide acquisition des structures orales.

L'approche communicative, qui s'est développée à partir des années 1970, a réhabilité la production écrite en l'intégrant dans une perspective fonctionnelle. L'écrit est alors envisagé comme un acte de communication authentique, répondant à des besoins spécifiques dans des situations variées. Moirand (1979) définit ainsi la compétence de communication

écrite comme « *la capacité de produire des textes adaptés à des situations précises, en tenant compte des paramètres de la communication : destinataire, objectif, contexte* ». (P. 29)

I.2.1.2. L'approche actionnelle

L'approche actionnelle, préconisée par le Cadre Européen Commun de Référence pour les langues (CECR) depuis 2001, représente l'évolution la plus récente. Elle considère l'apprenant comme un "acteur social" qui accomplit des tâches, non seulement langagières, dans des circonstances et un environnement donné. Comme l'explique Rosen (2009), « *la production écrite s'inscrit désormais dans une perspective actionnelle où l'apprenant mobilise l'ensemble de ses compétences et ressources pour réaliser des tâches significatives* ». (P. 38)

Cette évolution méthodologique reflète un changement profond dans la conception de l'écrit en langue étrangère : d'un écrit normé et décontextualisé, on est passé à un écrit fonctionnel, puis à un écrit social et actionnel, intégré dans des pratiques authentiques.

I.2.2. Influence des TIC sur les pratiques d'enseignement de l'écrit

Les technologies de l'information et de la communication (TIC) ont profondément transformé les pratiques d'enseignement de la production écrite en FLE. Comme le note Mangenot (2000), « *les TIC ont non seulement modifié les supports et les outils d'écriture, mais également les genres textuels, les processus rédactionnels et les situations de communication* ».

L'étude de Lukogho Vagheni (2019) sur l'utilisation du mail en didactique du FLE illustre parfaitement cette évolution. L'auteur démontre que le mail, en tant qu'outil didactique, a facilité les interactions et suscité l'engagement des apprenants qui, par la suite et hors leçon, ont pu rédiger des mails dans des contextes divers. Cette recherche souligne comment les outils numériques peuvent créer des situations de communication authentiques et motivantes.

Les TICE ont également favorisé l'émergence de nouvelles pratiques pédagogiques :

- L'écriture collaborative via des plateformes comme Google Docs ou les wikis
- La publication en ligne des productions des apprenants (blogs, sites web)
- L'utilisation de corpus numériques comme ressources pour l'écriture
- Le recours à des outils d'aide à la rédaction (correcteurs, dictionnaires en ligne)
- Les échanges interculturels via des projets de correspondance électronique

Comme le souligne Dejean-Thircuir et Mangenot (2006), « *les TIC permettent de décloisonner l'espace-classe et d'inscrire la production écrite dans des interactions sociales authentiques* ». (P. 49). Cette dimension sociale de l'écriture numérique rejoint les principes de l'approche actionnelle en plaçant l'apprenant dans des situations de communication réelles.

Toutefois, l'intégration des TICE dans l'enseignement de la production écrite soulève également des questions didactiques. Selon Barbot et Pugibet (2002), « *l'utilisation des outils numériques nécessite une réflexion sur les nouvelles compétences à développer chez les apprenants, notamment la littératie numérique et la capacité à s'adapter à différents environnements d'écriture* ». (P. 81)

I.2.3. Approche par compétences et production écrite

L'approche par compétences, qui s'est développée parallèlement à l'approche actionnelle, a considérablement influencé l'enseignement de la production écrite en FLE. Cette approche vise à développer chez l'apprenant des compétences transversales et transférables, mobilisables dans diverses situations.

Le CECR définit la compétence à communiquer langagièrement comme comprenant plusieurs composantes :

- La composante linguistique (*lexique, grammaire, phonologie, orthographe*)
- La composante sociolinguistique (*marqueurs de relations sociales, règles de politesse, etc.*)
- La composante pragmatique (*organisation discursive, cohérence, cohésion*)

Dans cette perspective, la production écrite n'est plus envisagée comme une simple application de règles grammaticales, mais comme la mobilisation intégrée de ces différentes composantes pour réaliser des tâches communicatives. Comme l'explique Beacco (2007), « *l'approche par compétences implique de décomposer la compétence de production écrite en savoir-faire identifiables et évaluables, tout en maintenant une vision holistique de l'acte d'écriture* ». (P. 29)

Cette approche a conduit à une redéfinition des objectifs d'apprentissage et des critères d'évaluation de la production écrite. Tagliante (2005) propose ainsi une évaluation Multidimensionnelle qui prend en compte :

- L'adéquation à la situation de communication (*respect de la consigne, prise en compte du destinataire*)
- La cohérence textuelle (*organisation logique, progression des idées*)
- La correction linguistique (*lexique, grammaire, orthographe*)
- La richesse et la précision des moyens linguistiques utilisés

L'approche par compétences a également favorisé l'intégration de la production écrite dans des séquences didactiques articulant les différentes activités langagières. Comme le souligne Cuq et Gruca (2005), « *la production écrite ne doit pas être isolée des autres compétences, mais s'inscrire dans un continuum d'activités qui permettent à l'apprenant de construire progressivement sa compétence scripturale* ». (P. 99)

I.2.4. Typologie des activités de production écrite

La diversité des activités de production écrite proposées en classe de FLE reflète la variété des objectifs pédagogiques, des niveaux de compétence et des contextes d'apprentissage. Une typologie de ces activités permet d'identifier les principes qui sous-tendent leur conception et leur mise en œuvre.

I.2.4.1. Activités de production écrite selon les niveaux du CECR

Le CECR propose une progression dans les activités de production écrite, adaptée aux différents niveaux de compétence :

Au niveau A1 (débutant), les activités se concentrent sur des productions simples et fonctionnelles :

- Remplir un formulaire avec des informations personnelles
- Écrire une carte postale ou un message court
- Rédiger des phrases simples sur soi-même ou sur son environnement immédiat

Comme le note Tagliante (2005), « *à ce niveau, les activités doivent être fortement guidées et contextualisées, avec un apport linguistique suffisant pour permettre à l'apprenant de réaliser la tâche* ». (P.105)

Au niveau A2 (*élémentaire*), les activités s'élargissent à des productions plus développées :

- Écrire des notes et messages simples relatifs à des besoins immédiats
- Rédiger une lettre personnelle très simple
- Décrire un événement, une expérience personnelle

L'étude de Lukogho Vagheni (2019) montre que les apprenants de niveau A2 peuvent produire des mails efficaces lorsque la consigne est claire et que le contexte de communication est bien défini.

Aux niveaux B1 et B2 (*intermédiaire*), les activités visent des productions plus élaborées et nuancées :

- Rédiger un texte articulé sur des sujets familiers ou d'intérêt personnel
- Écrire des lettres personnelles décrivant expériences et impressions
- Rédiger un essai ou un rapport qui développe une argumentation

Selon Cuq et Gruca (2005), ces niveaux correspondent à un saut qualitatif important dans la production écrite, avec l'introduction de la dimension argumentative et la prise en compte de la complexité textuelle.

Aux niveaux C1 et C2 (*avancé*), les activités ciblent des productions complexes et sophistiquées :

- Rédiger des textes bien structurés sur des sujets complexes
- Produire des écrits dans des styles appropriés à différents contextes
- Élaborer des résumés et des critiques d'ouvrages

Comme le souligne Beacco (2007), à ces niveaux, l'accent est mis sur la maîtrise des nuances stylistiques et rhétoriques, ainsi que sur la capacité à adapter sa production à des genres discursifs variés.

I.2.4.2. Stratégies d'enseignement efficaces de la production écrite

L'efficacité de l'enseignement de la production écrite en FLE repose sur la mise en œuvre de stratégies pédagogiques adaptées aux besoins des apprenants et aux objectifs d'apprentissage. Ces stratégies doivent prendre en compte les dimensions cognitive, affective et sociale de l'acte d'écriture.

I.2.4.2.1. Création d'un climat de confiance et motivation des apprenants

La dimension affective joue un rôle crucial dans l'apprentissage de la production écrite en langue étrangère. Comme le souligne l'ouvrage "*La production écrite*" (CLE International), « *créer un climat de confiance est essentiel pour que les élèves puissent s'accommoder des contraintes de l'écrit* ». Cette confiance se construit progressivement à travers plusieurs approches :

- La valorisation des réussites et des progrès, même modestes
- La dédramatisation de l'erreur, présentée comme une étape normale de l'apprentissage.
- L'instauration d'un rapport positif à l'écrit, notamment par des activités ludiques et créatives.
- La prise en compte des représentations et des expériences antérieures des apprenants.

La motivation constitue un autre facteur déterminant. Selon Viau (1994), la motivation en contexte scolaire dépend de trois perceptions : la valeur de l'activité, la compétence à l'accomplir et la contrôlabilité de son déroulement. Pour favoriser la motivation en production écrite, l'enseignant peut :

- Proposer des tâches d'écriture authentiques, ayant un sens pour l'apprenant
- Expliciter les objectifs et l'utilité des activités proposées
- Offrir des choix dans les sujets, les formats ou les destinataires
- Concevoir des défis adaptés, ni trop faciles ni trop difficiles
- Donner un feedback constructif et encourageant

Comme le note Cornaire et Raymond (1999), la motivation à écrire en langue étrangère est fortement liée au sentiment d'efficacité personnelle. Il est donc essentiel de permettre aux apprenants de vivre des expériences de réussite en production écrite.

I.2.4.3. Traitement de l'erreur en production écrite

La question du traitement de l'erreur occupe une place centrale dans l'enseignement de la production écrite en FLE. L'évolution des approches didactiques a conduit à reconsidérer le statut de l'erreur, désormais envisagée comme un révélateur des processus d'apprentissage et un levier pour la progression.

I.2.4.3.1. Typologie des erreurs en production écrite

L'identification et la catégorisation des erreurs constituent une étape préalable à leur traitement pédagogique. Plusieurs typologies ont été proposées, adaptées au contexte spécifique du FLE.

La typologie de Marquilló Larruy (2003) distingue :

- Les erreurs graphiques (*orthographe d'usage, accents, majuscules*)
- Les erreurs lexicales (*impropriétés, confusions, interférences avec la LI*)
- Les erreurs grammaticales (*morphologie, syntaxe, accords*)
- Les erreurs textuelles (*cohérence, cohésion, progression thématique*)
- Les erreurs pragmatiques (*adéquation à la situation de communication*)

L'étude de Guernine et Kacem (2024) sur l'exploitation de l'erreur dans la production écrite propose une catégorisation plus fine des erreurs grammaticales, distinguant notamment :

- Les erreurs de conjugaison
- Les erreurs d'accord sujet-verbe
- Les erreurs d'accord en genre et en nombre
- Les erreurs dans l'emploi des déterminants
- Les erreurs de construction syntaxique

Cette étude souligne que "*les erreurs syntaxiques sont parmi les plus fréquentes et les plus persistantes chez les apprenants de FLE, même à des niveaux avancés*".

Au-delà de ces catégorisations formelles, il est important de distinguer différents types d'erreurs selon leur origine et leur signification dans le processus d'apprentissage :

- Les erreurs de performance (*lapsus, inattention*) vs les erreurs de compétence (*méconnaissance de la règle*)
- Les erreurs interlinguales (*interférences avec la LI*) vs les erreurs intralinguales (*surgénéralisation des règles du français*)
- Les erreurs systématiques (*récurrentes*) vs les erreurs occasionnelles

Comme la note Corder (1980), "*l'analyse des erreurs permet de mieux comprendre les processus d'acquisition de la langue étrangère et d'adapter l'enseignement en conséquence*".

I.2.4.3.2. Exploitation pédagogique de l'erreur

L'approche contemporaine de l'erreur en didactique du FLE s'inscrit dans un perspectif constructiviste, considérant l'erreur comme une étape nécessaire de l'apprentissage et un indicateur des hypothèses élaborées par l'apprenant sur le fonctionnement de la langue.

L'exploitation pédagogique de l'erreur peut suivre plusieurs démarches complémentaires :

La démarche réflexive, qui implique l'apprenant dans l'identification et l'analyse de ses propres erreurs. Comme le souligne Astolfi (1997), transformer l'erreur en objet de réflexion permet de développer la conscience métalinguistique de l'apprenant. Cette approche peut se concrétiser par :

- Des activités de relecture guidée avec des grilles d'autocorrection
- Des séances collectives d'analyse d'erreurs anonymisées
- La constitution de "*banques d'erreurs*" classées par catégories
- La formulation d'hypothèses sur l'origine des erreurs

La démarche contrastive, qui met en relation les systèmes linguistiques de la L1 et de la L2 pour expliciter les interférences. Selon Marquilló Larruy (2003), la prise de conscience des différences entre les deux langues permet de prévenir certaines erreurs récurrentes. Cette approche est particulièrement pertinente dans des groupes linguistiquement homogènes.

La démarche différenciée, qui adapte le traitement de l'erreur au profil de l'apprenant et à son niveau de compétence. Comme le note Vigner (2001), la tolérance à l'erreur doit être modulée selon les objectifs de l'activité et le stade d'apprentissage. Ainsi, dans les activités de production libre visant la fluidité, une correction exhaustive peut être contre-productive.

L'étude de Guernine et Kacem (2024) démontre l'efficacité d'une approche systématique d'exploitation de l'erreur, comparant un groupe expérimental bénéficiant d'un traitement spécifique des erreurs à un groupe témoin. Les résultats montrent que *"l'exploitation pédagogique des erreurs a permis une amélioration significative de la qualité des productions écrites dans le groupe expérimental"*.

I.2.5. Techniques d'évaluation et de remédiation

Les techniques de correction et de remédiation en production écrite doivent s'inscrire dans une démarche cohérente, visant non seulement à signaler les erreurs mais à permettre à l'apprenant de progresser.

Les modalités de correction peuvent varier selon les objectifs pédagogiques :

- La correction directe, où l'enseignant fournit la forme correcte
- La correction indirecte, où l'enseignant signale l'erreur sans la corriger
- La correction codée, utilisant un système de symboles pour identifier le type d'erreur
- La correction sélective, ciblant certaines catégories d'erreurs selon les objectifs d'apprentissage
- La correction différée, intervenant après une phase d'autocorrection

Comme le souligne Ferris (2002), *"la correction indirecte et codée favorise davantage l'apprentissage à long terme que la correction directe, car elle implique activement l'apprenant dans le processus de révision"*.

Les activités de remédiation visent à traiter les erreurs récurrentes de manière systématique :

- Les exercices ciblés sur des points grammaticaux problématiques
- Les activités de reformulation et de paraphrase
- Les dictées négociées ou les dictées *"zéro faute"*
- La constitution de fiches personnelles de vigilance
- Les ateliers de réécriture ciblant des aspects spécifiques du texte

Selon Cornaire et Raymond (1999), la remédiation est d'autant plus efficace qu'elle s'inscrit dans un processus cyclique, où les erreurs traitées font l'objet d'un suivi dans les productions ultérieures.

L'approche par les pairs constitue également une modalité intéressante de correction et de remédiation. Comme le note Rollinson (2005), *la révision collaborative permet aux apprenants de développer leur capacité critique et leur conscience métalinguistique"*. Cette approche peut prendre la forme de :

- Sessions de révision en binômes ou en petits groupes
- Grilles d'évaluation mutuelle
- Discussions collectives sur des problèmes récurrents

- Réécritures collaboratives de passages problématiques

I.2.5.1. Évaluation de la production écrite

L'évaluation de la production écrite en FLE constitue un enjeu majeur, tant pour mesurer les progrès des apprenants que pour orienter l'enseignement. Cette évaluation doit être cohérente avec les approches didactiques adoptées et les objectifs d'apprentissage visés.

I.2.5.1.1. Critères d'évaluation selon les niveaux et objectifs

Les critères d'évaluation de la production écrite doivent être adaptés au niveau de compétence des apprenants et aux objectifs spécifiques des activités proposées. Le CECR fournit un cadre de référence pour élaborer ces critères, en définissant des descripteurs pour chaque niveau.

Pour les niveaux débutants (A1-A2), l'évaluation peut se concentrer sur :

- Le respect de la consigne et l'adéquation à la situation de communication
- L'intelligibilité globale du message
- La maîtrise d'un répertoire élémentaire de mots et de structures
- L'orthographe des mots courants
- L'utilisation de connecteurs simples (*et, mais, parce que*)

Comme le note Tagliante (2005), « à ces niveaux, la priorité est donnée à la réussite communicative plutôt qu'à la correction formelle ». (P. 107)

Pour les niveaux intermédiaires (B1-B2), les critères s'élargissent pour inclure :

- La cohérence et la cohésion du texte
- La richesse et la précision lexicales
- La variété des structures syntaxiques
- La maîtrise des temps verbaux et des accords
- L'adaptation au genre textuel et au registre de langue

Selon Cuq et Gruca (2005), ces niveaux correspondent à un saut qualitatif important, avec l'introduction de la dimension argumentative et la prise en compte de la complexité textuelle.

Pour les niveaux avancés (C1-C2), l'évaluation porte également sur :

- La finesse de l'argumentation et la nuance de l'expression
- La maîtrise des procédés rhétoriques et stylistiques

- L'originalité et la pertinence du contenu
- La précision dans l'articulation logique des idées
- La capacité à s'adapter à différents genres et contextes

Comme le souligne Beacco (2007), « à ces niveaux, l'accent est mis sur la maîtrise des nuances stylistiques et rhétoriques, ainsi que sur la capacité à adapter sa production à des genres discursifs variés ». (P. 36)

Au-delà de ces critères généraux, l'évaluation doit également prendre en compte les objectifs spécifiques de chaque activité d'écriture : s'agit-il de développer la fluidité, la précision linguistique, la créativité, la capacité argumentative ? Les critères et leur pondération doivent refléter ces priorités pédagogiques.

I.2.5.1.2. Outils d'évaluation formative et sommative

L'évaluation de la production écrite en FLE peut s'appuyer sur une diversité d'outils, adaptés aux différentes fonctions de l'évaluation.

L'évaluation formative, intégrée au processus d'apprentissage, vise à fournir un feedback constructif pour permettre à l'apprenant de progresser. Elle peut s'appuyer sur :

- Des grilles d'observation critériées, identifiant les points forts et les axes d'amélioration
- Des annotations qualitatives sur les productions, suggérant des pistes de réécriture
- Des entretiens individuels d'explicitation du processus d'écriture
- Des portfolios d'écriture, documentant les progrès sur une période donnée
- Des journaux de bord réflexifs, où l'apprenant analyse ses difficultés et ses stratégies

Comme le note Veltcheff et Hilton (2003), l'évaluation formative doit être conçue comme un dialogue entre l'enseignant et l'apprenant, orienté vers la progression.

L'évaluation sommative, qui intervient généralement en fin de séquence ou de parcours, vise à mesurer le niveau atteint par rapport à des objectifs prédéfinis. Elle peut s'appuyer sur :

- Des grilles d'évaluation analytiques, décomposant la compétence en critères précis
- Des grilles d'évaluation holistiques, proposant une appréciation globale de la production
- Des barèmes explicites, pondérant les différents aspects de la production

- Des référentiels standardisés, comme ceux des certifications officielles (DELF, DALF)

Selon Tagliante (2005), l'évaluation sommative gagne en fiabilité et en validité lorsqu'elle s'appuie sur des critères explicites et une pondération réfléchie des différentes dimensions de la production écrite.

Les outils numériques offrent également de nouvelles possibilités pour l'évaluation de la production écrite :

- Les plateformes d'écriture collaborative, qui permettent de suivre le processus de rédaction
- Les logiciels d'analyse textuelle, qui fournissent des données objectives sur certains aspects du texte
- Les portfolios numériques, qui facilitent la conservation et la comparaison des productions
- Les systèmes de feedback automatisé, qui peuvent compléter (mais non remplacer) l'évaluation humaine

I.2.5.1.3. Auto-évaluation et évaluation par les pairs

Le développement de l'autonomie des apprenants passe par leur implication active dans le processus d'évaluation. L'auto-évaluation et l'évaluation par les pairs constituent des modalités complémentaires à l'évaluation par l'enseignant.

L'auto-évaluation permet à l'apprenant de développer sa conscience métalinguistique et sa capacité à réguler son apprentissage. Elle peut être guidée par :

- Des questionnaires réflexifs sur le processus d'écriture
- Des grilles d'auto-correction ciblant des points spécifiques
- Des listes de vérification adaptées au genre textuel et au niveau
- Des journaux d'apprentissage documentant les difficultés rencontrées et les stratégies utilisées

Comme le souligne Little (2007), « *l'auto-évaluation contribue au développement de l'autonomie de l'apprenant, en l'impliquant activement dans la régulation de son apprentissage.* » (P. 34)

L'évaluation par les pairs offre une dimension sociale à l'évaluation et permet de multiplier les retours sur les productions. Elle peut prendre diverses formes :

- Les ateliers de révision collaborative

- Les grilles d'évaluation mutuelle
- Les commentaires croisés sur des productions anonymisées
- Les discussions collectives sur des critères de qualité

Selon Rollinson (2005), « *l'évaluation par les pairs développe non seulement les compétences critiques des apprenants, mais aussi leur capacité à recevoir et à intégrer des feedback constructifs* » (P.18)

Pour être efficaces, ces modalités d'évaluation nécessitent un encadrement approprié :

- Une formation explicite aux techniques d'évaluation
- Des critères clairs et compréhensibles par les apprenants
- Un climat de confiance et de bienveillance
- Un équilibre entre évaluation positive et suggestions d'amélioration
- Une articulation cohérente avec l'évaluation de l'enseignant

Comme la note Huver et Springer (2011), « *l'implication des apprenants dans l'évaluation contribue à développer leur autonomie et leur responsabilité face à leur apprentissage, tout en favorisant une conception plus dynamique et formative de l'évaluation* ». (P. 66)

I.2.5.1.4. Spécificités et valeur ajoutée de l'évaluation humaine

L'évaluation des productions écrites en français langue étrangère constitue un processus complexe qui mobilise de multiples compétences de la part de l'évaluateur. Contrairement aux systèmes d'évaluation automatisée, l'évaluation humaine se caractérise par sa capacité à appréhender les dimensions subjectives et qualitatives des textes produits par les apprenants.

I.2.5.1.4.1. Une approche holistique et contextualisée

L'évaluateur humain, généralement l'enseignant, aborde la production écrite dans sa globalité, en tenant compte simultanément de multiples dimensions. Comme le souligne Bazile (2020) dans son étude sur l'évaluation des productions écrites, « *l'enseignant évalue diversement différentes composantes de la production : la dimension linguistique (connaissance des correspondances phonie/graphie), la cohérence textuelle, l'adéquation à la situation de communication, mais aussi la créativité et l'originalité du propos* » (P.29).

Cette approche holistique permet de saisir l'intention communicative de l'apprenant au-delà des erreurs formelles.

L'évaluation humaine se distingue également par sa capacité à contextualiser le jugement porté sur les productions. L'enseignant peut prendre en compte le parcours d'apprentissage de l'apprenant, ses progrès, ses difficultés spécifiques, et adapter son évaluation en conséquence. Selon Chelghoum (2023), « *l'évaluateur humain possède cette faculté d'ajuster ses critères en fonction du contexte d'apprentissage et des objectifs pédagogiques visés, ce qui confère à son évaluation une pertinence didactique que les systèmes automatisés ne peuvent atteindre* » (P.44).

I.2.5.1.4.2. La dimension interprétative et culturelle

Une caractéristique fondamentale de l'évaluation humaine réside dans sa dimension interprétative. L'enseignant ne se contente pas de vérifier l'application de règles linguistiques, mais interprète le sens du texte, apprécie sa cohérence interne, évalue la pertinence des arguments avancés. Cette dimension herméneutique est particulièrement importante en contexte de langue étrangère, où les productions écrites reflètent souvent des transferts culturels et des modes de pensée issus de la langue maternelle.

L'évaluateur humain est capable de déceler les références culturelles implicites, de comprendre les malentendus interculturels, et d'apprécier les efforts d'appropriation des conventions discursives de la langue cible. Comme l'explique Routel et Missoum (2023), « *l'évaluation de la production écrite en FLE ne peut faire l'économie d'une prise en compte des dimensions culturelles qui sous-tendent l'acte d'écriture, dimensions que seul un évaluateur humain, sensible aux nuances interculturelles, peut véritablement apprécier* » (P.97).

I.2.5.1.4.3. La dimension formative et dialogique

Au-delà de sa fonction certificative, l'évaluation humaine des productions écrites en FLE revêt une dimension formative essentielle. L'enseignant ne se limite pas à attribuer une note ou à signaler des erreurs, mais fournit des retours qualitatifs qui orientent l'apprenant dans son parcours d'apprentissage. Ces retours peuvent prendre diverses formes : commentaires en marge du texte, entretiens individuels, annotations ciblées, suggestions de réécriture.

Cette dimension dialogique de l'évaluation humaine est soulignée par Chelghoum (2023) : « *la communication entre l'évaluateur et l'apprenant constitue un moment privilégié d'apprentissage, où les critères d'évaluation deviennent des outils de médiation permettant à l'apprenant de développer une conscience métalinguistique et métadiscursive* » (P.99). L'évaluation s'inscrit ainsi dans un processus d'étayage, au sens vygotkien du terme, accompagnant l'apprenant dans sa zone proximale de développement.

I.2.5.1.5. Critères et grilles d'évaluation utilisés par les évaluateurs humains

L'évaluation humaine des productions écrites en FLE, pour être rigoureuse et équitable, s'appuie généralement sur des critères explicites et des grilles d'évaluation structurées. Ces outils permettent d'objectiver le jugement de l'évaluateur tout en préservant sa capacité d'appréciation qualitative.

I.2.5.1.5.1. Typologie des critères d'évaluation

Les critères d'évaluation utilisés par les enseignants de FLE peuvent être regroupés en plusieurs catégories, reflétant les différentes dimensions de la compétence scripturale.

Les critères linguistiques concernent la maîtrise du code de la langue :

- L'orthographe lexicale et grammaticale
- La morphosyntaxe (*conjugaison, accords, structure des phrases*)
- La richesse et la précision lexicales
- La ponctuation et la mise en page

Les critères textuels portent sur l'organisation et la cohérence du texte :

- La structure globale et la progression thématique
- L'utilisation des connecteurs logiques
- Les procédés de reprise et d'anaphore
- La cohérence temporelle et l'emploi des temps verbaux

Les critères pragmatiques évaluent l'adéquation à la situation de communication :

- Le respect de la consigne et du genre textuel demandé
- La prise en compte du destinataire
- L'adaptation du registre de langue
- L'efficacité communicative globale

Les critères socio-culturels apprécient l'intégration des dimensions culturelles :

- La connaissance des conventions culturelles liées au genre textuel
- La pertinence des références culturelles mobilisées
- L'adéquation aux attentes rhétoriques de la communauté francophone

Comme le note Bazile (2020), "*ces différentes catégories de critères ne sont pas étanches mais s'articulent dans une vision intégrée de la compétence scripturale, où les dimensions linguistique, textuelle, pragmatique et culturelle s'interpénètrent*".

I.2.5.1.5.2. Modèles de grilles d'évaluation

Plusieurs modèles de grilles d'évaluation ont été développés pour guider les enseignants de FLE dans leur appréciation des productions écrites. Le document du Réseau Canopé (2020) présente notamment trois types de grilles particulièrement pertinentes.

La "grille d'évolution" inspirée des travaux d'Emilia Ferreiro et Jacques Fijalkow propose une approche développementale de l'évaluation, qui situe la production de l'apprenant dans un continuum d'acquisition. Cette grille est particulièrement adaptée pour suivre les progrès des apprenants sur le long terme et identifier les étapes de leur développement scriptural.

Une deuxième grille, empruntée au groupe Écrire (recherche IFÉ), adopte une approche plus analytique, décomposant la compétence scripturale en dimensions précises et observables. Cette grille permet une évaluation fine des différentes composantes de la production écrite et facilite l'identification des points forts et des axes d'amélioration.

La troisième grille, inspirée des indicateurs décrits par Dominique Bucheton et Jean-Charles Chabanne, s'articule autour de trois dimensions fondamentales :

- La dimension énonciative et pragmatique (*qui parle à qui, pourquoi, comment*)
- La dimension sémantique et symbolique (*ce qui est dit, le contenu*)
- La construction d'un rapport à la norme (*correction linguistique et textuelle*)

Cette dernière grille présente l'avantage d'intégrer la dimension réflexive de l'écriture, en considérant la manière dont l'apprenant se positionne par rapport aux normes linguistiques et discursives.

I.2.5.1.5.3. Adaptation des critères selon les niveaux et objectifs

L'un des atouts majeurs de l'évaluation humaine réside dans sa capacité à adapter les critères d'évaluation en fonction du niveau des apprenants et des objectifs pédagogiques visés. Comme le souligne Routel et Missoum (2023), « *l'évaluation doit être équitable et tenir compte du niveau réel des apprenants, en modulant les exigences et en hiérarchisant les critères selon les priorités didactiques.* » (P. 36)

Pour les niveaux débutants (A1-A2), l'accent est généralement mis sur l'intelligibilité du message et le respect de la consigne, avec une tolérance plus grande vis-à-vis des erreurs formelles. Aux niveaux intermédiaires (B1-B2), les critères de cohérence textuelle et de richesse linguistique prennent davantage d'importance. Pour les niveaux avancés (C1-C2), l'évaluation intègre des dimensions plus subtiles comme la nuance stylistique, la finesse argumentative ou la maîtrise des implicites culturels.

Cette modulation des critères selon les niveaux s'accompagne également d'une adaptation aux objectifs spécifiques de l'activité d'écriture. Une production visant la fluidité et la créativité ne sera pas évaluée selon les mêmes critères qu'un texte académique ou professionnel. L'évaluateur humain peut ainsi ajuster sa grille en fonction de la visée communicative de la tâche proposée.

I.2.5.1.5.4. Enjeux de subjectivité et stratégies d'objectivation

L'évaluation humaine des productions écrites, malgré ses nombreux atouts, n'est pas exempte de biais et de subjectivité. La reconnaissance de ces limites et la mise en place de stratégies d'objectivation constituent des enjeux majeurs pour garantir la validité et la fiabilité de l'évaluation.

I.2.5.1.5.4.1. Sources de variabilité dans l'évaluation humaine

Plusieurs facteurs peuvent influencer le jugement de l'évaluateur humain et introduire une variabilité dans l'appréciation des productions écrites.

L'effet de halo désigne la tendance à se laisser influencer par une impression générale ou par certaines caractéristiques saillantes de la production (présentation, calligraphie, premières phrases) au détriment d'une analyse détaillée de l'ensemble du texte. Comme le note Chelghoum (2023), « *l'évaluateur peut être inconsciemment influencé par la qualité formelle de la présentation ou par ses attentes préalables concernant le niveau de l'apprenant* » (P.99).

Les variations inter-évaluateurs constituent une autre source importante de subjectivité. Différents enseignants peuvent interpréter diversement les mêmes critères ou leur accorder une pondération différente. L'étude comparative menée par Bazile (2020) montre que face à une même production écrite, les enseignants peuvent formuler des jugements sensiblement différents, notamment sur les aspects les plus qualitatifs comme la cohérence textuelle ou l'originalité.

La fatigue et les conditions d'évaluation peuvent également affecter la constance du jugement. Un même évaluateur peut varier dans ses appréciations selon le moment de la journée, le nombre de copies déjà corrigées ou le contexte institutionnel dans lequel s'inscrit l'évaluation.

I.2.5.1.5.4.2. Stratégies pour réduire la subjectivité

Face à ces sources de variabilité, diverses stratégies peuvent être mises en œuvre pour objectiver l'évaluation humaine des productions écrites en FLE.

L'explicitation et la communication préalable des critères constituent une première stratégie fondamentale. Comme le démontre la recherche de Chelghoum (2023), « *la communication préalable des critères d'évaluation aux apprenants non seulement améliore leurs performances en production écrite, mais contribue également à clarifier les attentes de l'évaluateur et à réduire la part d'implicite dans le jugement* » (P.99). Cette transparence des critères bénéficie tant aux apprenants qu'aux évaluateurs, en fournissant un cadre de référence commun.

La formation des évaluateurs représente une autre stratégie essentielle. Cette formation peut prendre diverses formes : sessions de standardisation où les enseignants évaluent collectivement des productions types, élaboration concertée de grilles d'évaluation, discussions sur les interprétations des critères. Selon Routel et Missoum (2023), la formation des évaluateurs permet d'harmoniser les pratiques et de construire une culture commune de l'évaluation au sein des équipes pédagogiques.

L'évaluation multiple, où une même production est évaluée par plusieurs enseignants, constitue une stratégie particulièrement efficace pour les évaluations à forts enjeux. Cette approche, courante dans les certifications officielles comme le DELF ou le DALF, permet

de neutraliser les biais individuels et d'accroître la fiabilité du jugement. Comme le souligne Bazile (2020), le croisement des regards évaluatifs enrichit l'appréciation de la production et garantit une plus grande équité dans l'évaluation.

I.2.5.1.6. Complémentarité entre évaluation humaine et outils numériques

Face aux défis de la subjectivité, les outils numériques peuvent constituer des auxiliaires précieux pour l'évaluateur humain, sans pour autant se substituer à son jugement qualitatif.

Les correcteurs orthographiques et grammaticaux permettent d'alléger la charge cognitive liée à la détection des erreurs formelles, libérant l'attention de l'évaluateur pour les aspects plus qualitatifs de la production. Toutefois, comme le note Bazile (2020), « *ces outils doivent être utilisés avec discernement, car ils ne détectent pas toutes les erreurs et peuvent parfois suggérer des corrections inappropriées au contexte* » (P. 68).

Les systèmes d'analyse textuelle automatisée peuvent fournir des indicateurs objectifs sur certains aspects de la production : diversité lexicale, complexité syntaxique, longueur moyenne des phrases, fréquence des connecteurs logiques. Ces données quantitatives viennent enrichir l'appréciation qualitative de l'évaluateur humain.

Les plateformes d'évaluation assistée par ordinateur permettent de structurer le processus d'évaluation en guidant l'enseignant à travers les différents critères et en facilitant la formulation de retours personnalisés. Comme le souligne l'étude comparative de Bazile et al. (2024) sur les pratiques d'évaluation en FLE, l'intelligence artificielle et l'évaluateur humain ne sont pas en concurrence mais en complémentarité, chacun apportant ses forces spécifiques au processus d'évaluation.

Cette complémentarité entre jugement humain et outils numériques représente une voie prometteuse pour concilier les exigences parfois contradictoires de l'évaluation : rigueur et sensibilité, objectivité et prise en compte de la singularité, efficacité et dimension formative.

I.2.5.1.7. Vers une éthique de l'évaluation humaine

Au-delà des compétences techniques, l'évaluation humaine des productions écrites en FLE soulève des questions éthiques fondamentales. L'évaluateur exerce un pouvoir considérable sur l'apprenant, à travers le jugement qu'il porte sur ses productions et les conséquences de ce jugement sur son parcours d'apprentissage.

Une éthique de l'évaluation implique plusieurs principes directeurs :

- L'équité, qui suppose de traiter tous les apprenants selon les mêmes critères explicites, tout en tenant compte de leurs spécificités
- La bienveillance, qui consiste à valoriser les réussites et à considérer les erreurs comme des étapes normales de l'apprentissage
- La transparence, qui exige d'explicitier les critères et les procédures d'évaluation
- L'humilité, qui invite l'évaluateur à reconnaître les limites de son jugement et à rester ouvert à la discussion

Comme le souligne Routel et Missoum (2023), « *l'évaluation n'est jamais un acte neutre, mais s'inscrit dans une relation pédagogique où l'enseignant assume la responsabilité d'un jugement qui peut avoir des effets durables sur la motivation et l'image de soi de l'apprenant* » (P. 88). Cette conscience des enjeux éthiques de l'évaluation constitue peut-être la dimension la plus fondamentale de l'expertise évaluative.

En définitive, l'évaluation humaine des productions écrites en FLE, malgré ses imperfections et sa part irréductible de subjectivité, demeure irremplaçable par sa capacité à saisir la complexité et la richesse de l'acte d'écriture. Les stratégies d'objectivation et les outils numériques ne visent pas à éliminer le jugement humain, mais à le rendre plus rigoureux et plus équitable, tout en préservant sa sensibilité aux dimensions qualitatives et culturelles de la production écrite.

Conclusion partielle

Au terme de cette exploration approfondie de la production écrite en classe de français langue étrangère, plusieurs constats s'imposent. La complexité intrinsèque de l'acte d'écriture, déjà considérable en langue maternelle, se trouve amplifiée dans le contexte de l'apprentissage d'une langue étrangère. L'apprenant de FLE doit en effet jongler simultanément avec l'acquisition du système linguistique français et le développement de compétences scripturales spécifiques, tout en s'appropriant les conventions socioculturelles qui régissent la production écrite dans l'espace francophone. Le cadre théorique que nous avons développé met en lumière les processus cognitifs complexes mobilisés lors de la production écrite en FLE. L'analyse des approches didactiques pour l'enseignement de la production écrite révèle une évolution significative des pratiques pédagogiques, depuis l'approche traditionnelle centrée sur la correction linguistique jusqu'aux approches plus récentes, communicative, actionnelle, par compétences, qui inscrivent l'acte d'écriture dans une perspective fonctionnelle et sociale. Cette évolution s'accompagne d'une diversification des activités proposées aux apprenants, allant des exercices structuraux aux tâches complexes ancrées dans des situations de communication authentiques. L'intégration des technologies numériques ouvre par ailleurs de nouvelles perspectives pour l'enseignement de la production écrite en FLE, en facilitant la collaboration, la révision et la publication des textes. Sur le plan des pratiques pédagogiques, nous avons souligné l'importance d'une progression cohérente dans les activités d'écriture, adaptée aux niveaux et aux besoins des apprenants. Le traitement de l'erreur apparaît comme un enjeu majeur, nécessitant une approche équilibrée entre correction et valorisation des acquis. Les modalités d'évaluation, qu'elles soient formatives ou sommatives, doivent s'appuyer sur des critères explicites et cohérents avec les objectifs d'apprentissage visés. L'implication des apprenants dans le processus d'évaluation, à travers l'auto-évaluation et l'évaluation par les pairs, contribue au développement de leur autonomie et de leur conscience métalinguistique. L'examen de l'évaluation humaine des productions écrites a mis en évidence ses spécificités et sa valeur ajoutée par rapport aux systèmes d'évaluation automatisée. L'approche holistique et contextualisée de l'évaluateur humain, sa capacité à saisir les dimensions interprétatives et culturelles des textes, ainsi que la dimension formative et dialogique de son évaluation, constituent des atouts majeurs pour accompagner efficacement les apprenants dans le développement de leurs compétences scripturales.

Chapitre II

L'évaluation automatisée

Des productions écrites en classe de FLE

Introduction partielle

Dans un contexte éducatif de plus en plus numérisé, l'évaluation des productions écrites en français langue étrangère (FLE) connaît une transformation profonde sous l'impulsion des technologies d'intelligence artificielle. L'émergence de systèmes d'évaluation automatisée soulève des questions fondamentales sur les modalités, la pertinence et l'efficacité de ces nouveaux outils dans l'appréciation des compétences scripturales des apprenants. Ce chapitre se propose d'explorer les multiples facettes de cette révolution technologique qui redéfinit progressivement les pratiques évaluatives dans le domaine du FLE.

L'évaluation des productions écrites a toujours constitué un défi majeur pour les enseignants de langues étrangères. La complexité de cette tâche réside dans la multiplicité des dimensions à prendre en compte : correction linguistique, cohérence textuelle, adéquation pragmatique, richesse lexicale, originalité du propos, entre autres. Traditionnellement, cette évaluation repose sur l'expertise humaine, mobilisant les compétences linguistiques, pédagogiques et interculturelles des enseignants. Cependant, cette approche présente certaines limites : subjectivité des jugements, variabilité inter-évaluateurs, temps considérable requis pour l'analyse des productions, difficulté à fournir un feedback détaillé et personnalisé à grande échelle.

Face à ces défis, les avancées récentes dans le domaine de l'intelligence artificielle offrent des perspectives inédites. Des systèmes comme FIDELIA, développé par France Éducation International, ou le projet CECR-FR Rator de l'Université catholique de Louvain, proposent désormais des solutions d'évaluation automatisée spécifiquement conçues pour les productions écrites en FLE. Plus récemment encore, l'avènement des modèles d'IA générative comme Gemini de Google, chatgpt et autres ouvrent de nouvelles possibilités dans l'analyse et l'évaluation des textes produits par les apprenants.

Ce chapitre vise à adopter une approche à la fois descriptive, analytique et critique. Nous explorerons d'abord les fondements technologiques de l'évaluation automatisée, en nous intéressant particulièrement aux principes de fonctionnement des différents systèmes et à l'apport spécifique des modèles d'IA générative comme Gemini. Nous examinerons ensuite les applications concrètes de ces technologies dans le domaine du FLE, en analysant leur capacité à évaluer les différentes dimensions de la compétence scripturale. La troisième partie sera consacrée aux avantages et aux limites des systèmes d'évaluation automatisée, en identifiant leurs points forts mais aussi les défis qu'ils doivent encore relever. Enfin, nous proposerons une comparaison approfondie entre l'évaluation humaine et l'évaluation automatisée, en explorant les voies d'une complémentarité fructueuse entre ces deux approches.

À travers cette analyse, nous souhaitons contribuer à une réflexion nuancée sur l'intégration des technologies d'intelligence artificielle dans les pratiques d'évaluation en FLE, en dépassant les postures technophiles ou technophobes pour envisager une approche raisonnée, éthique et pédagogiquement pertinente de ces nouveaux outils.

II.1. Fondements technologiques de l'évaluation automatisée des productions écrites

II.1.1. Evolution des technologies d'analyse automatique du langage

L'histoire de l'évaluation automatisée des productions écrites s'inscrit dans l'évolution plus large des technologies d'analyse automatique du langage. Cette évolution, marquée par plusieurs phases distinctes, témoigne d'une sophistication croissante des approches et des outils.

Les premiers outils d'analyse automatique du langage écrit, apparus dans les années 1970-1980, se limitaient essentiellement à la vérification orthographique. Ces correcteurs, basés sur des dictionnaires et des règles simples, permettaient uniquement d'identifier les erreurs d'orthographe lexicale par comparaison avec un lexique de référence. Leur fonctionnement reposait sur une logique binaire : un mot était soit correct (*présent dans le dictionnaire*), soit incorrect (*absent du dictionnaire*).

Dans les années 1990, ces outils se sont enrichis de fonctionnalités de vérification grammaticale, capables de détecter certaines erreurs d'accord, de conjugaison ou de syntaxe. Ces correcteurs grammaticaux, comme le célèbre Antidote pour le français, s'appuyaient sur des règles linguistiques formalisées et des analyses morphosyntaxiques plus élaborées. Toutefois, leur portée restait limitée à l'identification d'erreurs locales, sans véritable prise en compte du contexte global du texte.

Le tournant des années 2000 a vu l'émergence des premiers systèmes d'évaluation automatisée de textes complets, principalement pour l'anglais. Des outils comme E-rater (*Educational Testing Service*) ou Intelligent Essay Assessor (Pearson) proposaient pour la première fois une évaluation multidimensionnelle des productions écrites, prenant en compte non seulement la correction linguistique, mais aussi la cohérence textuelle, la richesse lexicale ou la pertinence du contenu. Ces systèmes, initialement développés pour l'évaluation de l'anglais langue maternelle, ont progressivement été adaptés à d'autres langues et contextes, notamment celui des langues étrangères.

Plus récemment, l'avènement de l'intelligence artificielle et particulièrement des modèles de langage de grande taille (LLM) a révolutionné le domaine. Des systèmes comme FIDELIA pour le français ou les applications basées sur des modèles comme GPT, Claude ou Gemini représentent une nouvelle génération d'outils d'évaluation, capables d'analyses, beaucoup plus fines et contextualisées des productions écrites.

II.1.2. Intelligence artificielle et apprentissage automatique

L'intelligence artificielle, et particulièrement l'apprentissage automatique, a transformé radicalement les possibilités d'évaluation automatisée des productions écrites. Cette évolution s'est faite en plusieurs étapes, chacune marquant un gain significatif en termes de performances et de capacités.

|Chapitre II L'évaluation automatisée Des productions écrites en classe de FLE

Les premières approches d'apprentissage automatique appliquées à l'évaluation de textes reposaient sur des méthodes supervisées classiques : régression linéaire, machines à vecteurs de support (SVM), forêts aléatoires. Ces modèles étaient entraînés sur des corpus de textes préalablement évalués par des humains, apprenant à associer certaines caractéristiques linguistiques (features) à des scores ou niveaux. L'outil FIDELIA, développé par France Éducation International, s'inscrit dans cette approche. Comme l'explique le site officiel de FEI, *"FIDELIA s'appuie sur un modèle d'intelligence artificielle directionnelle. La différence avec un modèle d'IA générative, c'est que le résultat donné à la requête sera toujours le même. C'est donc un modèle plus robuste."*

L'avènement de l'apprentissage profond (*deep learning*) a marqué un tournant décisif. Les réseaux de neurones récurrents (RNN), puis les architectures basées sur l'attention comme les Transformers, ont permis de modéliser des séquences textuelles de manière beaucoup plus fine, en tenant compte des dépendances à longue distance et des subtilités contextuelles. Ces avancées ont considérablement amélioré la capacité des systèmes à évaluer des aspects complexes comme la cohérence textuelle ou l'adéquation pragmatique.

Plus récemment, les modèles de langage de grande taille (LLM) comme GPT, Claude ou Gemini ont ouvert de nouvelles perspectives. Entraînés sur des corpus textuels gigantesques et dotés de milliards de paramètres, ces modèles possèdent une compréhension du langage sans précédent. Leur capacité à générer du texte cohérent et contextuel peut être mise à profit pour l'évaluation des productions écrites, soit directement (*en leur demandant d'évaluer un texte selon certains critères*), soit indirectement (en les utilisant comme composants d'un système d'évaluation plus large).

Dans le domaine spécifique du FLE, ces avancées technologiques permettent désormais d'envisager des évaluations automatisées beaucoup plus nuancées et contextualisées. Comme le note le rapport du projet CECR-FR Rator, *« l'intégration de méthodes d'IA avancées (deep learning) et de connaissances linguistiques permet une modélisation basée sur un large corpus qui représente différentes langues maternelles, âges, origines, niveaux d'acquisition, etc. »* (P.76)

II.1.3. Principes de fonctionnement des modèles génératifs

Les modèles d'IA générative, dont Gemini de Google ou chatgpt et autres sont un représentant emblématique, constituent une catégorie particulière de systèmes d'évaluation automatisée, dont les spécificités méritent une attention particulière. Ces modèles, initialement conçus pour générer du contenu textuel, peuvent également être utilisés pour analyser et évaluer des productions écrites, ouvrant ainsi de nouvelles perspectives dans le domaine du FLE.

Les modèles d'IA générative comme Gemini, ChatGPT (OpenAI) ou Claude (Anthropic) reposent sur des architectures de type Transformer, entraînées de manière auto-supervisée sur d'immenses corpus textuels. Leur objectif initial est de prédire le mot suivant dans une séquence, ce qui les amène à développer une représentation riche et nuancée du langage.

Ces modèles se distinguent par plusieurs caractéristiques fondamentales. Premièrement, leur taille : Gemini Ultra, par exemple, compte des centaines de milliards de paramètres, lui conférant une capacité de modélisation sans précédent. Deuxièmement, leur entraînement multimodal : contrairement aux premiers LLM, ChatGPT ou Gemini peuvent traiter et générer non seulement du texte, mais aussi des images, des vidéos et du code, ce qui enrichit sa compréhension contextuelle. Troisièmement, leur capacité d'apprentissage few-shot ou zéro-shot : ces modèles peuvent s'adapter à de nouvelles tâches avec peu ou pas d'exemples spécifiques.

Dans le contexte de l'évaluation des productions écrites, ces modèles fonctionnent différemment des systèmes traditionnels. Plutôt que d'appliquer des règles prédéfinies ou d'extraire des caractéristiques spécifiques, ils "*lisent*" le texte dans son intégralité et produisent une évaluation basée sur leur compréhension globale du langage et des conventions textuelles qu'ils ont assimilées durant leur entraînement.

Comme l'explique un article récent sur l'utilisation de l'IA générative dans l'enseignement du FLE, ces modèles ne se contentent pas d'identifier des erreurs isolées, mais peuvent appréhender la cohérence globale du texte, sa structure argumentative, et même sa pertinence par rapport à un contexte donné. Cette capacité à saisir les dimensions macro-textuelles représente une avancée significative par rapport aux approches antérieures.

II.1.4. Applications de l'évaluation automatisée dans le domaine du FLE

II.1.4.1. Évaluation des compétences linguistiques

L'évaluation des compétences linguistiques constitue le premier domaine d'application des systèmes d'évaluation automatisée en FLE. Ces systèmes se sont progressivement perfectionnés pour analyser avec une précision croissante les différentes dimensions de la maîtrise linguistique des apprenants.

II.1.4.1.1. Analyse lexicale et richesse du vocabulaire

L'analyse du lexique utilisé par les apprenants représente un aspect fondamental de l'évaluation des productions écrites en FLE. Les systèmes automatisés ont développé des méthodes de plus en plus sophistiquées pour évaluer non seulement la correction lexicale, mais aussi la richesse et la pertinence du vocabulaire employé.

Les approches traditionnelles s'appuient sur des mesures quantitatives comme la diversité lexicale (*ratio type-token*), la densité lexicale (*proportion de mots lexicaux par rapport aux mots grammaticaux*), ou la sophistication lexicale (*présence de mots rares ou complexes*). L'outil FABRA, développé dans le cadre du projet CECR-FR Rator, intègre ces différentes mesures pour caractériser le profil lexical des productions d'apprenants.

Les systèmes plus avancés, comme ceux basés sur ChatGPT, vont au-delà de ces mesures quantitatives pour évaluer la pertinence contextuelle du vocabulaire. Comme l'explique un chercheur spécialisé dans l'IA appliquée à l'enseignement des langues, les modèles génératifs peuvent apprécier si le lexique utilisé est approprié au registre, au genre textuel et à la situation de communication, dimensions essentielles dans l'apprentissage d'une langue étrangère.

Une innovation particulièrement intéressante concerne l'analyse de la progression lexicale des apprenants. Le projet CEFRLex, mentionné dans la documentation du CECR-FR Rator, a établi des distributions de fréquence de phénomènes linguistiques sur les six niveaux du CECR. Ces données permettent aux systèmes automatisés de situer le vocabulaire utilisé par un apprenant sur cette échelle développementale, fournissant ainsi une évaluation plus précise de son niveau de compétence lexicale.

Dans le contexte spécifique du FLE, ces systèmes peuvent également identifier les interférences lexicales avec la langue maternelle de l'apprenant (*faux amis, calques lexicaux*) ou les erreurs typiques liées à certaines familles linguistiques. Cette capacité à contextualiser l'analyse lexicale en fonction du profil de l'apprenant représente une avancée significative par rapport aux approches génériques.

II.1.4.1.2. Évaluation de la correction grammaticale et syntaxique

L'évaluation de la correction grammaticale et syntaxique constitue traditionnellement le cœur des systèmes d'évaluation automatisée. Dans ce domaine, les progrès technologiques ont permis de passer d'une simple détection d'erreurs à une analyse beaucoup plus fine et contextualisée des compétences grammaticales des apprenants.

Les premiers systèmes se limitaient à identifier des erreurs isolées (*accords, conjugaisons, prépositions*) par comparaison avec des règles prédéfinies. Les systèmes actuels, notamment ceux basés sur des architectures neuronales, peuvent analyser des structures syntaxiques complexes et évaluer leur correction en tenant compte du contexte global de la phrase et du texte.

FIDELIA, l'outil développé par France Éducation International, illustre cette approche avancée. Parmi les 48 variables linguistiques qu'il analyse, plusieurs concernent spécifiquement la correction grammaticale et syntaxique : complexité des phrases, diversité des structures syntaxiques, maîtrise des temps et modes verbaux, etc. Comme l'explique le site de FEI, ces variables ont été « *identifiées comme pertinentes pour différencier des niveaux de compétence en langue* ».

Les modèles d'IA générative comme ChatGPT apportent une dimension supplémentaire à cette évaluation. Leur capacité à comprendre le sens global du texte leur permet d'identifier des erreurs grammaticales qui, bien que formellement correctes, sont inappropriées dans le contexte. Par exemple, ils peuvent repérer un temps verbal grammaticalement correct mais inadapté à la situation narrative, ou une structure syntaxique qui, bien que correcte, crée une ambiguïté sémantique.

Les modèles d'IA générative comme ChatGPT apportent une dimension supplémentaire à l'analyse des erreurs. Leur exposition à de vastes corpus de textes, incluant potentiellement des productions d'apprenants de FLE, leur permet de reconnaître des patterns d'erreurs subtils et de les contextualiser. Comme le souligne un article sur l'utilisation de l'IA dans l'enseignement du FLE, "*ces modèles peuvent non seulement identifier l'erreur, mais aussi en comprendre la cause probable et suggérer des stratégies de remédiation adaptées*".

Cette capacité à analyser finement les erreurs spécifiques aux apprenants de FLE représente une valeur ajoutée considérable des systèmes d'évaluation automatisée. Elle permet non seulement une évaluation plus précise du niveau de l'apprenant, mais aussi un feedback plus pertinent et personnalisé, orienté vers la progression.

II.1.4.2. Évaluation des aspects discursifs

Au-delà des compétences linguistiques, les systèmes d'évaluation automatisée ont progressivement développé leur capacité à analyser les dimensions discursives et pragmatiques des productions écrites, aspects particulièrement importants dans l'apprentissage d'une langue étrangère.

II.1.4.2.1. Cohérence et cohésion textuelle

La cohérence et la cohésion textuelles constituent des dimensions essentielles de la compétence scripturale en langue étrangère. Les systèmes d'évaluation automatisée ont considérablement évolué dans leur capacité à analyser ces aspects complexes des productions écrites.

Les approches traditionnelles se concentraient principalement sur des marqueurs formels de cohésion : présence et diversité des connecteurs logiques, utilisation de pronoms et autres procédés anaphoriques, maintien des chaînes de référence. L'outil FABRA, par

exemple, intègre plusieurs variables mesurant ces aspects, comme la densité des connecteurs ou la diversité des procédés de reprise.

Les systèmes plus avancés, notamment ceux basés sur des architectures neuronales, peuvent désormais appréhender la cohérence globale du texte, au-delà de ces marqueurs formels. Comme l'explique un chercheur du projet CECR-FR Rator, « *les modèles de deep learning permettent d'analyser la progression thématique du texte et la logique argumentative, dimensions essentielles de la cohérence textuelle* ».

Les modèles d'IA générative comme Gemini représentent une avancée significative dans ce domaine. Leur capacité à comprendre le sens global d'un texte leur permet d'évaluer si les idées s'enchaînent logiquement, si les arguments sont bien articulés, si les différentes parties du texte forment un ensemble cohérent. Comme le note François T (2022) sur l'utilisation de ChatGPT dans l'évaluation des productions écrites, où il avance que le modèle peut identifier des ruptures de cohérence qui ne sont pas marquées par des erreurs formelles, mais par des incohérences sémantiques ou logiques.

Dans le contexte spécifique du FLE, ces systèmes peuvent également tenir compte des conventions rhétoriques propres au français. La structure argumentative "*thèse-antithèse-synthèse*", par exemple, très valorisée dans la tradition académique française, peut être reconnue et évaluée par ces systèmes. Cette sensibilité aux spécificités culturelles de l'organisation textuelle représente une avancée importante pour l'évaluation des productions en FLE.

II.1.4.2.2. Adéquation à la situation de communication

L'adéquation à la situation de communication constitue une dimension pragmatique essentielle de la compétence en langue étrangère. Les systèmes d'évaluation automatisée ont progressivement intégré cette dimension dans leur analyse des productions écrites.

Les premiers systèmes se limitaient généralement à vérifier le respect formel de la consigne (*nombre de mots, inclusion de certains éléments imposés*). Les systèmes actuels peuvent évaluer de manière beaucoup plus fine l'adéquation pragmatique du texte : pertinence par rapport au sujet, prise en compte du destinataire, adaptation du registre de langue, efficacité communicative globale.

FIDELIA, par exemple, intègre des variables qui permettent d'évaluer si la production répond aux attentes communicatives de la tâche proposée dans le TCF. Comme l'explique France Éducation International, le système a été entraîné sur un corpus diversifié pour « *ne pas privilégier une catégorie de candidats* » et pouvoir évaluer l'adéquation pragmatique dans différents contextes.

Les modèles d'IA générative comme ChatGPT apportent une dimension supplémentaire à cette évaluation pragmatique. Leur exposition à une grande variété de situations de communication leur permet de reconnaître les conventions associées à différents genres textuels (*lettre formelle, récit, argumentation, etc.*) et d'évaluer si la production de l'apprenant respecte ces conventions. Comme le soulignent Makhlouf, M et Legros, D (2024), ces modèles peuvent apprécier si le ton, le style et le registre sont appropriés au contexte communicatif, dimension souvent difficile à évaluer automatiquement.

Dans le contexte du FLE, cette évaluation de l'adéquation pragmatique est particulièrement importante car elle touche à des aspects culturels de la langue. Les conventions communicatives varient considérablement d'une culture à l'autre, et les apprenants de FLE doivent s'approprier non seulement le code linguistique, mais aussi les normes pragmatiques de la communication en français. Les systèmes d'évaluation automatisée les plus avancés peuvent désormais prendre en compte cette dimension interculturelle de la compétence communicative.

II.1.4.3. Études de cas de systèmes d'évaluation automatisée en FLE

Pour illustrer concrètement les applications de l'évaluation automatisée dans le domaine du FLE, il est pertinent d'examiner quelques systèmes représentatifs qui incarnent différentes approches et technologies.

II.1.4.3.1. FIDELIA et l'évaluation des productions du TCF

FIDELIA (*French Intelligent Digital Evaluation for Linguistic Instructional Assessment*) représente un cas emblématique d'application de l'évaluation automatisée dans le contexte du FLE. Développé par France Éducation International en collaboration avec le laboratoire CENTAL de l'Université catholique de Louvain, ce système est spécifiquement conçu pour l'évaluation des productions écrites du Test de Connaissance du Français (TCF).

Le fonctionnement de FIDELIA repose sur un modèle d'IA directionnelle, entraîné sur un corpus soigneusement sélectionné. Comme l'explique le site de France Éducation International, "27 000 productions ont ainsi été sélectionnées en veillant à la diversité du corpus afin ne pas privilégier une catégorie de candidats". Ces productions ont été réévaluées par 55 correcteurs du TCF pour constituer une base de référence fiable.

L'originalité de FIDELIA réside dans son approche hybride de l'évaluation. Le système ne remplace pas les correcteurs humains mais s'intègre dans un processus de double correction : « *A partir de mai 2024, l'évaluation des productions d'expression écrite passées sur ordinateur associe une correction humaine et une correction par FIDELIA. Le niveau délivré au candidat est calculé sur les résultats des deux corrections. En cas de désaccord, une troisième correction, humaine, sera assurée.* » (P.53)

Cette complémentarité entre évaluation humaine et automatisée constitue un point fort du système. Comme le souligne FEI, "dans 70% des cas, la correction des personnes correctrices et celle de FIDELIA est la même", ce qui témoigne de la fiabilité du système. Cependant, l'intervention humaine reste essentielle pour les cas atypiques : « *il se peut que l'intelligence artificielle n'arrive pas à estimer une production qui relève d'un type qu'il n'a jamais vu (production atypique). Un correcteur humain est alors plus à même d'estimer le niveau des individus évalués* ». (P. 66)

L'impact de FIDELIA sur les pratiques d'évaluation est significatif. D'une part, le système permet d'optimiser les ressources humaines : "*France Éducation internationale va ainsi pouvoir réallouer ses ressources à d'autres tâches pour améliorer la qualité des examens et du service.*" D'autre part, il contribue à standardiser l'évaluation, FIDELIA n'étant

"*jamais fatiguée*" et produisant "*des résultats constants : une même production corrigée plusieurs fois, obtiendra toujours le même résultat*".

Ce cas illustre parfaitement comment l'évaluation automatisée peut s'intégrer de manière pertinente dans un dispositif d'évaluation certifiante en FLE, en complémentarité avec l'expertise humaine.

II.1.4.3.2. Projet CECR-FR Rator et la prédiction des niveaux CECR

Le projet CECR-FR Rator, fruit de la collaboration entre le CENTAL (*Université catholique de Louvain*) et France Éducation International, représente une approche plus orientée recherche de l'évaluation automatisée en FLE. Son objectif principal est de décrire les compétences écrites des apprenants du FLE et, sur la base de cette description, à développer un algorithme permettant de prédire automatiquement le niveau du Cadre européen commun de référence pour les langues (CECR) de ces productions écrites.

Ce projet se distingue par sa double ambition, à la fois descriptive et prédictive. D'une part, il vise à décrire de façon généralisée et empirique (*à partir de données d'apprenants*) les compétences de production des apprenants du FLE. D'autre part, il cherche à mettre au point un système d'IA permettant d'analyser automatiquement ces productions en les reliant à un niveau du CECR.

L'approche méthodologique du CECR-FR Rator est particulièrement innovante. Le projet s'appuie sur l'outil FABRA (*French Aggregator-Based Readability Assessment toolkit*), qui permet d'extraire et calculer automatiquement plus de 400 variables linguistiques pertinentes pouvant être regroupées en quatre grandes familles : la longueur, le lexique, la syntaxe et le discours. Cette richesse de variables permet une caractérisation fine des productions d'apprenants.

Une des originalités du projet réside dans son ambition de fournir un diagnostic détaillé des productions analysées : « *nous prévoyons de proposer un diagnostic détaillé du texte analysé en associant un niveau CECR aux phénomènes linguistiques constitutifs de son niveau de compétence (mots, expressions polylexicales, structures syntaxiques, temps et mode des verbes, etc.)* » (P. 82). Cette approche va au-delà de la simple attribution d'un niveau global pour offrir une analyse granulaire des compétences.

Le projet CECR-FR Rator illustre également l'importance de la collaboration entre expertise linguistique et technologies d'IA : « *Il s'agit donc d'une belle collaboration entre les secteurs public et privé qui vise à développer une recherche de pointe sur l'évaluation automatique de la compétence écrite des apprenants du FLE. Une recherche de pointe tant en matière de combinaison de connaissances linguistiques et de méthodes d'IA avancées (deep learning) que de modélisation basée sur un large corpus* » (P. 91)

Ce cas démontre comment l'évaluation automatisée peut contribuer non seulement à des applications pratiques, mais aussi à une meilleure compréhension théorique des processus d'acquisition du FLE, illustrant ainsi la synergie possible entre recherche fondamentale et développement d'outils.

II.1.4.4. Applications pédagogiques de ChatGPT dans l'évaluation formative

Si FIDELIA et le CECR-FR Rator s'inscrivent principalement dans une logique d'évaluation certifiante ou de recherche, les modèles d'IA générative comme Gemini ouvrent des perspectives particulièrement intéressantes dans le domaine de l'évaluation formative en FLE.

Contrairement aux systèmes spécifiquement conçus pour l'évaluation, ChatGPT n'a pas été développé prioritairement pour cette fonction. Cependant, sa capacité à comprendre et générer du langage naturel en fait un outil potentiellement puissant pour l'évaluation formative des productions écrites. Comme l'explique un article sur l'utilisation de l'IA dans l'enseignement du FLE, ces modèles peuvent fournir un feedback immédiat, détaillé et personnalisé sur les productions des apprenants, ce qui est particulièrement précieux dans un contexte d'apprentissage autonome ou à distance.

Plusieurs applications pédagogiques de Gemini dans l'évaluation formative peuvent être identifiées. Premièrement, le modèle peut analyser une production écrite et fournir un feedback détaillé sur différents aspects : correction linguistique, cohérence textuelle, adéquation pragmatique, etc. Ce feedback peut être adapté au niveau de l'apprenant et à ses besoins spécifiques.

Deuxièmement, Gemini peut proposer des suggestions d'amélioration concrètes et contextualisées. Contrairement aux correcteurs traditionnels qui se limitent souvent à signaler les erreurs, le modèle peut expliquer pourquoi certaines formulations sont problématiques et proposer des alternatives plus appropriées. Cette dimension explicative est particulièrement précieuse dans un contexte d'apprentissage.

Troisièmement, le modèle peut s'engager dans un dialogue avec l'apprenant sur sa production, répondant à ses questions, clarifiant certains points, et l'aidant à réfléchir sur ses choix linguistiques et rhétoriques. Cette dimension interactive transforme l'évaluation en un véritable échange pédagogique.

Comme le souligne Cornillie, F (2023), l'intérêt de ces modèles réside moins dans leur capacité à attribuer une note ou un niveau que dans leur potentiel à accompagner le processus d'écriture et de révision. Cette approche s'inscrit parfaitement dans une conception de l'évaluation comme levier d'apprentissage plutôt que comme simple mesure de performance.

Cependant, l'utilisation de ChatGPT dans l'évaluation formative soulève aussi des questions importantes. Comme le note un article de CanoTech sur l'évaluation des productions assistées par IA, il est nécessaire pour l'élève de savoir utiliser les outils numériques de manière critique et constructive, en intégrant les IA comme un soutien à l'écriture plutôt qu'un remplacement de l'effort créatif. Cette préoccupation souligne l'importance d'une intégration réfléchie de ces outils dans les pratiques pédagogiques.

Ce cas illustre comment les modèles d'IA générative comme ChatGPT peuvent enrichir les pratiques d'évaluation formative en FLE, en offrant un feedback personnalisé et interactif qui soutient le développement des compétences scripturales des apprenants.

II.1.4.5. Avantages et limites de l'évaluation automatisée

II.1.4.5.1. Avantages des systèmes automatisés

L'évaluation automatisée des productions écrites en FLE présente de nombreux avantages qui expliquent l'intérêt croissant pour ces technologies dans les contextes éducatifs et certifiants.

II.1.4.5.1.1. Objectivité et constance des évaluations

L'un des avantages majeurs des systèmes d'évaluation automatisée réside dans leur objectivité et leur constance. Contrairement aux évaluateurs humains, dont les jugements peuvent être influencés par divers facteurs (*fatigue, effet de halo, variations d'humeur*), les systèmes automatisés appliquent systématiquement les mêmes critères à toutes les productions.

Cette constance est particulièrement mise en avant par France Éducation International à propos de FIDELIA : « *FIDELIA, de son côté, n'est jamais fatiguée les résultats produits par le modèle numérique sont constants : une même production corrigée plusieurs fois, obtiendra toujours le même résultat.* » (P. 48). Cette fiabilité représente un atout considérable dans les contextes d'évaluation à enjeux élevés, comme les certifications linguistiques.

L'objectivité des systèmes automatisés contribue également à réduire les biais potentiels liés aux caractéristiques personnelles des apprenants. Comme le souligne un chercheur du projet CECR-FR Rator, « *les systèmes automatisés évaluent uniquement la production textuelle, sans être influencés par des facteurs comme l'origine de l'apprenant, son parcours antérieur ou d'autres éléments contextuels qui pourraient inconsciemment affecter le jugement humain* » (P. 98).

Cette objectivité et cette constance sont particulièrement précieuses dans le contexte du FLE, où les apprenants proviennent d'horizons culturels et linguistiques très divers. Elles garantissent une équité de traitement, indépendamment de la familiarité de l'évaluateur avec la culture d'origine de l'apprenant ou des stéréotypes potentiels associés à certaines langues maternelles.

Cependant, comme nous le verrons dans la section sur les limites, cette objectivité apparente peut aussi masquer des biais algorithmiques hérités des données d'entraînement. La question n'est donc pas tant de remplacer le jugement humain par une évaluation supposément plus objective, mais plutôt de combiner les forces des deux approches.

II.1.4.5.1.2. Rapidité et efficacité du traitement

La rapidité et l'efficacité du traitement constituent un autre avantage majeur des systèmes d'évaluation automatisée. Ces systèmes peuvent analyser un grand nombre de productions en un temps très court, ce qui représente un gain considérable par rapport à l'évaluation humaine.

Cette efficacité est particulièrement précieuse dans les contextes d'évaluation à grande échelle, comme les certifications linguistiques internationales ou les cours en ligne ouverts et massifs (*MOOC*). Comme l'explique France Éducation International à propos de FIDELIA, cette automatisation permet de réallouer ses ressources à d'autres tâches pour améliorer la qualité des examens et du service.

Dans le contexte pédagogique, cette rapidité permet également de fournir un feedback immédiat aux apprenants, ce qui constitue un atout considérable pour l'apprentissage. Comme le souligne Moussouri, E (2023) sur l'utilisation de l'IA dans l'enseignement du FLE, la possibilité d'obtenir un retour instantané sur sa production écrite permet à l'apprenant d'identifier immédiatement ses erreurs et de les corriger, renforçant ainsi le processus d'apprentissage par essai-erreur.

Les modèles d'IA générative comme ChatGPT accentuent encore cet avantage en combinant rapidité et personnalisation. Contrairement aux systèmes traditionnels qui peuvent nécessiter un paramétrage spécifique pour chaque type d'évaluation, ces modèles peuvent s'adapter instantanément à différentes tâches et consignes, offrant ainsi une grande flexibilité.

Cette efficacité ne se mesure pas uniquement en termes de temps, mais aussi de ressources. Comme le note un chercheur en technologies éducatives, *"une fois développés et entraînés, ces systèmes peuvent traiter un volume considérable de productions à un coût marginal très faible, ce qui les rend particulièrement intéressants dans un contexte de contraintes budgétaires"*.

II.1.4.5.1.3. Capacité d'analyse multidimensionnelle

Les systèmes d'évaluation automatisée se distinguent également par leur capacité à analyser simultanément de multiples dimensions des productions écrites, offrant ainsi une évaluation potentiellement plus complète et détaillée que celle d'un évaluateur humain.

L'outil FABRA, développé dans le cadre du projet CECR-FR Rator, illustre parfaitement cette capacité d'analyse multidimensionnelle. Comme l'explique la documentation du projet, cet outil permet d'extraire et calculer automatiquement plus de 400 variables linguistiques pertinentes pouvant être regroupées en quatre grandes familles : la longueur, le lexique, la syntaxe et le discours. Cette richesse de variables dépasse largement ce qu'un évaluateur humain peut analyser systématiquement.

Cette analyse multidimensionnelle permet une caractérisation fine du profil linguistique de l'apprenant. Comme le souligne un chercheur de ce projet, le système peut identifier précisément les forces et faiblesses de l'apprenant dans différentes dimensions de la compétence scripturale : richesse lexicale, complexité syntaxique, maîtrise des connecteurs logiques, etc. Cette granularité de l'analyse constitue un atout majeur pour un feedback formatif ciblé.

Les modèles d'IA générative comme ChatGPT enrichissent encore cette capacité d'analyse multidimensionnelle en intégrant des aspects plus qualitatifs et contextuels. Comme on l'a expliqué dans le réseau Canopé (2023) sur l'utilisation de ces modèles dans l'évaluation des productions écrites, qu'au-delà des caractéristiques linguistiques formelles, ces systèmes peuvent appréhender des dimensions comme la créativité, la pertinence argumentative ou l'adéquation culturelle, aspects traditionnellement considérés comme relevant exclusivement du jugement humain.

Dans le contexte spécifique du FLE, cette capacité d'analyse multidimensionnelle permet de mieux appréhender la complexité du développement de la compétence scripturale en langue étrangère. Comme la note, encore une fois, Makhoul et Legros (2024), *« l'apprentissage de l'écrit en FLE implique la maîtrise simultanée de multiples sous-compétences qui évoluent souvent à des rythmes différents ; l'analyse automatisée permet de cartographier précisément ces différentes trajectoires développementales »*. (P. 56)

II.1.4.5.1.4. Feedback détaillé et personnalisé

La capacité à fournir un feedback détaillé et personnalisé constitue l'un des avantages les plus significatifs des systèmes d'évaluation automatisée, particulièrement dans une perspective formative.

Contrairement à l'évaluation humaine, souvent contrainte par le temps et les ressources, les systèmes automatisés peuvent générer des retours très détaillés sur chaque aspect de la production. Comme l'explique un chercheur du projet CECR-FR Rator, le système peut proposer un diagnostic détaillé du texte analysé en associant un niveau CECR aux phénomènes linguistiques constitutifs de son niveau de compétence (*mots, expressions polylexicales, structures syntaxiques, temps et mode des verbes, etc.*). Cette granularité du feedback permet à l'apprenant d'identifier précisément ses points forts et ses axes d'amélioration.

La personnalisation du feedback représente une autre avancée majeure. Les systèmes les plus sophistiqués peuvent adapter leurs retours au profil spécifique de l'apprenant : son niveau, sa langue maternelle, ses erreurs récurrentes, ses objectifs d'apprentissage. Comme le souligne un article sur l'utilisation de l'IA dans l'enseignement du FLE, ces systèmes peuvent construire progressivement un modèle de l'apprenant et personnaliser leurs retours en fonction de son parcours et de ses besoins spécifiques.

Les modèles d'IA générative comme ChatGPT poussent encore plus loin cette personnalisation. Leur capacité à engager un dialogue avec l'apprenant permet d'affiner progressivement le feedback en fonction des questions et des réactions de celui-ci. Comme l'explique Huver, E. (2001), *« ces modèles peuvent non seulement signaler une erreur et l'expliquer, mais aussi répondre aux interrogations de l'apprenant, clarifier certains points, et l'accompagner dans sa réflexion métalinguistique »*. (P. 28)

Dans le contexte du FLE, ce feedback détaillé et personnalisé est particulièrement précieux pour accompagner le développement de l'interlangue de l'apprenant. Comme le note Huver, E. (2001), « *l'apprentissage d'une langue étrangère est un processus hautement individualisé, influencé par de nombreux facteurs comme la langue maternelle, les autres langues connues, le style d'apprentissage, etc. ; un feedback personnalisé permet de mieux accompagner ce processus complexe.* » (P. 35)

II.1.4.5.2. Limites et défis actuels

Malgré leurs avantages indéniables, les systèmes d'évaluation automatisée présentent également des limites et des défis qui doivent être pris en compte pour une utilisation pertinente et éthique de ces technologies.

II.1.4.5.2.1. Difficultés face aux productions atypiques ou créatives

Une des limites majeures des systèmes d'évaluation automatisée concerne leur difficulté à appréhender les productions atypiques ou créatives, qui s'écartent des patterns habituels sur lesquels ils ont été entraînés.

Cette limite est explicitement reconnue par France Éducation International à propos de FIDELIA : « *il se peut que l'intelligence artificielle n'arrive pas à estimer une production qui relève d'un type qu'il n'a jamais vu (production atypique). Un correcteur humain est alors plus à même d'estimer le niveau des individus évalués car le cerveau humain est capable de généralisation et d'appliquer des procédures sur des éléments nouveaux qu'il n'a jamais vus.* » (P. 72)

Cette difficulté s'explique par le fonctionnement même des systèmes d'évaluation automatisée, qui reposent sur l'identification de patterns récurrents dans les données d'entraînement. Face à une production qui utilise des tournures inhabituelles, des métaphores originales, ou une structure argumentative non conventionnelle, ces systèmes peuvent se trouver désorientés et produire des évaluations inappropriées.

Dans le contexte du FLE, cette limite est particulièrement problématique car la créativité linguistique et l'originalité sont souvent valorisées aux niveaux avancés. Comme le souligne Shermis, M. D., & Burstein, J. (Eds.). (2013), « *paradoxalement, plus un apprenant progresse dans sa maîtrise du français, plus il est susceptible de produire des*

textes qui s'écartent des modèles standardisés et qui peuvent donc être mal évalués par les systèmes automatisés ». (P. 18)

Les modèles d'IA générative comme Gemini présentent potentiellement une meilleure capacité à gérer ces productions atypiques, grâce à leur exposition à une plus grande diversité de textes et à leur compréhension plus nuancée du langage. Cependant, même ces modèles avancés peuvent être désorientés par des usages très créatifs ou expérimentaux de la langue.

Cette limite souligne l'importance de maintenir une supervision humaine dans les processus d'évaluation automatisée, particulièrement pour les productions qui s'écartent des standards habituels. Comme le recommandent Chapelle, C. A., et Voss, E. (2016), « *les systèmes automatisés devraient idéalement intégrer des mécanismes de détection des cas atypiques, qui seraient alors redirigés vers des évaluateurs humains* ». (P. 06)

II.1.4.5.2.2. Sensibilité aux aspects culturels et contextuels

Une autre limite importante des systèmes d'évaluation automatisée concerne leur sensibilité aux dimensions culturelles et contextuelles de la communication, aspects particulièrement cruciaux dans l'apprentissage d'une langue étrangère.

La compétence en langue étrangère ne se limite pas à la maîtrise du code linguistique, mais implique également une compréhension des normes culturelles, des implicites, des références partagées qui sous-tendent la communication. Ces dimensions culturelles et contextuelles sont souvent difficiles à modéliser et à évaluer automatiquement.

Comme l'expliquent, encore une fois, Chapelle, C. A., et Voss, E. (2016), « *les productions écrites en FLE sont souvent marquées par des transferts culturels de la culture d'origine de l'apprenant. Ces transferts ne sont pas nécessairement des erreurs, mais peuvent refléter des différences légitimes dans les conventions rhétoriques ou les valeurs culturelles* ». (P.07). Les systèmes automatisés, entraînés principalement sur des normes occidentales ou françaises, peuvent avoir tendance à pénaliser ces expressions culturelles différentes.

Cette limite est particulièrement sensible dans l'évaluation de dimensions comme l'humour, l'ironie, les références culturelles ou les nuances pragmatiques. Ces aspects de la communication sont profondément ancrés dans un contexte culturel spécifique et requièrent souvent une connaissance implicite que les systèmes automatisés peinent à intégrer.

Les modèles d'IA générative comme ChatGPT, bien qu'entraînés sur des corpus multilingues et multiculturels, ne sont pas exempts de ces biais culturels. Leur compréhension des nuances culturelles reste limitée et peut refléter les déséquilibres présents dans leurs données d'entraînement, où certaines cultures et langues sont surreprésentées.

Dans le contexte du FLE, cette limite souligne l'importance d'une approche culturellement sensible de l'évaluation automatisée. Ces systèmes devraient idéalement être entraînés sur des corpus reflétant la diversité des apprenants de FLE et intégrer une sensibilité aux différentes traditions rhétoriques et culturelles.

II.1.4.5.2.3. Questions éthiques et risques de biais algorithmiques

L'utilisation de systèmes d'évaluation automatisée soulève également d'importantes questions éthiques, notamment concernant les risques de biais algorithmiques qui pourraient affecter l'équité de l'évaluation.

Ces biais peuvent provenir de plusieurs sources. Premièrement, les données d'entraînement elles-mêmes peuvent contenir des biais, reflétant des inégalités ou des préjugés présents dans la société. Si un système est entraîné principalement sur des productions d'apprenants issus de certains contextes éducatifs ou culturels, il risque de défavoriser les apprenants provenant d'autres horizons.

Deuxièmement, les critères d'évaluation intégrés dans ces systèmes peuvent refléter des normes linguistiques et culturelles dominantes, au détriment d'autres variétés légitimes du français. En effet, la définition même de ce qui constitue un 'bon français' est culturellement situé et peut varier selon les contextes. Les systèmes automatisés tendent à imposer une norme unique, souvent celle du français standard hexagonal.

Troisièmement, l'opacité de certains systèmes, particulièrement ceux basés sur des architectures neuronales complexes, pose un problème de transparence et d'explicabilité. Comme l'explique François, T (2022), « *si un apprenant reçoit une évaluation négative, il*

devrait pouvoir comprendre les raisons de cette évaluation et éventuellement la contester. Cette transparence est difficile à garantir avec les systèmes de type 'boîte noire' ». (P. 21)

Dans le contexte du FLE, ces questions éthiques sont particulièrement sensibles étant donné la diversité des apprenants et les enjeux potentiellement importants de l'évaluation (*accès à l'université, obtention d'un visa, opportunités professionnelles*). Comme l'ajoute François, T (2022), « *l'évaluation linguistique n'est jamais neutre ; elle s'inscrit dans des rapports de pouvoir et peut contribuer à reproduire ou à transformer des inégalités sociales* ». (P. 28)

Pour répondre à ces préoccupations éthiques, plusieurs principes ont été proposés : transparence des critères d'évaluation, diversité et représentativité des données d'entraînement, possibilité de recours humain, évaluation régulière des biais potentiels. Comme le recommande un rapport sur l'éthique de l'IA dans l'éducation, les systèmes d'évaluation automatisée devraient être conçus et déployés dans une perspective de justice sociale et d'équité, en veillant à ne pas désavantager certains groupes d'apprenants.

II.1.4.5.2.4. Acceptabilité par les apprenants et les enseignants

L'acceptabilité des systèmes d'évaluation automatisée par les principaux acteurs concernés – apprenants et enseignants – constitue un autre défi majeur pour leur intégration réussie dans les pratiques éducatives.

Du côté des apprenants, plusieurs facteurs peuvent influencer l'acceptabilité de ces systèmes. Premièrement, la confiance dans la fiabilité et l'équité de l'évaluation automatisée. Comme le montre une étude de l'université catholique de Louvain (2024) sur les perceptions des apprenants, « *beaucoup d'étudiants expriment des doutes quant à la capacité d'une machine à évaluer justement la qualité et la créativité de leur production* ». Deuxièmement, la qualité et la pertinence du feedback fourni. Les apprenants sont généralement plus réceptifs aux systèmes qui offrent des explications claires et des suggestions constructives, plutôt qu'une simple note ou un niveau.

Du côté des enseignants, l'acceptabilité dépend largement de la manière dont ces systèmes s'intègrent dans leurs pratiques pédagogiques. Comme le souligne une enquête auprès d'enseignants de FLE, beaucoup craignent que ces technologies ne réduisent leur rôle ou ne standardisent excessivement l'évaluation, au détriment d'une approche plus holistique

et personnalisée. Les enseignants s'inquiètent également de leur propre capacité à comprendre et à utiliser efficacement ces outils, soulignant l'importance de la formation et de l'accompagnement.

Dans le contexte spécifique du FLE, ces questions d'acceptabilité sont influencées par des facteurs culturels et institutionnels. En effet, les traditions éducatives varient considérablement selon les pays et les contextes, avec des attitudes très différents vis-à-vis de l'innovation technologique et de l'automatisation dans l'éducation.

Pour favoriser l'acceptabilité de ces systèmes, plusieurs approches ont été suggérées :

Impliquer les enseignants et les apprenants dans la conception et l'évaluation des outils, présenter ces technologies comme des compléments plutôt que des substituts à l'évaluation humaine, assurer la transparence des processus d'évaluation, et proposer des formations adaptées. Donc, l'introduction de ces technologies devrait s'inscrire dans une démarche progressive et participative, respectueuse des cultures éducatives locales et des préoccupations légitimes des acteurs concernés.

II.1.5. Comparaison entre évaluation humaine et évaluation automatisée des productions écrites

II.1.5.1. Analyse comparative des processus d'évaluation

Pour comprendre pleinement les enjeux de l'évaluation des productions écrites en FLE, il est essentiel de comparer méthodiquement les processus d'évaluation humaine et automatisée, en examinant leurs approches respectives, leurs forces et leurs limites.

II.1.5.1.1. Critères et méthodes d'évaluation

Les critères et méthodes d'évaluation constituent un premier axe de comparaison fondamental entre l'évaluation humaine et l'évaluation automatisée des productions écrites en FLE.

L'évaluation humaine traditionnelle s'appuie généralement sur des grilles critériées, comme celles proposées par le Cadre Européen Commun de Référence pour les Langues (CECR). Ces grilles définissent plusieurs dimensions à évaluer : correction linguistique, richesse lexicale, cohérence textuelle, adéquation pragmatique, etc. L'évaluateur humain

attribue des scores pour chaque dimension, souvent selon une approche holistique qui prend en compte l'impression générale produite par le texte.

En effet, l'évaluateur humain procède généralement par une lecture globale suivie d'analyses plus ciblées, dans un va-et-vient constant entre perception d'ensemble et attention aux détails. Cette approche permet une appréciation nuancée qui intègre des dimensions qualitatives difficiles à formaliser : originalité, créativité, sensibilité culturelle, etc.

L'évaluation automatisée, quant à elle, repose sur des critères plus explicitement formalisés et quantifiables. Les systèmes comme FIDELIA s'appuient sur des variables linguistiques précisément définies (*48 dans le cas de FIDELIA*), analysées de manière systématique et objective. Comme l'indique France Éducation International, ces variables ont été identifiées comme pertinentes pour différencier des niveaux de compétence en langue.

Cette différence d'approche se reflète dans la méthodologie d'évaluation. L'évaluateur humain procède souvent de manière non linéaire, avec une attention variable selon les aspects du texte qui lui paraissent saillants. Les systèmes automatisés, en revanche, appliquent systématiquement les mêmes analyses à l'ensemble du texte, sans variation d'attention ou de priorité.

Les modèles d'IA générative comme ChatGPT introduisent une nuance intéressante dans cette comparaison. Ces modèles tentent de reproduire certains aspects de l'évaluation humaine, notamment la capacité à appréhender globalement le texte avant d'en analyser les détails. Cette approche hybride pourrait potentiellement combiner la systématisme des systèmes automatisés traditionnels avec la sensibilité contextuelle de l'évaluation humaine.

Dans le contexte spécifique du FLE, cette différence de critères et de méthodes a des implications importantes. L'évaluation humaine peut mieux prendre en compte la dimension interculturelle de l'apprentissage du français, en valorisant des expressions qui, bien qu'inhabituelles pour un locuteur natif, témoignent d'une appropriation créative de la langue par l'apprenant. Les systèmes automatisés, en revanche, tendent à privilégier la conformité à des normes linguistiques standardisées.

II.1.5.1.2. Fiabilité et validité des évaluations

La fiabilité (*constance des résultats*) et la validité (*pertinence de l'évaluation par rapport à ce qu'elle prétend mesurer*) constituent un deuxième axe de comparaison essentiel entre l'évaluation humaine et l'évaluation automatisée.

L'évaluation humaine présente des défis importants en termes de fiabilité. De nombreuses études ont mis en évidence la variabilité inter-évaluateurs (*différences entre plusieurs correcteurs*) et intra-évaluateur (*variations chez un même correcteur selon le moment ou les conditions*). Comme le résume un chercheur en docimologie de FEI, « *même avec des grilles d'évaluation détaillées et une formation approfondie, des écarts significatifs persistent entre les évaluateurs humains, reflétant la part inévitable de subjectivité dans le jugement* ». (P. 44)

Cette variabilité a conduit à l'adoption de pratiques comme la double correction à l'aveugle, utilisée notamment par France Éducation International pour le TCF. Comme l'explique le site de FEI, « *mettre en place il y a 10 ans la double correction était le prérequis au développement d'un système d'aide à la correction automatisée* ». Cette approche améliore la fiabilité mais augmente considérablement les ressources nécessaires.

L'évaluation automatisée, en revanche, offre une fiabilité parfaite en termes de constance : comme le souligne FEI à propos de FIDELIA, « *une même production corrigée plusieurs fois, obtiendra toujours le même résultat* ». (P. 99). Cette constance représente un avantage majeur dans les contextes d'évaluation à grande échelle ou à enjeux élevés.

Cependant, la question de la validité est plus complexe. L'évaluation humaine, malgré sa variabilité, peut appréhender des dimensions subtiles de la compétence scripturale qui échappent encore aux systèmes automatisés : créativité, finesse argumentative, sensibilité culturelle, etc. La validité de l'évaluation humaine repose sur sa capacité à saisir la complexité et la richesse de l'expression écrite dans toutes ses dimensions, y compris celles qui résistent à la formalisation.

Les systèmes automatisés, bien que constants dans leurs résultats, peuvent présenter des limites en termes de validité, particulièrement pour les aspects qualitatifs et contextuels de l'écriture. Comme le note un chercheur du projet CECR-FR Rator, « *la question n'est pas tant de savoir si ces systèmes sont fiables (ils le sont, par définition), mais s'ils mesurent effectivement ce que nous voulons mesurer : la compétence scripturale dans toute sa complexité* ». (P. 84)

Dans le contexte du FLE, cette tension entre fiabilité et validité est particulièrement sensible. Comme le souligne Moirand, S. (1979), l'apprentissage d'une langue étrangère implique non seulement l'acquisition de compétences linguistiques formelles, mais aussi l'appropriation de modes de pensée et d'expression culturellement situés. L'évaluation doit pouvoir capturer cette dimension culturelle, ce qui reste un défi pour les systèmes automatisés.

II.1.5.1.3. Temps et ressources nécessaires

Le temps et les ressources nécessaires constituent un troisième axe de comparaison particulièrement pertinent entre l'évaluation humaine et l'évaluation automatisée.

L'évaluation humaine est généralement chronophage. En tant qu'enseignante, l'évaluation détaillée d'une production écrite de niveau avancé peut, parfois, me prendre 20 à 30 minutes par copie, sans compter le temps nécessaire pour formuler un feedback constructif. Cette contrainte temporelle limite considérablement le nombre et la fréquence des évaluations possibles, particulièrement dans les contextes d'enseignement avec des effectifs importants.

Au-delà du temps, l'évaluation humaine mobilise des ressources humaines qualifiées, dont la formation et la coordination représentent un investissement significatif. Comme le souligne France Éducation International à propos du TCF, la mise en place d'un système de double correction implique une logistique complexe et coûteuse.

L'évaluation automatisée, en revanche, offre des gains considérables en termes de temps et de ressources une fois le système développé et déployé. Comme l'explique Barbot, M.-J., et Pugibet, V. (2002), deux chercheurs en technologies éducatives, « *le coût marginal d'une évaluation supplémentaire est pratiquement nul, ce qui permet d'envisager des évaluations beaucoup plus fréquentes et systématiques* ». (P. 23). Cette efficacité représente un atout majeur dans les contextes d'évaluation à grande échelle ou pour le feedback formatif régulier.

Cependant, il convient de nuancer cette comparaison en prenant en compte les coûts de développement et de maintenance des systèmes automatisés. Comme le note un expert en IA éducative que la création d'un système comme FIDELIA représente un investissement initial considérable en termes de recherche, de développement et de constitution de corpus

annotés. Ces coûts peuvent être prohibitifs pour des contextes éducatifs disposant de ressources limitées.

Dans le cas spécifique des modèles d'IA générative comme ChatGPT, la question des ressources se pose différemment. Ces modèles généraux, développés pour de multiples usages, peuvent être adaptés à l'évaluation des productions écrites sans nécessiter le développement d'un système spécifique. Cependant, leur utilisation implique généralement des coûts d'abonnement ou d'API qui peuvent représenter une charge récurrente significative.

II.1.5.2. Complémentarité des approches

Au-delà de la simple comparaison, l'analyse des évaluations humaine et automatisée révèle des complémentarités potentiellement fécondes, qui ouvrent la voie à des approches hybrides plus riches et plus pertinentes.

II.1.5.2.1. Systèmes hybrides d'évaluation

Les systèmes hybrides, combinant évaluation humaine et automatisée, représentent une voie prometteuse pour tirer parti des forces complémentaires de ces deux approches.

FIDELIA illustre parfaitement cette approche hybride. Comme l'explique France Éducation International, « *l'évaluation des productions d'expression écrite passées sur ordinateur associe une correction humaine et une correction par FIDELIA. Le niveau délivré au candidat est calculé sur les résultats des deux corrections. En cas de désaccord, une troisième correction, humaine, sera assurée.* » (P. 33). Cette méthodologie reconnaît à la fois la valeur de la constance de l'évaluation automatisée et l'importance du jugement humain pour les cas complexes ou atypiques.

Plusieurs modèles de systèmes hybrides peuvent être envisagés. Dans un premier modèle, l'évaluation automatisée sert de "premier filtre", traitant l'ensemble des productions et identifiant celles qui nécessitent une attention humaine particulière. Comme l'explique, encore une fois, Barbot, M.-J., et Pugibet, V. (2002), « *le système peut facilement identifier les cas 'standards' et les évaluer de manière fiable, permettant aux évaluateurs humains de concentrer leur attention sur les productions atypiques, créatives ou particulièrement complexes* ». (P. 25)

Dans un deuxième modèle, l'évaluation automatisée et humaine se déroulent en parallèle, chacune se concentrant sur des aspects différents de la production. Par exemple, le système automatisé pourrait évaluer principalement les aspects formels (*correction grammaticale, richesse lexicale, etc.*) tandis que l'évaluateur humain se concentrerait sur les dimensions plus qualitatives (*pertinence argumentative, créativité, sensibilité culturelle*). Comme le suggère Mahfoud, Z. (2024), cette répartition des tâches permet d'optimiser les ressources en exploitant les forces respectives des deux approches.

Dans un troisième modèle, l'évaluation automatisée sert de support à l'évaluateur humain, lui fournissant des analyses détaillées qui enrichissent et objectivent son jugement. Comme l'explique un chercheur du projet CECR-FR Rator, « *le système peut identifier systématiquement certains patterns linguistiques que l'évaluateur humain pourrait manquer, tout en laissant à ce dernier la responsabilité du jugement final* ». (P. 71)

Dans le contexte du FLE, ces approches hybrides sont particulièrement pertinentes étant donné la complexité et la multi dimensionnalité de la compétence scripturale en langue étrangère. Comme le souligne Tagliante, C. (2005), l'évaluation du FLE implique à la fois des aspects techniques facilement formalisables et des dimensions culturelles et créatives qui requièrent une sensibilité humaine. Les systèmes hybrides permettent de concilier ces deux exigences.

II.1.5.2.2. Répartition optimale des tâches

La complémentarité entre évaluation humaine et automatisée invite à réfléchir à une répartition optimale des tâches, qui exploite au mieux les forces respectives de chaque approche.

Les systèmes automatisés excellent dans certaines tâches spécifiques : analyse systématique de caractéristiques linguistiques formelles, traitement rapide de grands volumes de productions, application constante de critères explicites, détection de patterns récurrents. Ces systèmes peuvent analyser en quelques secondes des centaines de variables linguistiques avec une précision et une constance qu'aucun évaluateur humain ne pourrait atteindre.

Les évaluateurs humains, quant à eux, conservent un avantage décisif pour d'autres aspects de l'évaluation : appréciation de la créativité et de l'originalité, sensibilité aux nuances culturelles et pragmatiques, compréhension des intentions communicatives, évaluation de la pertinence argumentative dans des contextes complexes. Comme le soulignent Veltcheff, C., et Hilton, S. (2003), « *l'évaluateur humain peut saisir intuitivement des dimensions de la communication écrite qui échappent encore largement à la formalisation algorithmique* ». (P. 19)

Une répartition optimale des tâches pourrait donc confier aux systèmes automatisés l'analyse des aspects formels et quantifiables de la production, tout en réservant aux évaluateurs humains le jugement sur les dimensions plus qualitatives et contextuelles. Comme le suggèrent, encore une fois, Veltcheff, C., & Hilton, S. (2003), « *le système pourrait fournir une analyse détaillée de la correction grammaticale, de la richesse lexicale ou de la complexité syntaxique, tandis que l'évaluateur humain se concentrerait sur l'adéquation pragmatique, la pertinence argumentative ou la créativité* ». (P. 19)

Cette répartition peut également s'envisager selon les niveaux de compétence. Comme l'explique un didacticien du FLE, "*aux niveaux débutants (A1-A2), où l'évaluation porte largement sur la maîtrise de structures linguistiques de base, les systèmes automatisés peuvent être particulièrement pertinents. Aux niveaux avancés (B2-C2), où la finesse d'expression et la sensibilité culturelle deviennent cruciales, l'intervention humaine reste indispensable*".

Dans le contexte spécifique du FLE, cette répartition des tâches doit également prendre en compte la diversité des profils d'apprenants et des contextes d'apprentissage. Comme le notent Mangenot, F., et Soubrié, T. (2014)., l'équilibre optimal entre évaluation humaine et automatisée peut varier considérablement selon qu'il s'agit d'un contexte universitaire, professionnel, ou d'intégration sociale, chacun valorisant différemment certains aspects de la compétence scripturale.

II.1.5.2.3. Cas de FIDELIA : double correction humaine et automatisée

Le cas de FIDELIA, développé par France Éducation International, offre un exemple concret et instructif de complémentarité entre évaluation humaine et automatisée dans le contexte spécifique du FLE.

Comme l'explique le site de FEI, FIDELIA s'intègre dans un processus de double correction : A partir de mai 2024, l'évaluation des productions d'expression écrite passées sur ordinateur associe une correction humaine et une correction par FIDELIA. Le niveau délivré au candidat est calculé sur les résultats des deux corrections. En cas de désaccord, une troisième correction, humaine, sera assurée.

Cette approche repose sur plusieurs principes intéressants. Premièrement, elle reconnaît la valeur complémentaire des deux types d'évaluation, sans hiérarchiser a priori l'une par rapport à l'autre. Le niveau final est calculé en prenant en compte les deux évaluations, ce qui suggère une confiance équivalente dans les deux approches.

Deuxièmement, elle prévoit un mécanisme de résolution des désaccords qui fait appel à une expertise humaine supplémentaire. Comme le précise FEI, « *en cas de désaccord, une troisième correction, humaine, sera assurée* ». (P. 39). Ce choix reflète une reconnaissance des limites potentielles de l'évaluation automatisée face à certaines productions atypiques.

Troisièmement, elle s'appuie sur une analyse empirique de la concordance entre les deux types d'évaluation. Comme l'indique FEI, dans 70% des cas, la correction des personnes correctrices et celle de FIDELIA est la même. Cette statistique témoigne à la fois de la fiabilité du système automatisé et de l'existence d'un espace significatif (30%) où les deux approches divergent, justifiant leur complémentarité.

L'expérience de FIDELIA met également en lumière les bénéfiques pratiques de cette complémentarité. D'une part, elle permet d'optimiser les ressources humaines : France Éducation international va ainsi pouvoir réallouer ses ressources à d'autres tâches pour améliorer la qualité des examens et du service. D'autre part, elle renforce la robustesse et la légitimité de l'évaluation : FIDELIA nous permet d'obtenir une évaluation similaire à ce que nous obtiendrions si la production écrite était corrigée par un ensemble de correcteurs.

Dans le contexte plus large de l'évaluation en FLE, l'approche de FIDELIA offre un modèle potentiellement transposable à d'autres situations, avec les adaptations nécessaires. Comme le suggère Lukogho Vagheni, G. (2019), ce modèle de double correction pourrait inspirer des pratiques similaires dans d'autres contextes, comme l'évaluation universitaire ou la certification professionnelle, en ajustant la pondération respective des deux évaluations selon les enjeux et les objectifs spécifiques.

II.1.5.3. Impact sur les pratiques pédagogiques

Au-delà des aspects techniques et méthodologiques, la complémentarité entre évaluation humaine et automatisée transforme progressivement les pratiques pédagogiques dans l'enseignement du FLE.

II.1.5.3.1. Évolution du rôle de l'enseignant-évaluateur

L'intégration de systèmes d'évaluation automatisée dans les pratiques pédagogiques entraîne une évolution significative du rôle de l'enseignant-évaluateur de FLE.

Traditionnellement, l'enseignant assume l'intégralité du processus d'évaluation, de la conception des tâches à la correction détaillée des productions, en passant par la formulation du feedback. Cette approche, bien que riche pédagogiquement, représente une charge de travail considérable qui limite souvent la fréquence et la profondeur des évaluations.

Avec l'introduction de systèmes d'évaluation automatisée, le rôle de l'enseignant se transforme. L'enseignant peut déléguer certaines tâches répétitives et chronophages aux systèmes automatisés, pour se concentrer sur des aspects plus qualitatifs et personnalisés de l'évaluation. Cette redistribution des tâches peut permettre un accompagnement plus individualisé des apprenants.

Cette évolution s'accompagne d'un déplacement de l'expertise de l'enseignant. Au-delà de sa compétence linguistique et didactique traditionnelle, l'enseignant doit désormais développer une "*littératie numérique*" spécifique : capacité à choisir les outils appropriés, à interpréter leurs résultats, à identifier leurs limites, à les intégrer de manière pertinente dans sa pratique pédagogique. L'enseignant devient un médiateur entre les systèmes automatisés et les apprenants, un rôle qui requiert de nouvelles compétences professionnelles.

Cette transformation suscite des réactions contrastées parmi les enseignants de FLE. Certains y voient une opportunité d'enrichir leur pratique et de se concentrer sur des aspects plus gratifiants de leur métier. Comme le témoigne un enseignant cité dans une étude récente en France, les outils automatisés me libèrent du temps pour des interactions plus qualitatives avec mes étudiants et pour une réflexion plus approfondie sur ma pédagogie. D'autres expriment des inquiétudes quant à une possible dévaluation de leur expertise ou à une standardisation excessive de l'évaluation.

Dans le contexte spécifique du FLE, cette évolution du rôle de l'enseignant-évaluateur doit prendre en compte la dimension interculturelle de l'enseignement. L'enseignant de FLE n'est pas seulement un évaluateur de compétences linguistiques, mais aussi un médiateur culturel qui accompagne l'apprenant dans son appropriation d'une nouvelle langue-culture. Cette dimension de médiation culturelle reste difficilement automatisable.

II.1.5.3.2. Nouvelles approches de l'enseignement de l'écrit

La complémentarité entre évaluation humaine et automatisée favorise également l'émergence de nouvelles approches dans l'enseignement de l'écrit en FLE.

Traditionnellement, l'enseignement de l'écrit s'organise souvent autour d'un nombre limité de productions substantielles, évaluées de manière approfondie mais peu fréquente. Cette approche, contrainte par les ressources disponibles, limite les opportunités de pratique et de feedback régulier.

L'intégration de systèmes d'évaluation automatisée permet d'envisager des approches plus itératives et progressives. Les apprenants peuvent désormais produire des textes plus fréquemment, recevoir un feedback immédiat sur certains aspects de leur production, réviser leur texte en conséquence, et progresser plus rapidement grâce à ces cycles courts d'écriture-feedback-révision.

Cette évolution s'inscrit dans une conception plus processuelle de l'écriture, qui met l'accent sur les différentes étapes du processus rédactionnel plutôt que sur le seul produit final. Comme le souligne un article de CanoTech sur l'évaluation des productions assistées par IA, « *au-delà de l'attention portée à la conformité de la production finale vis-à-vis des*

attendus, suivre l'ensemble du processus d'écriture depuis la planification jusqu'à la révision vise à mieux situer et accompagner l'élève dans sa démarche ». (P. 77)

Les modèles d'IA générative comme ChatGPT enrichissent encore ces possibilités en offrant un feedback dialogique et personnalisé. Ces modèles peuvent engager une conversation avec l'apprenant sur sa production, répondre à ses questions, suggérer des pistes d'amélioration, et l'accompagner tout au long du processus d'écriture. Cette dimension interactive transforme l'évaluation en un véritable échange pédagogique.

Dans le contexte spécifique du FLE, ces nouvelles approches doivent s'adapter à la diversité des traditions éducatives et des styles d'apprentissage. Les apprenants de FLE proviennent de contextes éducatifs très divers, avec des conceptions parfois très différentes de l'écrit, de l'évaluation, et de l'autorité pédagogique. L'introduction de ces nouvelles approches doit tenir compte de ces différences culturelles.

II.1.5.3.3. Autonomisation des apprenants face à l'évaluation

La complémentarité entre évaluation humaine et automatisée ouvre également des perspectives intéressantes pour l'autonomisation des apprenants de FLE face à l'évaluation de leurs productions écrites.

Traditionnellement, l'évaluation est largement perçue comme une activité externe à l'apprenant, réalisée par l'enseignant ou l'institution. Cette conception peut favoriser une certaine passivité face au processus d'évaluation, l'apprenant se contentant de recevoir un jugement sur sa production sans nécessairement s'approprier les critères ou participer activement à l'analyse de son texte.

L'intégration de systèmes d'évaluation automatisée, particulièrement ceux offrant un feedback détaillé et immédiat, peut transformer cette dynamique. Comme l'explique Lukogho Vagheni, K. (2020), ces outils permettent à l'apprenant de recevoir un retour instantané sur sa production, d'identifier lui-même ses erreurs et ses points d'amélioration, et de développer progressivement une capacité d'auto-évaluation.

Cette autonomisation peut prendre plusieurs formes. Premièrement, l'utilisation régulière de ces outils peut familiariser l'apprenant avec les critères d'évaluation et

développer sa conscience métalinguistique. Comme le souligne Fulep, J. (2023), en voyant systématiquement quels aspects de sa production sont évalués et comment, l'apprenant intériorise progressivement ces critères et développe un regard plus analytique sur sa propre écriture.

Deuxièmement, certains systèmes permettent à l'apprenant de paramétrer lui-même l'évaluation, en choisissant les aspects sur lesquels il souhaite recevoir un feedback prioritaire. Cette personnalisation favorise une approche plus stratégique et réflexive de l'apprentissage. Comme l'explique, encore une fois Fulep, J. (2023), l'apprenant devient coconcepteur de son parcours d'apprentissage, en identifiant lui-même ses besoins et en orientant l'évaluation en conséquence.

Troisièmement, les modèles d'IA générative comme ChatGPT offrent la possibilité d'un dialogue avec l'apprenant sur sa production. Cette dimension interactive peut stimuler une réflexion métacognitive sur le processus d'écriture. Comme le notent Cornillie, F., Desmet, P., & Baten, K. (2023), « *l'apprenant peut questionner le système sur les raisons de certaines évaluations, explorer des alternatives, et développer ainsi une compréhension plus profonde des mécanismes de la langue* ». (P. 39)

Dans le contexte spécifique du FLE, cette autonomisation face à l'évaluation représente un enjeu particulièrement important. L'apprentissage d'une langue étrangère est un processus à long terme qui se poursuit bien au-delà du cadre formel de l'enseignement. Développer la capacité des apprenants à évaluer et améliorer leurs propres productions est donc essentiel pour un apprentissage durable.

Conclusion partielle

Au terme de cette exploration approfondie de l'évaluation automatisée des productions écrites en français langue étrangère, plusieurs constats s'imposent et des perspectives se dessinent pour l'avenir de ces technologies dans le domaine de l'enseignement et de l'évaluation du FLE.

L'évaluation automatisée des productions écrites en FLE, propulsée par l'Intelligence Artificielle, est un domaine en pleine effervescence qui offre des outils prometteurs pour enrichir les pratiques évaluatives. Toutefois, son déploiement doit être guidé par une approche pragmatique et critique, reconnaissant ses limites actuelles et s'assurant qu'elle reste au service des objectifs pédagogiques et du développement global des compétences des apprenants. La clé résidera dans la capacité des acteurs de l'éducation à orchestrer une collaboration intelligente entre la machine et l'humain, où la technologie augmente les capacités de l'enseignant sans jamais chercher à le supplanter dans ce qui fait l'essence même de son métier : l'accompagnement humain et éclairé de chaque apprenant sur le chemin de la maîtrise de la langue.

L'évolution rapide des technologies d'intelligence artificielle a considérablement transformé le paysage de l'évaluation des productions écrites. Des systèmes comme FIDELIA, développé par France Éducation International, ou le projet CECR-FR Rator de l'Université catholique de Louvain, témoignent de la sophistication croissante de ces outils, capables désormais d'analyser des dimensions multiples de la compétence scripturale. Plus récemment, l'émergence des modèles d'IA générative comme ChatGPT a ouvert de nouvelles perspectives, notamment en termes de feedback personnalisé et interactif.

Ces avancées technologiques s'accompagnent d'avantages indéniables : objectivité et constance des évaluations, rapidité et efficacité du traitement, capacité d'analyse multidimensionnelle, possibilité d'un feedback détaillé et personnalisé. Dans un contexte d'internationalisation de l'enseignement du français et de diversification des modalités d'apprentissage, ces atouts répondent à des besoins réels et croissants.

Chapitre III

L'expérimentation

Introduction partielle

Dans ce dernier chapitre, nous allons passer à l'expérimentation, vérifier la validation de nos hypothèses et essayer de répondre à notre problématique, dans le sens de la comparaison entre les deux types de l'évaluation humaine et artificielle d'un lot de productions écrites d'une classe de 1^{ère} AM. Nous allons expliquer comment l'IA a contribué à faciliter et à améliorer la tâche d'évaluation à un enseignant.

Nous allons, d'abord, commencer par les objectifs fixés dans cette recherche, la démarche méthodologique que nous avons adoptée, en évoquant la méthode que nous avons utilisée, en précisant le public ciblé, le cadre spatio-temporel de notre étude, ainsi que les outils que nous avons mobilisés. Nous présenterons ensuite les résultats des données recueillies et procéderons à leur analyse.

III.1. Démarche méthodologique

III.1.1. Objectifs de la recherche

Nous examinerons les implications pédagogiques liées à l'utilisation des outils d'intelligence artificielle dans l'évaluation des productions écrites en français. Cette réflexion nous amènera à explorer l'impact de ces outils sur l'objectivité des résultats d'évaluation et sur la perception des évaluations effectuées par l'IA, en les comparant aux méthodes traditionnelles (*de l'enseignant*).

III.1.2. La méthode utilisée

Nous avons choisi la méthode expérimentale et qualitative qui donne la priorité à l'expérience, à l'analyse des contenus et à l'interprétation des résultats non chiffrés :

« La méthode expérimentale consiste à soumettre un phénomène à des conditions contrôlées afin d'en observer les réactions et d'en dégager des lois générales. Elle permet ainsi de passer de l'observation à l'explication, voire à la prédiction. »

(Jean-Pierre Astolfi, L'erreur, un outil pour enseigner, 1997)

III.1.3. Le public visé

Il s'agit d'un groupe éducatif de 1^{ère} année moyenne. Une classe dont le nombre est de 38 apprenants, âgés entre 11 et 13 ans.

III.1.4. Présentation de l'établissement

Le collège d'enseignement moyen est nommé, Belouaer Belgassem, c'est un ancien établissement construit à la commune d'Al Anasser-wilaya de Bordj Bou-Arréridj. Ce collège contient 32 classes. Dont neuf groupes de 1^{ère} AM, le nombre des élèves dans chaque classe est environ 37 et 47 et 58 enseignants, dont huit enseignants de FLE.

III.1.5. Le cadre spatio-temporel

Notre étude s'est déroulée pendant l'année scolaire 2024/2025, en tant qu'enseignante de langue française dans l'établissement scolaire de Belouaer Belgassem, dans la commune d'El Anasser, wilaya de Bordj Bou-Arréridj.

J'ai établi l'expérimentation avec une classe de 1^{ère} année moyenne durant le premier trimestre, pendant la 3^{ème} séquence qui concerne les sports, intitulée « l'importance de bouger régulièrement ».

III.1.6. Le matériel utilisé

Pour réaliser notre expérimentation, nous avons suivi le programme scolaire lors de la séance de production écrite. Ainsi, nous avons utilisé un tableau pour exposer la situation d'écriture, les consignes et les critères de réussite, ainsi qu'une fiche pédagogique pour l'enseignant. Par ailleurs, les élèves n'ont de recours qu'à une double-feuille pour effectuer leurs rédactions.

III.1.7. Le choix des critères de réussite

➤ Selon ChatGPT

Nous avons questionné **ChatGPT** à propos des origines et des références dont il s'est servi lors de la création de la grille d'évaluation, et sa réponse a été que : Les principales inspirations viennent **des programmes de l'Éducation nationale française**, surtout pour le cycle 3, qui insistent sur l'importance de bien maîtriser la langue, d'assurer la cohérence

textuelle, ainsi que sur l'orthographe et la syntaxe. Il s'est également appuyé sur le **CECRL** (*cadre Européen Commun de Référence pour les Langues*) pour définir les compétences langagières attendues propres à chaque niveau. En plus, il s'est inspiré des **grilles d'évaluation déjà utilisées** par des enseignants de FLE/FLS ou de Français au primaire et au collège, intégrant des critères comme la compréhension du sujet, la structure du texte, la richesse des idées, la correction linguistique, ainsi que la maîtrise des temps verbaux et de la ponctuation. Enfin, il mentionne les ressources de formation pédagogique comme celles des **INSPE** (*France*) ou de guides enseignants belges et québécois qui proposent des grilles à plusieurs domaines selon le type de production écrite.

➤ Selon l'enseignante

En règle générale, les grilles d'évaluation employées dans le programme scolaire reposent sur les points linguistiques de la séquence à laquelle se rapporte la production écrite, ainsi que sur l'ensemble des règles grammaticales, y compris la conjugaison ; la syntaxe et l'orthographe. Cependant, certains critères d'évaluation sont souvent utilisés par les enseignants de manière implicite, sans être explicitement mentionnés, tels que le style personnel de l'élève (*qu'il s'agisse d'une production originale ou d'un plagiat*) ainsi que la qualité des explications, qui varient en fonction du niveau de chaque élève. Par exemple, un travail jugé moyen par un élève de niveau excellent pourra être considéré comme très bon s'il provient d'un élève de niveau faible ou moyen. Cela reflète les efforts fournis par chaque élève dans leurs travaux ainsi que leurs intentions respectives. Ces deux critères ont été intégrés à notre grille d'évaluation afin de mettre en évidence l'apport humain dans notre travail.

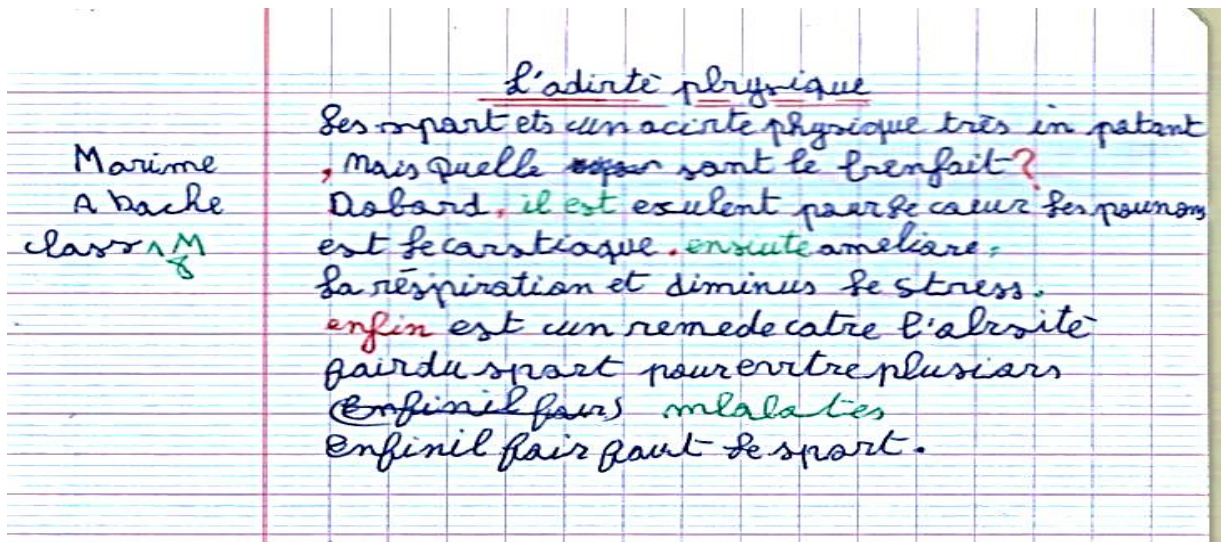
III.2. Déroulement de l'expérimentation

III.2.1. Corpus

III.2.1.1. Evaluation de ChatGPT des copies des élèves

Nous avons recopié l'évaluation authentique telle qu'elle a été faite par ChatGPT, aucune modification.

Voici l'évaluation de la copie 1 manuscrite intitulée « *L'activité physique* ». Le texte a été écrit par **Marieme Abache**, en **classe de 5e M**.



Q 1. Observation du contenu

✓ Points positifs :

- Le sujet est bien compris : l'élève traite les bienfaits du sport sur la santé.
- Utilisation de connecteurs logiques : *d'abord, ensuite, enfin*.
- L'idée générale est claire : le sport est bénéfique pour le corps et la santé mentale.

△ Points à améliorer :

- Des fautes de grammaire et d'orthographe fréquentes :
 - "Les sport" → Le sport
 - "très in portant" → très important
 - "les poumons et le carastique" → probablement voulait dire les poumons et le cœur (*cardiaque* ?)
 - "diminue le stress." → correct, mais mal relié au reste de la phrase.
 - "enfin est un remède" → il manque un sujet ou la tournure est maladroite.
 - "confinié" → probablement voulait dire confiné
 - "fais maladies" → fait face à plusieurs maladies
- Des phrases mal construites ou incomplètes : certaines idées sont mal reliées.
- Ponctuation mal placée : une seule phrase avec un point final.

✎ Correction possible

L'activité physique

Le sport est une activité physique très importante.

Mais quels sont ses bienfaits ?

D'abord, il est excellent pour le cœur et les poumons.

Ensuite, il améliore la respiration et diminue le stress.

Enfin, c'est un remède contre l'obésité.

Faire du sport permet d'éviter plusieurs maladies.

En conclusion, il faut faire du sport.

▮ Grille d'évaluation (sur 20 points)

Critère	Note /20	Commentaire
Respect du sujet	4 / 4	Sujet bien compris
Organisation des idées	3 / 4	Bonne logique, mais quelques confusions
Richesse du vocabulaire	2 / 4	Vocabulaire simple et parfois imprécis
Correction grammaticale et orthographe	2 / 4	Nombreuses fautes à corriger
Ponctuation et syntaxe	2 / 4	Phrases parfois mal construites, ponctuation à revoir
Total	13 / 20	Niveau satisfaisant, mais améliorable

Souhaites-tu une fiche de remédiation ou un exercice pour travailler les fautes repérées ?

Oui

Parfait ! Voici une fiche de remédiation personnalisée pour aider Mariane à progresser sur les points faibles identifiés dans sa copie.

✂ FICHE DE REMÉDIATION – Améliorer son expression écrite

⚡ 1. Accords et grammaire

▶ Problème observé :

- *Les sport* → le nom « *sport* » est au singulier, il faut écrire le sport.
- *il est exulent* → il faut écrire il est excellent.

🔄 À s'entraîner :

Complète les phrases avec les bons articles et accords :

1. ___ sport est bon pour la santé. (*Le / Les*)
 2. Il est ___ pour le cœur. (*excellent / excellents*)
 3. ___ respiration est importante. (*La / Le*)
-

□ 2. Lexique et vocabulaire spécifique

▶ Problème observé :

- *carastique* → mot inventé ou mal orthographié. Probablement voulait dire cardiaque ou caractéristique.
- *fais maladies* → mauvaise construction. Il faut dire fait face à plusieurs maladies.

🔄 À mémoriser :

Mauvais mot	Mot correct	Sens
carastique	cardiaque	Relatif au cœur
fais maladies	fait face aux maladies	Combat les maladies

✂ 3. Ponctuation et construction des phrases

► Problème observé :

- Une seule phrase avec un point à la fin.
- Mauvaise liaison entre les idées.

↻ À réviser :

Utilise des connecteurs logiques pour structurer ton texte :

- D'abord / Ensuite / Enfin / En conclusion
- Exemple de phrase correcte :

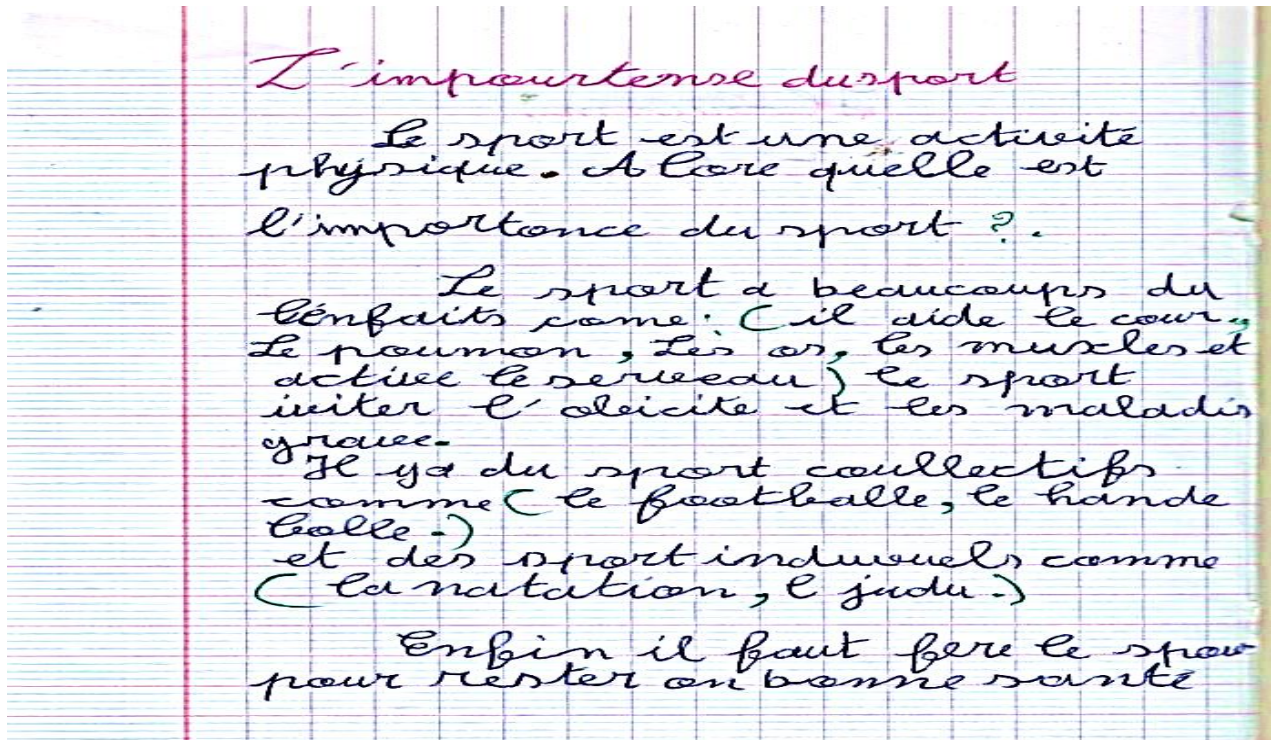
Ensuite, le sport diminue le stress et améliore la respiration.

□ 4. Exercice de réécriture

Récris ce texte en corrigeant les fautes et en améliorant les phrases :

Les sport est un acte phisique très in portant. il est exulent pour le caur les poumon. ensuite ameliore la respirasion et diminus le stresse. enfin est un remède pour les maladie.

Voici l'évaluation complète de la copie 2 intitulée « *L'importance du sport* ».



Q 1. Observation du contenu

✓ Points positifs :

- Sujet bien compris : le texte explique clairement pourquoi le sport est important.
- Les bienfaits du sport sont bien détaillés : effets sur le cœur, les poumons, les muscles, le cerveau, la santé générale.
- Présence de connecteurs logiques : *Enfin, comme.*
- L'élève distingue bien sports collectifs et sports individuels.
- L'écriture est lisible et soignée.

△ Points à améliorer :

- **Orthographe :**
 - *L'improurtence* → L'importance
 - *beaucoup du bénéfices* → beaucoup de bienfaits
 - *Les or* → les os

- *induvuel* → individuel
- *resté en bonne santé* → rester en bonne santé
- **Syntaxe :**
 - Phrases parfois trop longues, mal ponctuées.
 - Emploi répétitif de la conjonction *et* dans une seule phrase.

Proposition de correction

L'importance du sport

Le sport est une activité physique. Alors, quelle est l'importance du sport ?

Le sport a beaucoup de bienfaits, comme : il aide le cœur, les poumons, les os, les muscles et il active le cerveau. Le sport permet aussi d'éviter l'obésité et les maladies graves.

Il y a des sports collectifs, comme le football et le handball, et des sports individuels, comme la natation et le judo.

Enfin, il faut faire du sport pour rester en bonne santé.

Grille d'évaluation (sur 20 points)

Critère	Note /20	Commentaire
Respect du sujet	4 / 4	Sujet bien compris et bien traité
Organisation des idées	4 / 4	Les paragraphes sont bien structurés
Richesse du vocabulaire	3 / 4	Bon lexique, mais pourrait être un peu plus varié
Correction grammaticale et orthographe	2 / 4	Fautes d'orthographe fréquentes
Ponctuation et syntaxe	2 / 4	Phrases parfois mal ponctuées ou surchargées
Total	15 / 20	Bonne copie, avec du fond, mais à revoir au niveau de la langue

✦ Conseils de remédiation personnalisés

► Travailler l'orthographe des mots suivants :

- importance, bienfaits, individuel, rester

► Réduire les longues phrases :

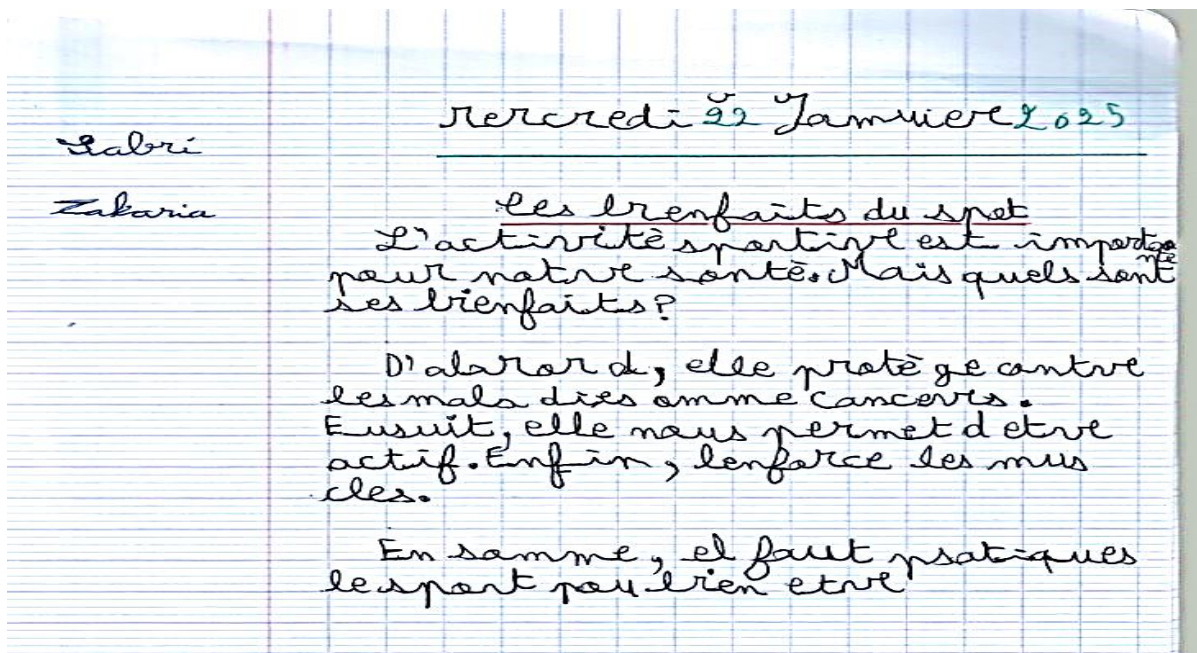
- Penser à faire des phrases plus courtes pour mieux respirer et éviter les erreurs.

► Ponctuation :

- Utiliser le point final plus souvent.
- Éviter de mettre une virgule avant « et ».

Souhaites-tu que je fasse une fiche d'exercices pour cette copie aussi ?

Merci, voici l'évaluation complète de la copie 3 intitulée « Les bienfaits du sport » écrite par Gabri et Zakaria.



🔍 1. Observation du contenu

✓ Points positifs :

- Sujet bien compris.
 - Structure claire avec une introduction, un développement et une conclusion.
 - Utilisation de connecteurs logiques : *D'abord, Ensuite, Enfin, En somme.*
 - Effort d'écriture propre et soignée.
-

⚠ Points à améliorer :

✍ Orthographe et grammaire :

- *L'activité sportife* → L'activité sportive
- *elle protè ge* → elle protège (*erreur d'espacement*)
- *les mals dites* → les maladies
- *permèt* → permet
- *lenforte* → renforce
- *les muscles* → les muscles
- *pratiques* → pratiquer
- *pour bien être* → pour le bien-être

☐ Syntaxe et reformulation :

- Les phrases sont simples mais parfois maladroitement ou trop courtes.
 - Il manque un peu de vocabulaire varié.
-

📝 Proposition de correction

Les bienfaits du sport

L'activité sportive est importante pour notre santé. Mais quels sont ses bienfaits ?

D'abord, elle protège contre les maladies comme le cancer.

Ensuite, elle nous permet d'être actifs.

Enfin, elle renforce les muscles.

En somme, il faut pratiquer le sport pour le bien-être.

▣ Grille d'évaluation (sur 20 points)

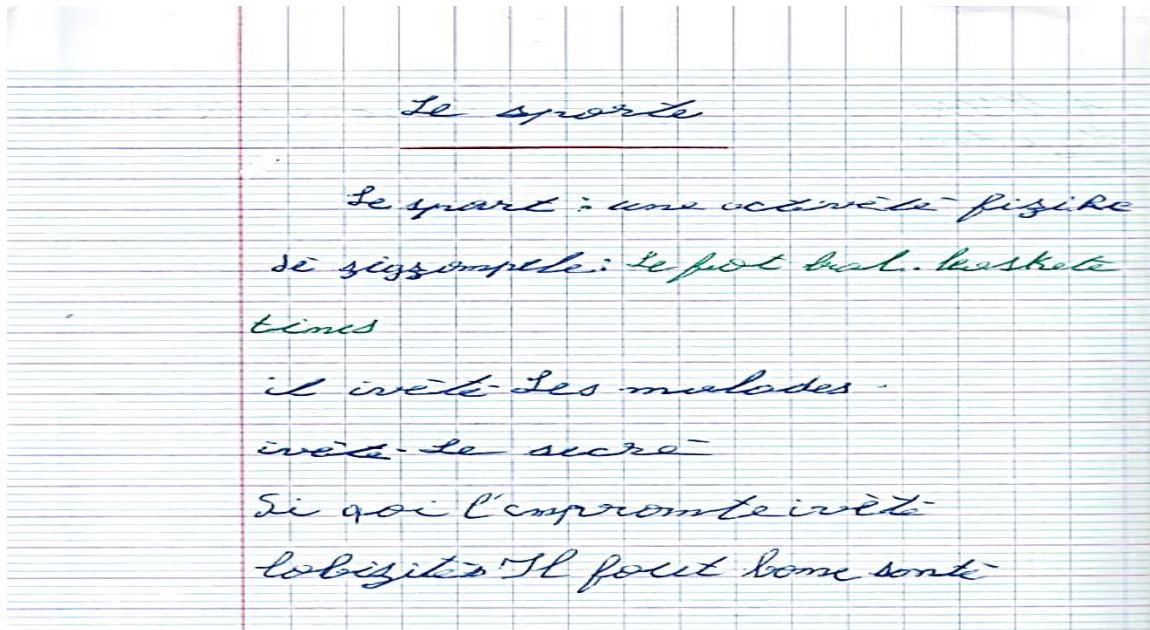
Critère	Note /20	Commentaire
Respect du sujet	4 / 4	Sujet bien traité
Organisation des idées	3 / 4	Structure claire, mais développement un peu court
Richesse du vocabulaire	2 / 4	Vocabulaire très simple
Correction grammaticale et orthographe	2 / 4	Plusieurs fautes importantes
Ponctuation et syntaxe	2 / 4	Syntaxe correcte mais sans complexité
Total	13 / 20	Bon début, mais attention à l'orthographe et à la richesse des phrases

✦ Conseils de remédiation personnalisés

➤ À revoir :

- Les accords dans les verbes et l'orthographe de mots courants comme : *sportive*, *maladies*, *muscles*.
- Faire attention aux espacements : « *protège* » ne doit pas être coupé.
- Travailler l'expression orale et écrite pour mieux enchaîner les idées.

Voici l'évaluation complète de la copie 4 intitulée « Le sport ».



Q 1. Observation du contenu

✓ Points positifs :

- Sujet identifié : l'importance du sport.
- Tentative de structure (*exemples + bienfaits + conclusion*).
- Présentation propre et aérée.
- Écriture appliquée.

⚠ Points à améliorer :

✍ Orthographe et grammaire :

- *fizike* → physique
- *sigsonplé* → par exemple
- *le fot bol* → le football
- *tinés* → difficile à lire, probablement tennis
- *ivété* → évite
- *melodes* → maladies

- *sucre* → sans précision, devrait être le diabète ou le sucre en trop
- *l'emprunte vité* → sans sens clair, probablement l'améliore la forme ou l'endurance
- *lobizités* → l'obésité
- *Il faut bone santé* → manque un verbe, devrait être Il faut être en bonne santé

□ **Syntaxe et formulation :**

- Phrases souvent incomplètes ou mal formulées.
- Manque de connecteurs logiques (*pas de « d'abord », « ensuite », etc.*).
- Très peu de phrases complètes.

✎ **Proposition de correction (version améliorée)**

Le sport

Le sport est une activité physique.

Par exemple : le football, le basket, le tennis.

Il évite les maladies comme le diabète.

Il lutte aussi contre l'obésité.

C'est pourquoi le sport est important pour être en bonne santé.

▣ **Grille d'évaluation (sur 20 points)**

Critère	Note /20	Commentaire
Respect du sujet	3 / 4	Sujet respecté
Organisation des idées	2 / 4	Idées présentes mais mal enchaînées
Richesse du vocabulaire	1 / 4	Vocabulaire pauvre et répétitif
Correction grammaticale et orthographe	1 / 4	Nombreuses fautes importantes
Ponctuation et syntaxe	1 / 4	Très peu de phrases complètes
Total	8 / 20	Travail insuffisant mais bonne volonté visible

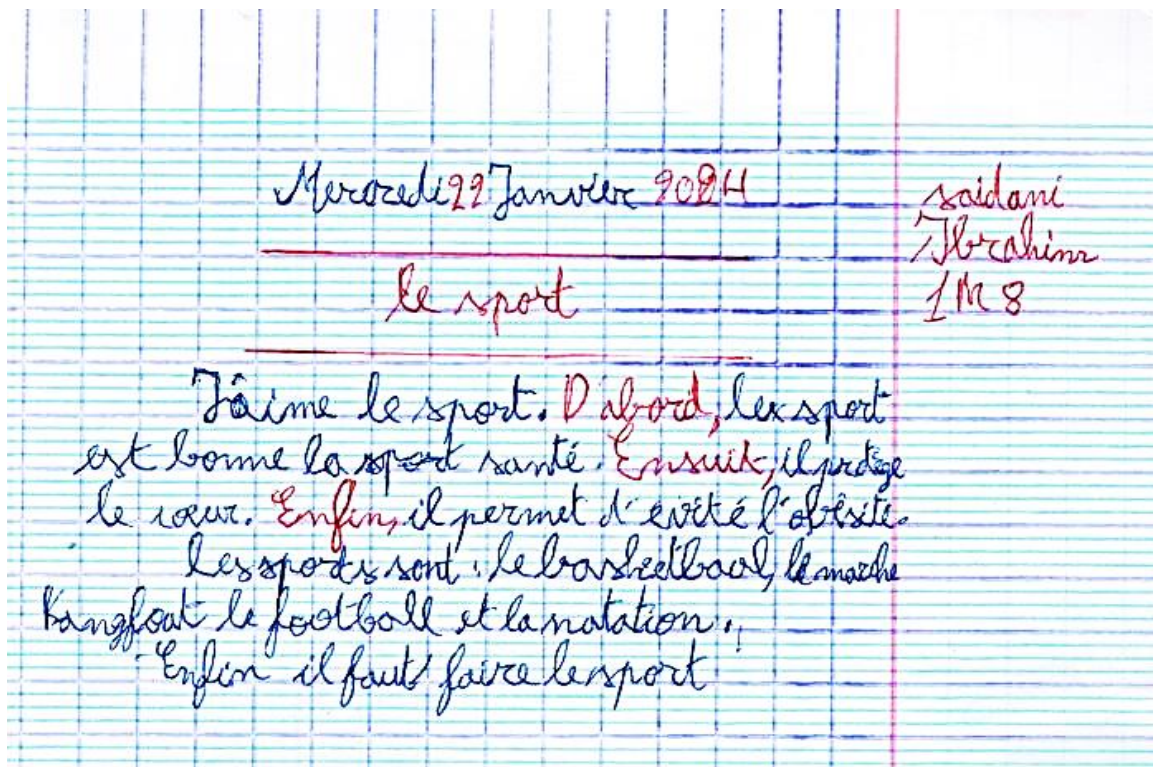
★ Conseils de remédiation personnalisés

À travailler en priorité :

- L'orthographe de mots fréquents (*sport, physique, maladies...*).
- La construction de phrases simples sujet + verbe + complément.
- L'usage de connecteurs : *D'abord, Ensuite, Enfin...*

Souhaites-tu que je prépare une fiche d'exercices ciblée sur les erreurs de cette copie ?

Merci pour cette nouvelle copie 5. Voici son évaluation complète.



☞ Texte recopié :

J'aime le sport.

D'abord, le sport est bon pour la santé.

Ensuite, il protège le cœur.

Enfin, il permet d'éviter l'obésité.

Les sports sont : le basketball, la marche, le handball, le football et la natation.

Enfin, il faut faire le sport.

✓ **Points positifs :**

- Sujet bien compris : les bienfaits du sport.
 - Structure claire avec les connecteurs D'abord / Ensuite / Enfin.
 - Présence d'une introduction, d'exemples, d'une conclusion.
 - Écriture soignée, présentation propre.
-

△ **Axes de progrès :**

✍ **Langue (orthographe/grammaire) :**

- « *faire le sport* » → on dit plutôt faire du sport.
- Une répétition inutile de "Enfin" dans la conclusion (*déjà utilisé juste avant*).
- Quelques noms de sports pourraient être mieux présentés dans une phrase plus naturelle : "*Parmi les sports, on peut citer...*"

□ **Style et fluidité :**

- Phrase très courte en introduction : « *J'aime le sport* » → Cela pourrait être intégré dans une phrase plus riche.
 - Répétition du mot « *sport* » dans presque chaque phrase.
-

□ **Proposition de version améliorée :**

Le sport est une activité physique que j'aime beaucoup.

D'abord, il est bon pour la santé.

Ensuite, il protège le cœur et permet d'éviter certaines maladies comme l'obésité.

Parmi les sports, on peut pratiquer le football, le basketball, la natation, la marche ou

encore le handball.

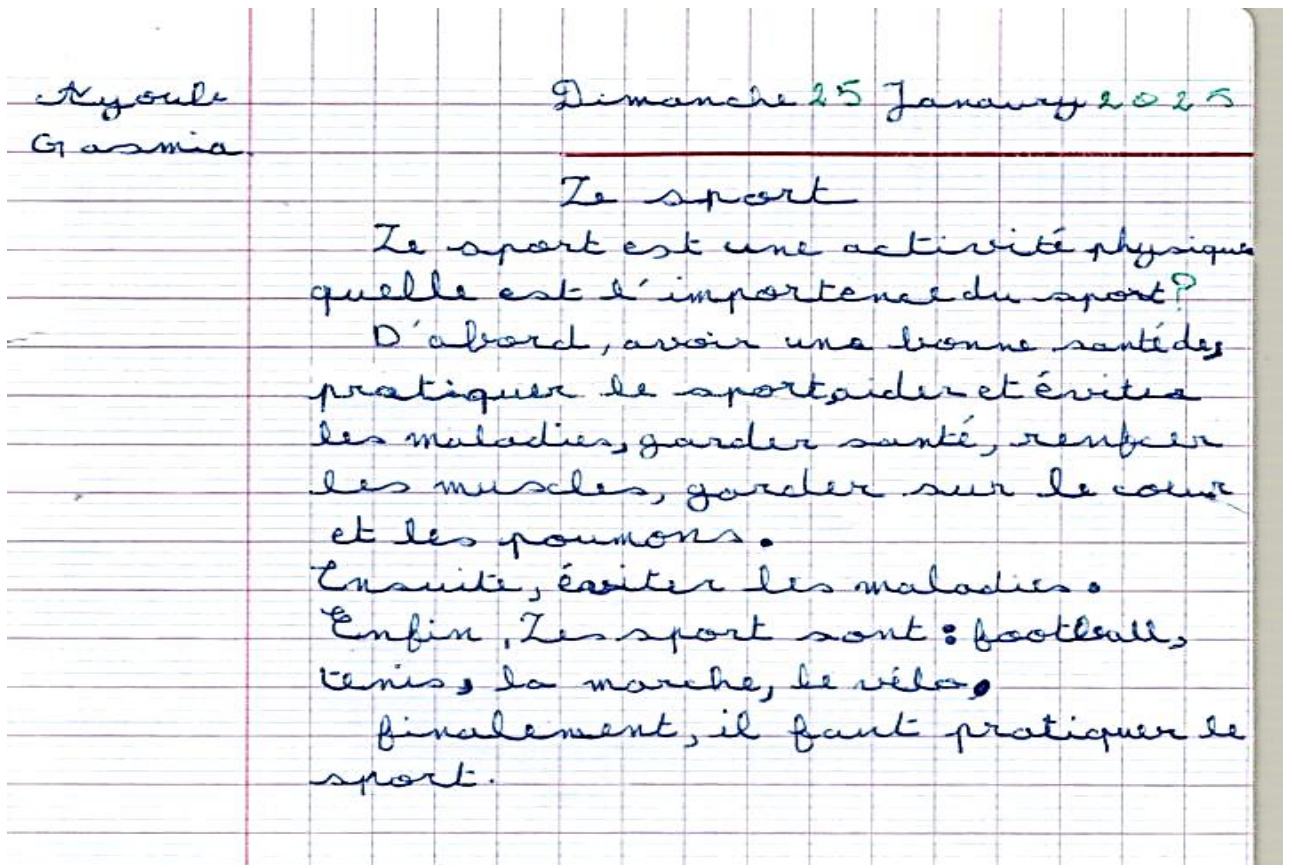
En conclusion, il est important de faire du sport pour rester en bonne santé.

▣ Grille d'évaluation (sur 20 points)

Critère	Note /20	Commentaire
Respect du sujet	4 / 4	Sujet très bien traité
Organisation des idées	3.5 / 4	Bon usage des connecteurs
Richesse du vocabulaire	2.5 / 4	Lexique simple, mais pertinent
Correction grammaticale et orthographe	3 / 4	Quelques erreurs mineures
Ponctuation et syntaxe	3 / 4	Phrases correctes mais parfois répétitives
Total	16 / 20	Très bon travail, à approfondir légèrement

Souhaites-tu que je prépare une fiche de révision ciblée sur les connecteurs logiques et les noms de sports ?

Merci ! Voici l'évaluation complète de cette copie 6 manuscrite.



📄 Texte recopié :

Le sport

Le sport est une activité physique.

Quelle est l'importance du sport ?

D'abord, avoir une bonne santé, pratiquer du sport aide à éviter les maladies, garder santé, renforcer les muscles, garder sur le cœur et les poumons.

Ensuite, éviter les maladies.

Enfin, les sports sont : football, tennis, la marche, le vélo.

Finalement, il faut pratiquer le sport.

✓ Points positifs :

- Sujet bien introduit par une question rhétorique (*bonne idée !*).
- Utilisation de connecteurs logiques : D'abord / Ensuite / Enfin / Finalement.

- Vocabulaire varié : cœur, muscles, poumons, noms de plusieurs sports.
 - Présentation soignée, écriture claire et bien aérée.
-

△ Axes de progrès :

Langue (*grammaire et formulation*) :

- « *pratiquer du sport* » → la forme correcte est pratiquer le sport ou faire du sport.
- Répétition de l'idée « *éviter les maladies* ».
- Formulation parfois maladroite :
 - « *garder santé* » → dire garder la santé.
 - « *garder sur le cœur et les poumons* » → dire protéger le cœur et les poumons.
 - La liste des sports pourrait être mieux intégrée dans une phrase.

□ Organisation :

- Redondances entre les phrases.
 - La phrase centrale est un peu longue et mériterait d'être découpée pour plus de clarté.
-

✦⁺ Proposition de version améliorée :

Le sport est une activité physique très importante.

D'abord, il permet de rester en bonne santé. Il aide à renforcer les muscles, le cœur et les poumons.

Ensuite, il permet d'éviter certaines maladies.

Parmi les sports, on trouve le football, le tennis, la marche ou le vélo.

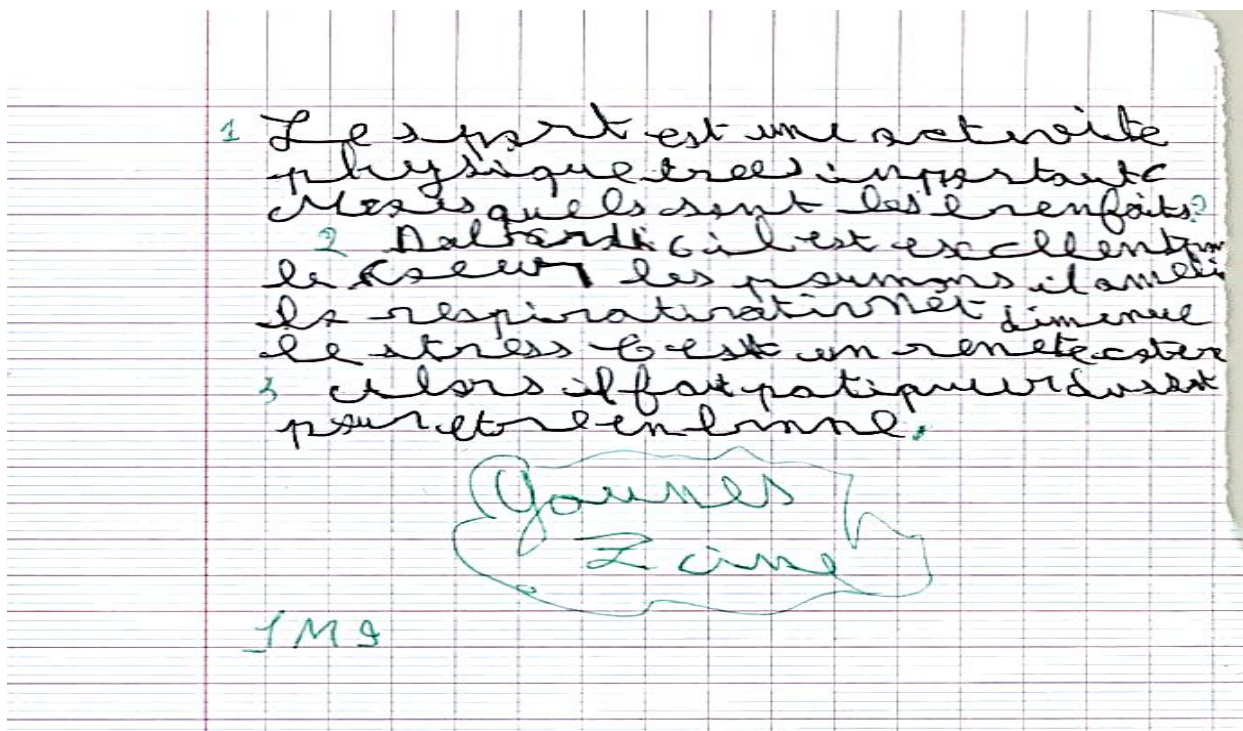
Finalement, il est essentiel de pratiquer une activité sportive régulièrement.

III Grille d'évaluation (sur 20 points)

Critère	Note /20	Commentaire
Respect du sujet	4 / 4	Sujet bien respecté
Organisation des idées	3 / 4	Bonne logique, mais des redondances
Richesse du vocabulaire	3 / 4	Lexique varié (organes, sports)
Correction grammaticale et orthographe	2.5 / 4	Quelques erreurs de formulation
Ponctuation et syntaxe	3 / 4	Lisible, mais améliorabile dans la structure
Total	15.5 / 20	Bon travail, encore quelques maladresses à corriger

Souhaites-tu un tableau comparatif des copies déjà corrigées ?

Merci pour la nouvelle copie 7 ! Voici l'analyse complète de cette production écrite.



📌 Texte recopié :

Le sport est une activité physique très importante.

Quels sont les bienfaits ?

D'abord, le sport est excellent pour le cœur, la tête, et surtout pour la concentration. Il calme le stress. C'est un remède naturel contre la fatigue morale.

Pour être en bonne santé, il faut pratiquer du sport.

✓ Points positifs :

- Bonne structure argumentative en réponse à une question d'introduction.
- Utilisation de connecteurs logiques : D'abord / Pour être en bonne santé.
- Vocabulaire précis et pertinent : cœur, concentration, stress, remède, fatigue morale.
- Le style est fluide et naturel.
- Une formulation personnelle qui dénote une bonne compréhension du sujet.

△ Axes de progrès :

▬ Langue :

- « *Quels sont les bienfaits ?* » : bonne question, mais elle gagnerait à être un peu reformulée dans une tournure complète, par exemple :
 ➤ *Quels sont les bienfaits du sport ?*
- La dernière phrase pourrait être un peu développée pour une meilleure conclusion.

□ Organisation :

- Bonne logique d'ensemble, mais le texte est court : il manque peut-être un exemple concret ou une liste de sports pour enrichir davantage.

✦ Proposition de version améliorée :

Le sport est une activité physique très importante pour la santé.

Quels sont les bienfaits du sport ?

D'abord, il est excellent pour le cœur, la concentration et le moral. Il aide à réduire le stress et la fatigue.

Par exemple, la marche, le vélo ou la natation sont des sports bénéfiques.

En conclusion, pour rester en bonne santé, il est essentiel de pratiquer une activité sportive régulièrement.

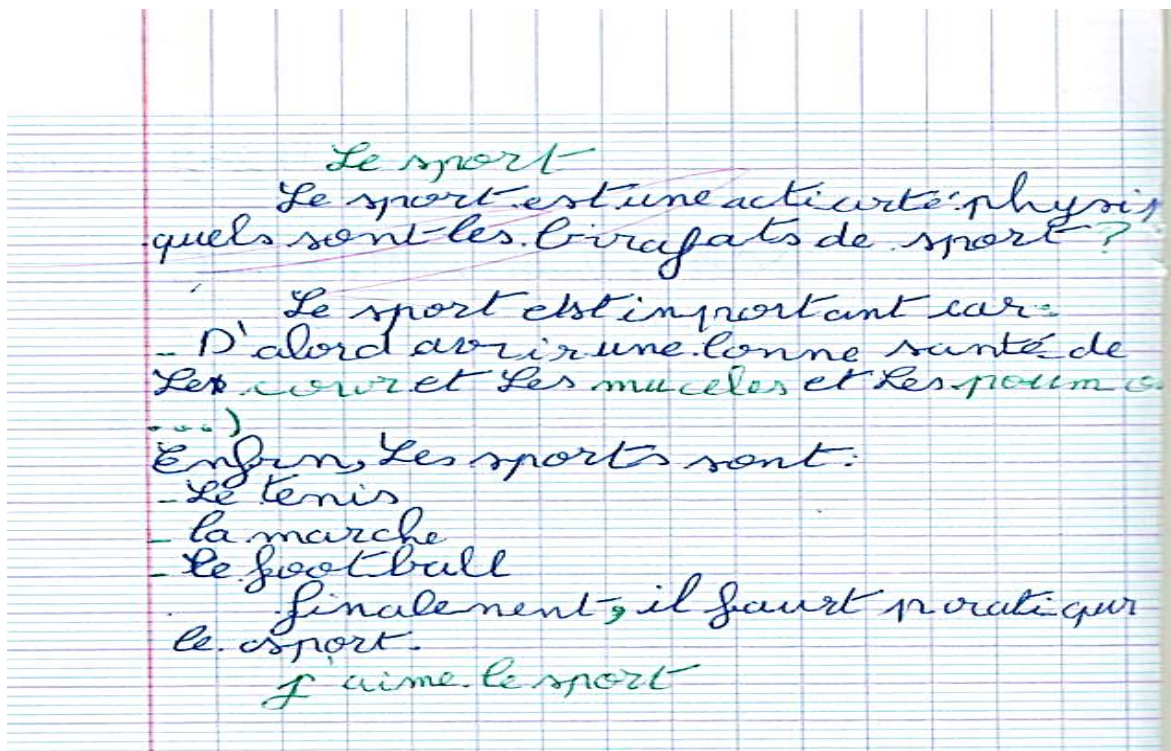
▮ Grille d'évaluation (sur 20 points)

Critère	Note /20	Commentaire
Respect du sujet	4 / 4	Sujet bien traité
Organisation des idées	3.5 / 4	Cohérence bonne, mais un peu court
Richesse du vocabulaire	3.5 / 4	Vocabulaire intéressant et précis
Correction grammaticale et orthographe	3.5 / 4	Très peu d'erreurs
Ponctuation et syntaxe	3.5 / 4	Fluidité correcte, ponctuation juste

Total	18 / 20	Une copie très solide et bien rédigée
-------	---------	---------------------------------------

Souhaites-tu que je prépare un tableau comparatif des 4 dernières copies corrigées ?

Merci ! Voici l'analyse complète de cette nouvelle copie 8 manuscrite.



🐾 Texte recopié :

Le sport

Le sport est une activité physique.

Quels sont les bienfaits du sport ?

Le sport est important car :

D'abord, avoir une bonne santé de cœur, les muscles et les poumons.

Enfin, les sports sont :

- le tennis
- la marche

- le football
Finalement, il faut pratiquer le sport.
J'aime le sport.
-

✓ **Points positifs :**

- Bonne volonté de structuration du texte avec une introduction, un développement et une conclusion.
 - L'élève suit les consignes du sujet.
 - Les exemples sont bien listés.
 - La phrase finale "*J'aime le sport.*" est personnelle et engageante.
-

△ **Axes de progrès :**

☛ **Langue et syntaxe :**

- Phrase incomplète ou maladroite :

D'abord, avoir une bonne santé de cœur, les muscles et les poumons.

☛ Suggestion : *Le sport permet d'avoir une bonne santé, notamment pour le cœur, les muscles et les poumons.*

- Petite erreur dans l'emploi de « *Finalement* » : cela fonctionne, mais « *En conclusion* » serait plus clair ici.

□ **Organisation et développement :**

- Le développement reste très court : il manque un argument ou un exemple expliqué.
 - La liste des sports aurait pu être intégrée dans une phrase complète.
-

✦ **Proposition de version améliorée :**

Le sport est une activité physique bénéfique pour la santé.

Quels sont les bienfaits du sport ?

D'abord, il aide à garder un cœur en bonne santé, à renforcer les muscles et à bien faire fonctionner les poumons.

Ensuite, il existe plusieurs sports comme le tennis, la marche ou le football.

En conclusion, il est important de pratiquer une activité physique régulièrement.

Moi, j'aime le sport !

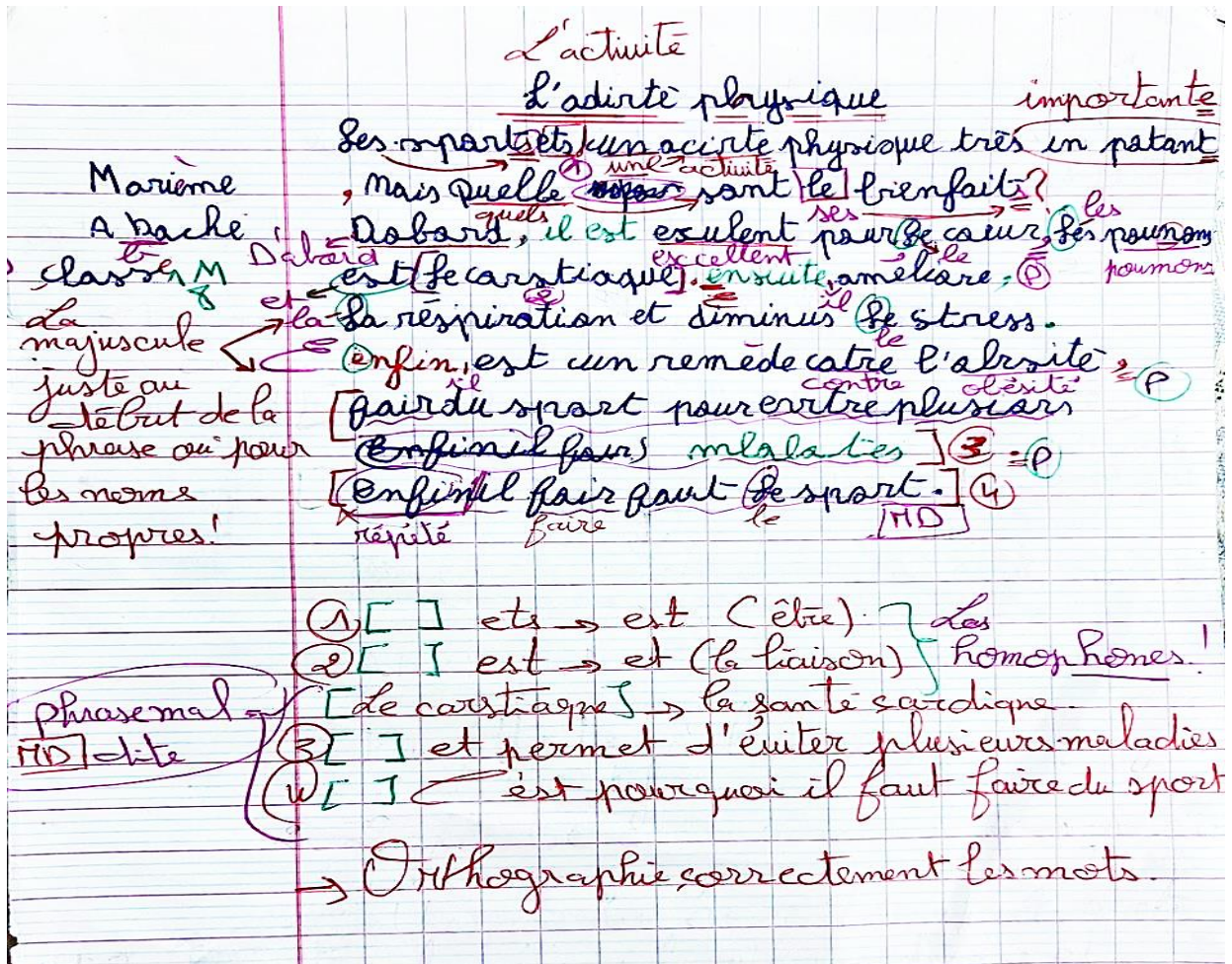
▣ **Grille d'évaluation** (sur 20 points)

Critère	Note /20	Commentaire
Respect du sujet	4 / 4	Sujet compris et traité
Organisation des idées	3 / 4	Bonne structure mais développement un peu faible
Richesse du vocabulaire	2.5 / 4	Simple mais clair
Correction grammaticale et orthographe	3 / 4	Quelques formulations à corriger
Ponctuation et syntaxe	3 / 4	Syntaxe correcte, un peu à améliorer
Total	15.5 / 20	Bon effort, à enrichir pour gagner en précision et fluidité

Souhaites-tu que je prépare maintenant le tableau comparatif des copies 8 à 11 pour voir l'évolution ou comparer les forces/faiblesses ?

III.2.1.2. Evaluation de l'enseignante des mêmes copies

Copie 1 :



Grille d'évaluation

critère	Points	Commentaire
Compréhension du sujet et des intentions communicatives	3/3	L'élève a compris le sujet et a tenté de fournir des informations pertinentes sur le thème du sport, et est débarrassé de hors-sujets.
Organisation du texte	2/3	L'introduction, le développement et la conclusion sont nettement structurés ; avec un enchaînement logique. Mais la conclusion est trop courte et n'est pas Claire.
Qualité de l'explication	3/4	L'ensemble des idées est clair, précis, bien développé mais elles manquent des exemples

Style personnel /créatif	0,5/1	Le texte évite la platitude, et réussit à se distinguer par une manière vivante, fluide d'expliquer.
Conjugaison au présent de l'indicatif	1/2	Le présent de l'indicatif est globalement maîtrisé.
Ponctuation	1/2	La ponctuation est limitée et parfois incorrecte.
Phrases déclaratives	1/1	L'élève a majoritairement utilisé des phrases déclaratives simples.
Orthographe	0,25/2	Plusieurs fautes d'orthographe dans l'ensemble du texte.
Syntaxe	0,5/2	Présence d'erreurs syntaxiques mais la compréhension reste globalement possible.

➤ Note: 12,25 /20

✓ Copie 2 :

L'importance du sport
 & d'importance

Le sport est une activité physique. À l'ère quelle est l'importance du sport ?
 on dit : mais l'importance

Le sport a beaucoup du bienfaits comme : (il aide le cœur, le poumon, les os, les muscles et les artères le système) Le sport évite l'obésité et les maladies graves.

Il y a du sports collectifs comme (le football, le handball, le tennis) et des sports individuels comme (la natation, le judo).

Enfin il faut faire le sport pour rester en bonne santé.

Phrases mal dites : [] : D'abord, il fortifie le cœur
 [] : le sport permet d'éviter

Ⓟ Utilise la ponctuation.

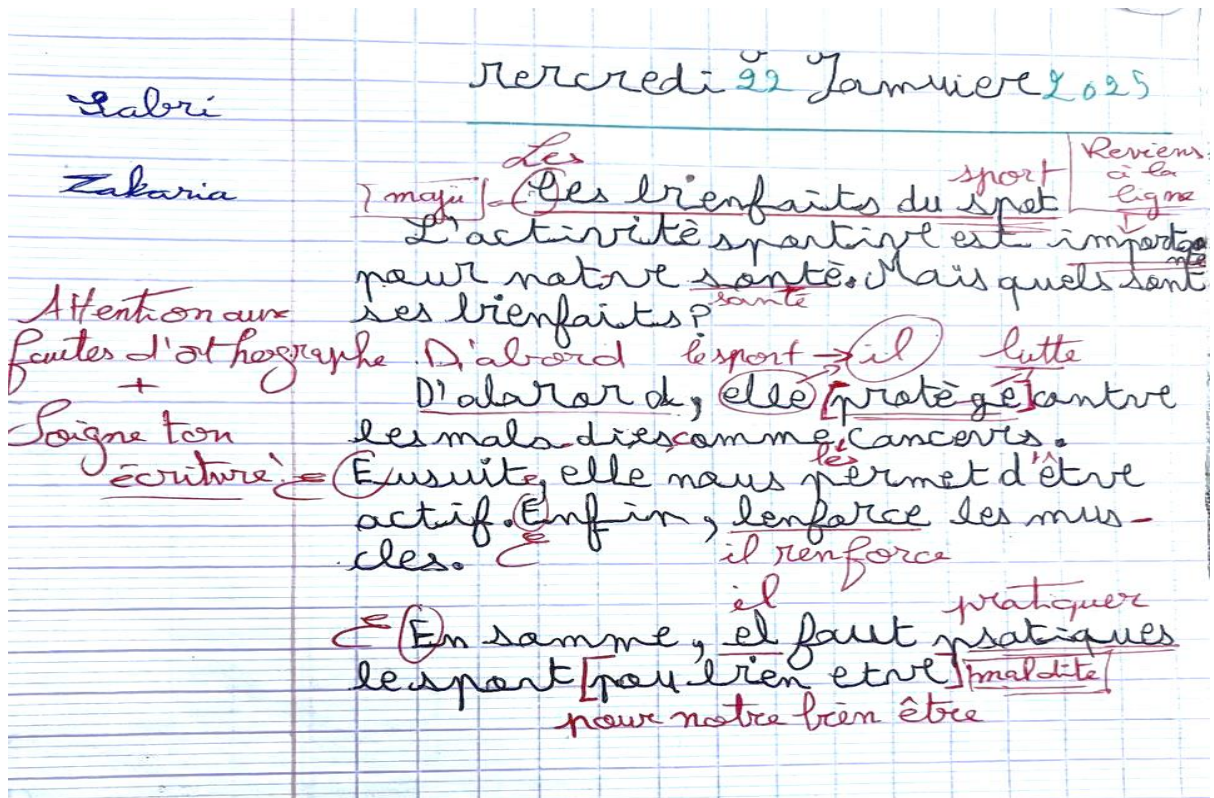
Annotations marginales :
 - mimex pluriel
 - N'oublie pas la marque de pluriel
 - Attention aux règles d'accord!

Grille d'évaluation

critère	Points	Commentaire
Compréhension du sujet et des intentions communicatives	3/3	L'élève a compris le sujet et a tenté de fournir des informations pertinentes sur le thème du sport, et est débarrassé de hors-sujets.
Organisation du texte	2/3	L'introduction, le développement et la conclusion sont nettement structurés ; avec un enchaînement logique. Mais la conclusion est trop courte et n'est pas claire et il manque des connecteurs logiques.
Qualité de l'explication	4/4	L'ensemble des idées est clair, précis, bien développé avec des exemples bien choisis.
Style personnel /créatif	1/1	Le texte évite la platitude, et réussit à se distinguer par une manière vivante, fluide d'expliquer.
Conjugaison au présent de l'indicatif	2/2	Le présent de l'indicatif est globalement maîtrisé.
Ponctuation	0,5/2	La ponctuation est limitée et incorrecte.
Phrases déclaratives	0,5/1	L'élève a utilisé des phrases déclaratives qui manquent de structuration (<i>majuscule + point</i>)
Orthographe	0,5/2	Plusieurs fautes d'orthographe dans l'ensemble du texte.
Syntaxe	1/2	Présence d'erreurs syntaxiques mais la compréhension reste globalement possible.

➤ Note: 14,5 /20

✓ Copie 3 :



Grille d'évaluation

critère	Points	Commentaire
Compréhension du sujet et des intentions communicatives	3/3	L'élève a compris le sujet et a tenté de fournir des informations pertinentes sur le thème du sport, et est débarrassé de hors-sujets.
Organisation du texte	2/3	L'introduction, le développement et la conclusion sont structurés ; avec un enchaînement logique. Mais la conclusion est trop courte et n'est pas Claire.
Qualité de l'explication	4/4	L'ensemble des idées est clair et simple.
Style personnel /créatif	1/1	Le texte évite la platitude, et réussit à se distinguer par une manière fluide d'expliquer.
Conjugaison au présent de l'indicatif	2/2	Le présent de l'indicatif est globalement maîtrisé.
Ponctuation	2/2	La ponctuation est limitée et incorrecte.

Phrases déclaratives	1/1	L'élève a utilisé des phrases déclaratives qui manquent de structuration (majuscule + point)
Orthographe	1/2	Plusieurs fautes d'orthographe dans l'ensemble du texte.
Syntaxe	2/2	Présence d'erreurs syntaxiques mais la compréhension reste globalement possible.

➤ Note: 18/20

✓ Copie 4 :

Le sport

Le sport : une activité physique :
 activité physique P

de signifie : le foot bal, basket
 par exemple football P le basketbal

+ (et) tennis P
 le tennis P

la phrase commence par une majuscule → il invite Les malades ;
 Il invite les maladies P

[invite - le sucre] → comme le diabète
 phrase mal dite

mal-dite [Si qui l'emprunte invite
 c'est qu'on mal dite invite
 l'obésité Il fait] bonne sont P
 l'obésité pour avoir une bonne P

Attention !!

- Orthographe correctement les mots
- Soigne ton écriture
- Rédige des phrases complètes et cohérentes !
- Utilise les connecteurs d'énumération

Grille d'évaluation

critère	Points	Commentaire
Compréhension du sujet et des intentions communicatives	3/3	L'élève a compris le sujet et a tenté de fournir des informations pertinentes sur le thème du sport, et est débarrassé de hors-sujets.
Organisation du texte	0,5/3	Le texte manque d'organisation d'un texte explicatif (<i>un seul paragraphe</i>)
Qualité de l'explication	2/4	Les idées ne sont pas claires et manquent d'explication et d'organisation, mais il a donné des idées quand même.
Style personnel /créatif	1/1	Le texte évite la platitude, et réussit à se distinguer par des idées uniques.
Conjugaison au présent de l'indicatif	0,5/2	La conjugaison n'est pas maîtrisée totalement
Ponctuation	0,5/2	La ponctuation n'est pas utilisée totalement ; que quelques deux points.
Phrases déclaratives	0,5/1	L'élève a utilisé des phrases déclaratives qui manquent de structuration (majuscule + point)
Orthographe	0,5/2	Pleine de fautes d'orthographe dans l'ensemble du texte, majoritairement à cause de l'écriture phonétique déformée, ou plus précisément l'orthographe phonétique.
Syntaxe	0,5/2	Présence d'erreurs syntaxiques et la compréhension est difficile.

➤ Note: 9/20

✓ Copie 5 :

Le nom propre commence par la majuscule

Saïdani
Saïdani
Abraham
1M8

Mercredi 22 Janvier 2024

Le sport

sport → masculin
bon

① [] (J'aime le sport.) D'abord, le sport est bon [] la santé. Ensuite, il protège le cœur. Enfin, il permet d'éviter l'obésité.

Les sports sont : le basket ball, le football, le badminton, le tennis, le football et la natation.

Enfin, il faut faire le sport.

connecteur répété

Après les propositions (de) les verbes se mettent à l'infinitif d'entrer

Attention!!

- des connecteurs d'énumération sont mal-placés dans l'ensemble du texte
- [] ① Introduction trop courte + absence de questionnement
- [] @ est bon pour la santé → le sport → mal dite
- Soigne ton écriture et sépare entre les paragraphes.

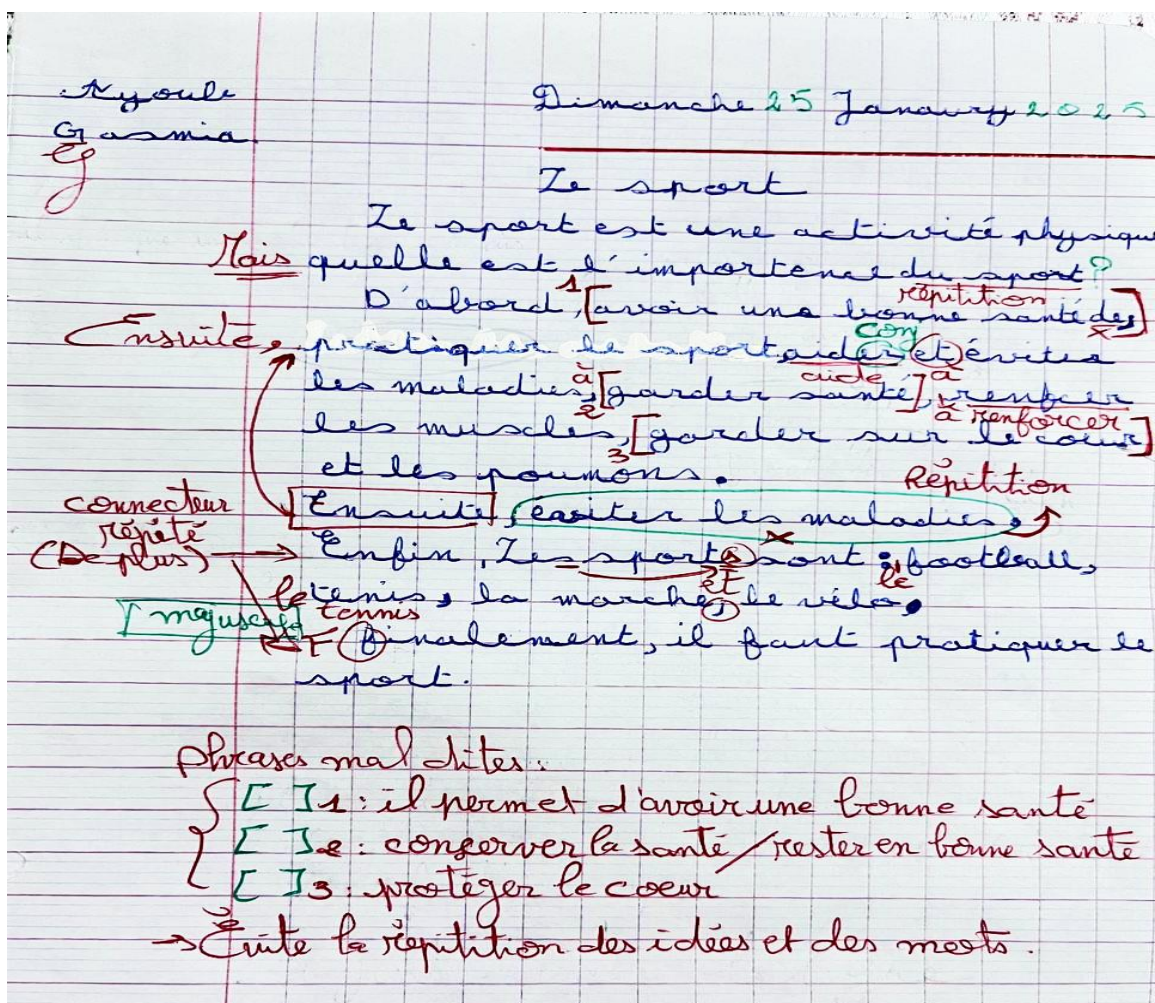
Grille d'évaluation

critère	Points	Commentaire
Compréhension du sujet et des intentions communicatives	3/3	L'élève a compris le sujet et a tenté de fournir des informations pertinentes sur le thème du sport, et est débarrassé de hors-sujets.
Organisation du texte	2,5/3	L'introduction, le développement et la conclusion sont structurés ; avec un enchaînement logique, mais l'introduction est trop courte avec un manque de questionnement.
Qualité de l'explication	4/4	L'ensemble des idées est clair et simple.

Style personnel /créatif	1/1	Le texte évite la platitude, et réussit à se distinguer par une manière fluide d'expliquer.
Conjugaison au présent de l'indicatif	2/2	Le présent de l'indicatif est globalement maîtrisé.
Ponctuation	2/2	La ponctuation est juste et bien placée
Phrases déclaratives	1/1	L'élève a utilisé des phrases déclaratives justes (<i>majuscule + point</i>)
Orthographe	1,5/2	Un nombre limité de fautes d'orthographe dans l'ensemble du texte.
Syntaxe	1,5/2	Il y a une seule mauvaise construction grammaticale et deux phrases mal dites dans l'ensemble du texte.

➤ Note: 18,5/20

✓ Copie 6 :



Grille d'évaluation

critère	Points	Commentaire
Compréhension du sujet et des intentions communicatives	3/3	L'élève a compris le sujet et a tenté de fournir des informations pertinentes sur le thème du sport, et est débarrassé de hors-sujets.
Organisation du texte	2/3	L'introduction, le développement et la conclusion sont structurés ; avec un enchaînement logique. le développement est un peu mal organisé avec une répétition des idées
Qualité de l'explication	4/4	L'ensemble des idées est clair et simple.
Style personnel /créatif	0,5/1	Le texte évite la platitude, il suit la même structure proposée avec des idées uniques.
Conjugaison au présent de l'indicatif	2/2	Le présent de l'indicatif est globalement maîtrisé.
Ponctuation	2/2	La ponctuation est juste et bien placée.
Phrases déclaratives	1/1	L'élève a utilisé des phrases déclaratives justes.
Orthographe	1,5/2	Quelques fautes d'orthographe dans l'ensemble du texte.
Syntaxe	1,5/2	La construction syntaxique est correcte globalement avec plusieurs phrases mal dites.

➤ Note: 17,5/20

✓ Copie 7 :

1 Le sport est une activité physique très importante.
 Mais quels sont les bienfaits?

D'abord → 2. D'abord, il est excellent pour le coeur, les poumons et la respiration.
 Il diminue le stress. C'est un remède contre les maladies.

[?] 3. alors il faut pratiquer regularment pour être en bonne santé. [?] [?]

idée
 inachevée
 ↓
 contre les
 maladies
 par exemple.

Jour
Zine

[MS]
 ① [] pratiquer du sport

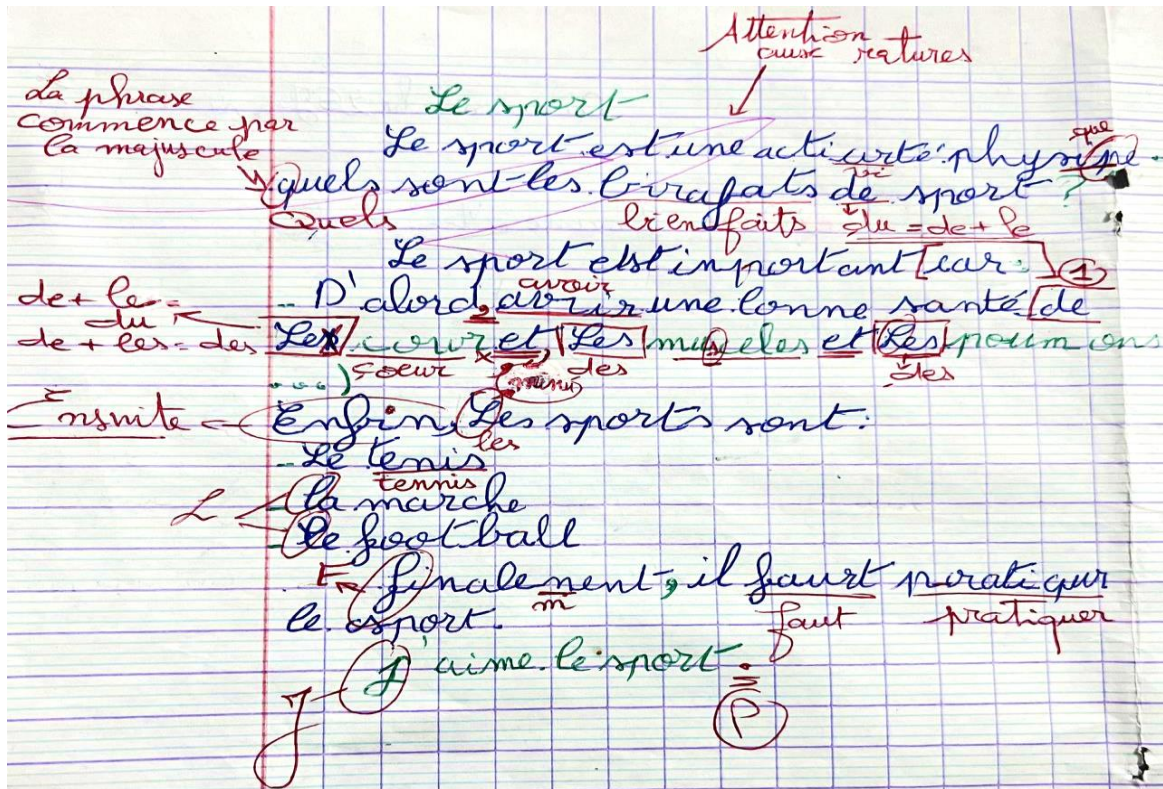
Attention }
 → Soigne ton écriture !!
 → Utilise correctement les connecteurs d'énumération.
 → Du est le titre
 → Complète les phrases.

Grille d'évaluation

critère	Points	Commentaire
Compréhension du sujet et des intentions communicatives	3/3	L'élève a compris le sujet et a tenté de fournir des informations pertinentes sur le thème du sport, et est débarrassé de hors-sujets.
Organisation du texte	1/3	Le texte ne suit pas une structure organisée, une petite introduction, un développement inachevé et sans conclusion.
Qualité de l'explication	2/4	L'ensemble des idées est simple et inachevé, il manque des exemples et d'explication.
Style personnel /créatif	0,5/1	Le texte suit une structure répétée avec une production unique.
Conjugaison au présent de l'indicatif	2/2	Le présent de l'indicatif est globalement maîtrisé.
Ponctuation	0,5/2	La ponctuation est rarement utilisée, un texte avec trois phrases
Phrases déclaratives	0,5/1	L'élève a utilisé des phrases mais sans ponctuation, qui qualifie la phrase comme déclarative
Orthographe	1/2	Un nombre limité des fautes d'orthographe avec une écriture illisible.
Syntaxe	1/2	La construction syntaxique est correcte avec quelques constructions maladroitement et des phrases incomplètes.

➤ Note: 11,5/20

✓ Copie 8 :



Grille d'évaluation

critère	Points	Commentaire
Compréhension du sujet et des intentions communicatives	3/3	L'élève a compris le sujet et a tenté de fournir des informations pertinentes sur le thème du sport, et est débarrassé de hors-sujets.
Organisation du texte	2,5/3	Le texte suit une structure organisée, avec une mauvaise utilisation des connecteurs logiques.
Qualité de l'explication	3/4	L'ensemble des idées est clair, organisé et simple.
Style personnel /créatif	1/1	Un style et une structure unique, ajoutant ces émotions sur le thème
Conjugaison au présent de l'indicatif	2/2	Le présent de l'indicatif est globalement maîtrisé.
Ponctuation	1,5/2	La ponctuation est rarement utilisée, un texte avec trois phrases
Phrases déclaratives	1/1	Elles sont globalement maîtrisées

Orthographe	0,5/2	Peu de fautes d'orthographe et d'utilisation du majuscule/minuscule
Syntaxe	1,5/2	La construction syntaxique est correcte avec une mauvaise utilisation des déterminants partitifs.

➤ Note: 15,5/20

III.2.2. Analyse Comparative Détaillée : Evaluation Humaine/Evaluation ChatGPT

Ce document présente une analyse comparative détaillée entre les évaluations fournies par un enseignant (*transcription textuelle*) et celles générées par ChatGPT pour huit productions écrites d'élèves sur le thème du sport.

L'analyse se concentre sur les points suivants pour chaque copie :

Identification de la copie (*Numéro/Nom élève, Score Enseignant, Score ChatGPT*)

Appréciation générale et notation (*Convergence/Divergence*)

Critères d'évaluation (*Similitudes/Différences*)

Points forts identifiés (*Convergence/Divergence*)

Erreurs et points faibles identifiés (*Convergence/Divergence, Niveau de détail*)

Feedback et propositions de correction/remédiation (*Nature, Pertinence, Personnalisation*)

✓ Copie 1 (*Marieme Abache*)

Identification : Texte "*L'activité physique*" - Enseignant : 12.25/20 - ChatGPT : 13/20

Appréciation générale et notation : Convergence notable. Les deux évaluateurs donnent une note très similaire (12.25 vs 13) et considèrent la copie comme satisfaisante mais améliorable. L'enseignant note 12.25, ChatGPT arrondit à 13 mais qualifie le niveau de "*satisfaisant, mais améliorable*".

Critères d'évaluation : Différences partielles. L'enseignant utilise une grille plus détaillée (9 critères incluant *style, conjugaison, types de phrases*) tandis que ChatGPT utilise une grille plus standard (5 critères : *sujet, organisation, vocabulaire, grammaire/orthographe, ponctuation/syntaxe*). Les deux couvrent la compréhension du sujet, l'organisation, l'orthographe/grammaire et la syntaxe/ponctuation.

Points forts identifiés : Convergence. Les deux relèvent la bonne compréhension du sujet et la présence d'une structure logique (*connecteurs logiques*).

Erreurs et points faibles identifiés : Forte convergence. Les deux identifient des fautes d'orthographe et de grammaire fréquentes ("*Les sport*", "*in potant*"/"*important*", "*exulent*"/"*excellent*", "*coeur*", "*pounons*"/"*poumons*", "*corstiaque*"/"*carastique*", "*réspiration*", "*diminus*"/"*diminue*", "*remède*", "*absité*"/"*obésité*", "*fairdu*"/"*faire du*", "*mlalates*"/"*maladies*"). Ils notent aussi des phrases mal construites ou incomplètes et des problèmes de ponctuation.

Feedback et propositions de correction/remédiation : Approches différentes mais complémentaires. L'enseignant propose la correction juste en dessous des mots fautifs. Il ajoute des remarques, des commentaires, et reformule parfois des phrases entières. Cette approche crée un schéma mental chez l'élève, l'aide à identifier la nature de ses erreurs et à mieux les corriger. Pour l'enseignant, cela permet une lecture plus analytique et ciblée des productions écrites. ChatGPT propose une correction rédigée complète, une fiche de remédiation détaillée avec exercices ciblés sur les erreurs (*accords, lexique, ponctuation, connecteurs*) et un exercice de réécriture.

✓ Copie 2 (*L'impourtanse du sport*)

Identification : Texte "*L'importance du sport*" - Enseignant : 14.5/20 - ChatGPT : 15/20

Appréciation générale et notation : Convergence. Notes très proches (14.5 vs 15). Les deux considèrent la copie comme bonne sur le fond mais nécessitant des améliorations sur la forme. ChatGPT : "*Bonne copie, avec du fond, mais à revoir au niveau de la langue*".

Critères d'évaluation : Différences partielles similaires à la Copie 1 (*grille enseignant plus détaillée*). Les deux évaluent le sujet, l'organisation, le vocabulaire (*enseignant : 4/4, ChatGPT : 3/4*), l'orthographe/grammaire, la ponctuation/syntaxe.

Points forts identifiés : Forte convergence. Les deux soulignent la bonne compréhension du sujet, la bonne organisation des idées (*paragraphes structurés*), la distinction entre sports collectifs et individuels, et la clarté des idées. L'enseignant note aussi le style "*vivant et fluide*". ChatGPT note l'écriture "*lisible et soignée*".

Erreurs et points faibles identifiés : Forte convergence. Les deux relèvent les fautes d'orthographe ("*L'impourtense*"/"*L'improurtence*", "*beaucoup du lienfaits*"/"*beaucoup du bénéfices*", "*Les or*"/"*Les os*", "*induvuel*", "*fere*"/"*faire*", "*rester*"/"*resté*", "*coullectifs*",

"footballe", "hande bolle", "judu", "sante"). Les deux notent des problèmes de ponctuation et de syntaxe (*phrases longues, emploi de la virgule*).

Feedback et propositions de correction/remédiation : Approches différentes. L'enseignant annoté généralement à l'aide de stylos de couleurs différentes (*souvent rouge, vert, bleu ou violet*) pour distinguer les types d'interventions (*corrections, commentaires, rappels de règles, suggestions...*). Il souligne, entoure, barre ou insère des indications directement sur la copie de l'élève. Il peut aussi utiliser des flèches, crochets, ou encadrés pour attirer l'attention sur une erreur précise ou proposer une reformulation. ChatGPT propose une correction rédigée, identifie des mots spécifiques à travailler en orthographe ("*importance*", "*bienfaits*", "*individuel*", "*rester*"), et donne des conseils précis sur la longueur des phrases et la ponctuation (*point final, virgule avant "et"*).

✓ Copie 3 (Sabri Zakaria)

Identification : Texte "*Les bienfaits du sport*" - Enseignant : 18/20 - ChatGPT : 13/20

Appréciation générale et notation : Divergence majeure. L'enseignant attribue une excellente note (18/20) tandis que ChatGPT donne une note moyenne (13/20). L'appréciation globale diffère : l'enseignant semble très satisfait, ChatGPT voit un "bon début" mais pointe des faiblesses importantes.

Critères d'évaluation : Différences notables dans l'application. Bien que les grilles diffèrent (*enseignant plus détaillée*), la divergence vient de l'évaluation des critères communs. Ex : Orthographe (*Ens : 1/2, ChatGPT : 2/4*), Syntaxe (*Ens : 2/2, ChatGPT : 2/4*), Richesse vocabulaire (*Ens : 4/4 Qualité explication, ChatGPT : 2/4 Richesse vocabulaire*). L'enseignant semble plus indulgent ou valorise différemment certains aspects.

Points forts identifiés : Convergence partielle. Les deux notent la bonne compréhension du sujet, la structure claire (*intro/dév/conclusion*) et l'utilisation de connecteurs. L'enseignant valorise la clarté et la simplicité des idées (4/4), ChatGPT trouve le vocabulaire "très simple" (2/4) et le développement "*un peu court*". L'enseignant note le style "*fluide*", ChatGPT la "*syntaxe correcte mais sans complexité*".

Erreurs et points faibles identifiés : Convergence sur les erreurs, divergence sur leur poids. Les deux identifient les fautes d'orthographe ("*sportife*"/"*sportive*", "*protè ge*", "*mals dites*"/"*maladies*", "*permèt*"/"*permet*", "*lenforte*"/"*renforce*", "*musles*"/"*muscles*", "*psatiques*"/"*pratiquer*", "*pou bien etre*"/"*pour le bien-être*"). ChatGPT note aussi des

phrases "*maladroites ou trop courtes*" et un manque de vocabulaire varié, aspects moins pénalisés par l'enseignant.

Feedback et propositions de correction/remédiation : Approches différentes. L'enseignant utilise des couleurs et des formes visuelles pour rendre son feedback plus clair, immédiat et mémorable pour l'élève. ChatGPT propose une correction rédigée, des conseils sur les accords, l'orthographe de mots courants ("*sportive*", "*maladies*", "*muscles*"), les espacements, et suggère de travailler l'enchaînement des idées.

✓ Copie 4 (*Le sporte*)

Identification : Texte "*Le sporte*" - Enseignant : 9/20 - ChatGPT : 8/20

Appréciation générale et notation : Forte convergence. Notes très basses et très proches (9 vs 8). Les deux évaluateurs jugent le travail insuffisant. ChatGPT : "*Travail insuffisant mais bonne volonté visible*".

Critères d'évaluation : Application convergente malgré des grilles différentes. Les deux attribuent des notes faibles sur l'organisation (*Ens* : 0.5/3, *ChatGPT* : 2/4), le vocabulaire (*Ens* : 2/4 *Qualité explication*, *ChatGPT* : 1/4 *Richesse vocabulaire*), l'orthographe (*Ens* : 0.5/2, *ChatGPT* : 1/4) et la syntaxe/ponctuation (*Ens* : 0.5/2 *Syntaxe* + 0.5/2 *Ponctuation*, *ChatGPT* : 1/4 *Ponctuation/Syntaxe*).

Points forts identifiés : Convergence minimale. Les deux notent que le sujet est compris/identifié. L'enseignant relève des "*idées uniques*" (*style* 1/1), ChatGPT note une "*bonne volonté visible*" et une "*écriture appliquée*".

Erreurs et points faibles identifiés : Forte convergence. Les deux identifient de très nombreuses fautes d'orthographe ("*sporte*", "*fizike*", "*zigzompele*"/"*sigsonplé*", "*fot bol*", "*baskete*", "*tines*"/"*tennis*", "*ivété*"/"*évite*", "*molades*"/"*maladies*", "*secré*"/"*sucre*", "*lobizité*"/"*lobizités*", "*fout*"/"*faut*", "*bone santé*"). Les deux notent le manque d'organisation, l'absence de connecteurs, des phrases incomplètes ou mal formulées, un vocabulaire pauvre et répétitif, et une ponctuation quasi absente.

Feedback et propositions de correction/remédiation : Approches différentes. L'enseignant utilise des couleurs et des formes visuelles pour rendre son feedback plus clair, immédiat et mémorable pour l'élève, pour identifier par exemple la majuscule, la ponctuation ou l'accord. ChatGPT propose une correction rédigée (*version très simplifiée mais correcte*), et des conseils clairs sur les priorités : orthographe des mots fréquents, construction de phrases simples (*S+V+C*), usage des connecteurs.

✓ **Copie 5 (Saidani Ibrahim)**

Identification : Texte "*J'aime le sport*" - Enseignant : 18.5/20 - ChatGPT : 16/20

Appréciation générale et notation : Divergence modérée. L'enseignant donne une note excellente (18.5), ChatGPT une très bonne note (16). Les deux apprécient la copie mais ChatGPT est légèrement plus critique sur le style. ChatGPT : "*Très bon travail, à approfondir légèrement*".

Critères d'évaluation : Application convergente sur le fond, divergente sur la notation. Les deux évaluent positivement la compréhension du sujet, l'organisation (*connecteurs*), la correction orthographique et syntaxique (*peu d'erreurs*). La différence de note vient peut-être du critère "Richesse du vocabulaire" (*Ens : 4/4 Qualité explication, ChatGPT : 2.5/4*) et "Style" (*Ens : 1/1, ChatGPT note des répétitions*).

Points forts identifiés : Forte convergence. Les deux notent la bonne compréhension du sujet, la structure claire avec connecteurs, la présence d'exemples, la correction globale de la langue et la présentation soignée.

Erreurs et points faibles identifiés : Convergence partielle. L'enseignant ne relève que très peu d'erreurs d'orthographe, d'accord, de conjugaison et d'une seule phrase mal dite. ChatGPT identifie l'usage de "*faire le sport*" (*au lieu de "faire du sport"*), la répétition de "*Enfin*", la répétition du mot "*sport*", et suggère une meilleure intégration de la liste de sports et de la phrase d'introduction.

Feedback et propositions de correction/remédiation : Approches différentes. L'enseignant utilise des remarques ; des commentaires accompagnés de petits rappels des règles, placés juste à côté des erreurs. ChatGPT propose une version améliorée du texte (*plus fluide, moins répétitive*) et suggère une fiche de révision sur les connecteurs et les noms de sports.

✓ **Copie 6 (Ayoub Gasmia)**

Identification : Texte "*Le sport*" (*avec question initiale*) - Enseignant : 17,5/20 - ChatGPT : 15.5/20

Appréciation générale et notation : Divergence modérée. L'enseignant donne une excellente note (17,5), ChatGPT une bonne note (15.5). Les deux apprécient la copie mais ChatGPT identifie plus de points de friction dans la formulation. ChatGPT : "*Lisible, mais quelques maladresses*".

Critères d'évaluation : Application globalement convergente mais notation différente. Les deux notent positivement le sujet, la conjugaison, la ponctuation, la syntaxe globale. Divergences sur l'organisation (*Ens : 2/3 avec répétition, ChatGPT : 3/4 avec redondances*), le vocabulaire (*Ens : 4/4 Qualité explication, ChatGPT : 3/4 Richesse vocabulaire*), l'orthographe (*Ens : 1.5/2, ChatGPT : 2.5/4*).

Points forts identifiés : Forte convergence. Les deux apprécient l'introduction par une question, l'utilisation de connecteurs, le vocabulaire varié (*organes, sports*), la correction globale de la langue (*conjugaison, ponctuation, syntaxe*), et la présentation soignée.

Erreurs et points faibles identifiés : Convergence sur les points, formulation différente. Les deux notent la répétition de l'idée "*éviter les maladies*". L'enseignant note des erreurs sur "*importance*", "*pratiquer le sport*", "*aider et éviter*", "*garder santé*", "*garder sur le coeur*", "*tenis*", "*finaleme*", il signale la majuscule et la minuscule, les accords, les connecteurs d'énumérations ; les ratures et les phrases mal dites. ChatGPT pointe des formulations maladroitement ("*pratiquer du sport*", "*garder santé*", "*garder sur le cœur et les poumons*"), la répétition, et la phrase centrale jugée trop longue et qui mériterait d'être découpée.

Feedback et propositions de correction/remédiation : Approches différentes. L'enseignant utilise des couleurs et des formes visuelles pour rendre son feedback plus clair, immédiat et mémorable pour l'élève. ChatGPT propose une version améliorée (*phrases plus courtes, formulations corrigées, moins de redondances*).

✓ Copie 7 (Ziane Younes)

Identification : Texte "*Le sport*" - Enseignant : 11,5/20 - ChatGPT : 11.5/20

Appréciation générale et notation : Forte convergence (*la même note*), Notes faibles. Les deux identifient des lacunes importantes notamment en organisation et orthographe.

Critères d'évaluation : Application convergente. Les deux notent positivement la compréhension du sujet et la conjugaison. Ils pénalisent fortement l'organisation (*Ens : 1/3, ChatGPT : 2/4*), la qualité de l'explication/vocabulaire (*Ens : 2/4, ChatGPT : 2/4*), l'orthographe (*Ens : 1/2, ChatGPT : 1.5/4*) et la ponctuation/syntaxe (*Ens : 0.5/2 Ponctuation + 1/2 Syntaxe, ChatGPT : 2/4*).

Points forts identifiés : Convergence minimale. Les deux notent la compréhension du sujet et la maîtrise globale du présent de l'indicatif.

Erreurs et points faibles identifiés : Forte convergence. Les deux identifient de nombreuses fautes d'orthographe ("*activite*", "*trees*", "*inportante*", "*Meois*"/"*Mais*", "*benfaits*",

"Dobordi"/"D'abord", "ameli"/"améliore", "respiratiratirn"/"respiration", "dimenue"/"diminue", "renete coter"/"remède contre", "fau"/"faut", "patiquur"/"pratiquer", "dusont"/"du sport", "bnne"/"bonne"). Les deux notent le manque d'organisation (*pas de conclusion, développement inachevé*), le manque d'exemples et d'explication, la ponctuation limitée et des constructions syntaxiques maladroites.

Feedback et propositions de correction/remédiation : Approches différentes. L'enseignant utilise des couleurs et des formes visuelles pour rendre son feedback plus clair, immédiat et mémorable pour l'élève. ChatGPT propose une correction rédigée, et des conseils ciblés sur l'orthographe, la structure du texte (*plan*), l'ajout d'exemples et la ponctuation.

✓ Copie 8 (*Le sport*)

Identification : Texte "*Le sport*" - Enseignant : 15.5/20 - ChatGPT : 14/20

Appréciation générale et notation : Forte convergence. Notes très proches (15.5 vs 14). Les deux évaluateurs considèrent la copie comme moyenne, avec des points forts mais aussi des erreurs notables.

Critères d'évaluation : Application globalement convergente. Les deux notent positivement la compréhension du sujet, l'organisation générale (*connecteurs*), la conjugaison. Ils pénalisent l'orthographe (*Ens : 0.5/2, ChatGPT : 2/4*) et la syntaxe/ponctuation (*Ens : 1,5/2 Syntaxe + 1/2 Ponctuation, ChatGPT : 2/4*).

Points forts identifiés : Convergence. Les deux notent la compréhension du sujet, la tentative de structuration avec des connecteurs et des listes.

Erreurs et points faibles identifiés : Forte convergence. Les deux identifient de nombreuses fautes d'orthographe ("*activite*", "*physipe*"/"*physique*", "*birafats*"/"*bienfaits*", "*avir*"/"*avoir*", "*santé de*", "*cour*"/"*cœur*", "*muceles*"/"*muscles*", "*tenis*"/"*tennis*", "*finalenent*"/"*finalelement*", "*faurt*"/"*faut*", "*pratiqur*"/"*pratiquer*"). Les deux notent des problèmes de construction de phrases (*manque majuscule/point, listes mal intégrées*), de ponctuation et de syntaxe (*utilisation de "car :", formulation "avoir une bonne santé de"*).

Feedback et propositions de correction/remédiation : Approches différentes. L'enseignant utilise des remarques ; des commentaires accompagnés de petits rappels des règles, placés juste à côté des erreurs. ChatGPT propose une correction rédigée, des conseils sur l'orthographe de mots spécifiques, la construction des phrases (*éviter les listes abruptes*), et la ponctuation.

III.2.3. Synthèse de l'Analyse Comparative : Évaluation Humaine/Evaluation ChatGPT

Cette synthèse résume les principales convergences et divergences observées lors de l'analyse comparative détaillée des évaluations fournies par un enseignant et celles générées par ChatGPT pour huit productions écrites d'élèves sur le thème du sport.

Convergences Générales

1. Compréhension du Sujet : Les deux évaluateurs s'accordent quasi systématiquement sur le fait que les élèves ont compris le sujet demandé. Ce critère fondamental est validé de manière similaire par l'humain et l'IA.

2. Identification des Erreurs Majeures (*Orthographe, Grammaire*) : Il existe une forte convergence dans la détection des erreurs factuelles de langue (*orthographe d'usage, accords de base, conjugaisons simples*). Les fautes récurrentes comme "sport" vs "sporte", les erreurs sur les mots courants ("*important*", "*physique*", "*bienfaits*", "*maladies*") ou les accords simples sont repérées par les deux.

3. Repérage de la Structure Globale : L'enseignant et ChatGPT identifient généralement de manière similaire la présence (*ou l'absence*) d'une structure de base (*introduction, développement, conclusion*) et l'utilisation de connecteurs logiques.

4. Appréciation des Copies Très Faibles ou Très Fortes : Pour les copies manifestement très faibles (*Copie 4, notes 9 et 8*) ou très correctes sur le plan formel (*Copie 5, notes 18.5 et 16*), l'appréciation générale tend à converger, même si la notation peut légèrement différer.

Divergences Générales

1. Notation et Sévérité : C'est un point de divergence majeur. L'enseignant semble globalement plus indulgent et parfois l'inverse dans sa notation (*cf. Copie 3 : 18 vs 13 ; Copie 5 : 18.5 vs 16 ; Copie 6 : 17.5 vs 15.5 Copie 7 : 11.5 vs 18*). ChatGPT applique sa grille de manière plus constante, conduisant parfois à des écarts de notes significatifs (*jusqu'à 5 points*).

2. Grille et Pondération des Critères : Les grilles utilisées diffèrent. Celle de l'enseignant est plus détaillée, incluant des aspects comme le "*style personnel/créatif*". La maîtrise particulière de la "*conjugaison*" ou des "*phrases déclaratives*" constitue les éléments linguistiques à réinvestir durant la séquence liée à cette production écrite. ChatGPT utilise

une grille plus standard. La pondération implicite ou explicite des critères varie, l'enseignant semblant parfois valoriser davantage le fond ou l'effort perçu que la stricte correction formelle par rapport à ChatGPT.

3. Analyse de la Syntaxe et du Style : Bien que les erreurs syntaxiques de base soient souvent repérées par les deux, ChatGPT semble plus apte à identifier les répétitions, les maladresses stylistiques, les phrases trop longues ou mal articulées, et à proposer des reformulations. L'enseignant note parfois ces aspects (*via codes ou commentaires*) mais de manière moins systématique ou détaillée dans la transcription fournie.

4. Nature et Format du Feedback : C'est une différence fondamentale. Lorsqu'il s'agit de l'évaluation d'une production écrite, l'enseignant assume une fonction dynamique de soutien à la progression de l'élève. Il propose des corrections dans le texte, reformule des phrases douteuses et opère des notes pédagogiques en marge, telles que « phrase maladroite », « majuscule oubliée », « améliore ton écriture ». Pour clarifier ses retours, l'enseignant recourt à des stylos de couleurs différentes (rouge, vert, bleu, violet) en fonction du type de retour : comment traditionnel, correction, rappel de règle, remarque ou proposition. Le retour opère une signalisation sur la copie de l'élève qui prend des formes variées (soulignement, encadrement, barré, note). Il a recours aussi à des flèches, crochets ou autres types d'encadrements pour « signaler » une erreur ou suggérer une amélioration. Cette trace écrite qui allie visuellement une forme et des retours plus spécifiquement caractérisés permet à l'élève de comprendre ce qui doit être ajusté et comment progresser. Elle constitue donc un levier pédagogique de la compétence scripturale acquise progressivement. ChatGPT fournit un feedback entièrement rédigé, incluant systématiquement : une liste de points positifs et négatifs, une proposition de correction complète du texte de l'élève, une grille d'évaluation commentée, et souvent des conseils de remédiation personnalisés voire des exercices ciblés.

5. Prise en Compte du Contexte/Effort : L'enseignant, par sa connaissance des élèves et du contexte, peut implicitement intégrer une appréciation de l'effort ou de la progression, ce qui peut influencer sa notation (*cf. indulgence apparente sur certaines copies bien notées malgré des erreurs*). ChatGPT réalise une évaluation plus "objective" ou "décontextualisée", basée uniquement sur le texte fourni.

6. La note attribuée à la copie 7 est insuffisante en raison d'une méprise de la part de ChatGPT, provoquée par une méconnaissance des consignes et des objectifs d'apprentissage relatifs à la production écrite, comme l'organisation d'un texte explicatif avec l'utilisation de connecteurs d'énumération, de phrases déclaratives et de la ponctuation. Bien que ChatGPT ait présenté une approche générale, il n'a pas considéré le contexte et la situation d'apprentissage que l'enseignant, lui, privilégie.

Conclusion partielle

Conclusion Partielle

L'analyse comparative menée entre l'évaluation humaine réalisée par l'enseignant et l'évaluation automatisée effectuée par ChatGPT sur un corpus de huit productions écrites d'élèves a permis de mettre en lumière des points de convergence significatifs mais aussi des divergences fondamentales entre les deux approches. Les convergences observées sont rassurantes sur certains aspects fondamentaux. Les deux évaluateurs s'accordent largement sur la compréhension du sujet par les élèves et sur l'identification des erreurs factuelles de langue les plus manifestes (orthographe d'usage, accords de base). La structure globale des textes et l'appréciation générale des copies situées aux extrêmes (très faibles ou très correctes formellement) tendent également à être évaluées de manière similaire. Toutefois, les divergences sont marquées et soulèvent des questions importantes pour la pratique évaluative. La notation constitue un écart majeur, l'enseignant manifestant une tendance à la clémence ou à une valorisation différente des efforts, tandis que ChatGPT applique sa grille de manière plus systématique et parfois plus sévère. Les critères et leur pondération diffèrent également, l'enseignant intégrant potentiellement des dimensions qualitatives ou contextuelles que l'IA ne peut appréhender. Si les erreurs de base sont repérées par les deux, ChatGPT démontre une capacité supérieure à identifier les maladresses stylistiques, les répétitions et les problèmes de syntaxe plus fins. La différence la plus flagrante réside dans la nature et le format du feedback : codes d'annotation nécessitant une interprétation pour l'enseignant contre feedback textuel détaillé, corrections complètes et propositions de remédiation pour ChatGPT. Enfin, l'évaluation de l'enseignant reste intrinsèquement contextualisée, tandis que celle de l'IA est purement basée sur le texte soumis. En conclusion de cette analyse spécifique, il apparaît clairement que si l'IA, et notamment ChatGPT, peut offrir un soutien précieux pour l'évaluation des aspects formels de la langue et la génération d'un feedback détaillé et rapide, elle ne saurait se substituer au jugement nuancé et contextualisé de l'enseignant. Les forces et faiblesses respectives des deux approches plaident fortement en faveur d'une articulation réfléchie et d'une complémentarité : l'IA comme outil d'aide à la détection des erreurs de surface et à la production de feedback formatif, l'enseignant conservant la responsabilité de l'évaluation sommative, de l'appréciation des dimensions qualitatives profondes et de l'adaptation pédagogique au contexte spécifique de chaque apprenant. Cette étude renforce l'idée que l'avenir de l'évaluation réside probablement dans des approches hybrides intelligemment conçues.

Conclusion Générale

Conclusion Générale

Au terme de ce mémoire consacré à l'évaluation des productions écrites en Français Langue Étrangère (*FLE*), il convient de synthétiser les principaux apports de notre réflexion et de notre étude expérimentale, afin de répondre à la problématique qui a guidé notre démarche. L'évaluation, processus essentiel à l'apprentissage, se trouve aujourd'hui à la croisée des chemins, entre les pratiques établies de l'évaluation humaine menée par l'enseignant et les potentialités offertes par l'émergence fulgurante des outils d'Intelligence Artificielle (*IA*), notamment les modèles de langage génératifs comme ChatGPT. Notre question centrale était de déterminer dans quelle mesure l'intelligence artificielle peut-elle compléter et améliorer l'évaluation humaine des productions écrites en FLE sans altérer la qualité du jugement porté sur ces textes ? Cette interrogation soulevait des questions subsidiaires quant à la capacité de l'IA à réduire les erreurs, à sa place potentielle (*complément ou remplacement*) et à la complémentarité pédagogique entre les retours automatisés et ceux de l'enseignant.

Notre parcours nous a d'abord conduits à explorer les fondements et les enjeux de l'évaluation humaine dans le premier chapitre. Nous avons souligné la richesse de l'analyse qualitative que permet l'expertise de l'enseignant, sa capacité à appréhender la cohérence globale, l'originalité, l'intention communicative, et à adapter son jugement au contexte spécifique de l'apprenant. Cependant, nous avons aussi rappelé les limites inhérentes à cette approche : la subjectivité potentielle, la lourde charge de travail que représente la correction de nombreuses copies, les effets parasites pouvant influencer le jugement, et la difficulté, faute de temps, à fournir systématiquement un feedback détaillé et personnalisé à chaque élève.

Le deuxième chapitre s'est attaché à décrypter les mécanismes et les promesses de l'évaluation automatisée, propulsée par les avancées du Traitement Automatique du Langage Naturel et de l'IA. Nous avons mis en lumière les forces potentielles de ces outils : une objectivité accrue sur les critères formels (*orthographe, grammaire de base*), une rapidité d'exécution permettant un feedback quasi immédiat, une constance dans l'application des critères, et, avec les IA génératives, la capacité à produire des retours explicatifs et structurés. Néanmoins, les limites de ces technologies sont également apparues clairement : leur difficulté à saisir la créativité, les nuances stylistiques, l'ironie, la pertinence culturelle ou

|Conclusion Générale

l'intention profonde de l'auteur, ainsi que les risques liés aux biais algorithmiques et au manque de transparence de certains modèles.

L'étude comparative menée dans le troisième chapitre, confrontant l'évaluation d'une enseignante à celle de ChatGPT sur un corpus de huit copies d'élèves, a permis d'ancrer ces considérations théoriques dans une réalité pratique. Les résultats ont confirmé empiriquement plusieurs des forces et faiblesses anticipées. Une convergence notable a été observée dans l'identification des erreurs factuelles de langue les plus évidentes, la reconnaissance de la compréhension du sujet par les élèves, et le repérage de la structure textuelle globale. Cependant, des divergences significatives sont apparues, particulièrement concernant la notation et la sévérité. L'enseignant se montrant globalement plus indulgent que l'IA, l'analyse des aspects stylistiques et syntaxiques fins, où ChatGPT a démontré une capacité supérieure à repérer répétitions et maladresses et surtout la nature et le format du feedback. L'enseignant guide l'élève en corrigeant directement sur sa copie : il reformule, commente et utilise des couleurs, signes (*flèches, encadrés, soulignements*) pour clarifier les erreurs et proposer des améliorations, tandis que ChatGPT fournissait systématiquement un retour entièrement rédigé, incluant points positifs et négatifs, une proposition de correction intégrale, une grille commentée et des pistes de remédiation. Enfin, l'étude a souligné l'incapacité de l'IA à prendre en compte le contexte ou l'effort perçu, dimension implicite mais présente dans l'évaluation humaine.

Ces résultats nous permettent de répondre de manière nuancée à notre problématique centrale. L'intelligence artificielle, telle qu'incarnée par ChatGPT dans notre étude, apparaît comme un complément potentiellement précieux à l'évaluation humaine. Elle peut effectivement améliorer certains aspects du processus évaluatif, notamment en systématisant la détection des erreurs formelles et en offrant un feedback rapide, structuré et détaillé qui peut favoriser l'autonomie de l'apprenant. Sa capacité à proposer des corrections et des exercices ciblés constitue un atout indéniable pour la remédiation. Toutefois, notre analyse confirme qu'elle ne saurait remplacer le jugement de l'enseignant. Son évaluation, bien que constante, peut manquer de la souplesse et de la contextualisation nécessaires, conduisant à une sévérité parfois excessive ou à une incompréhension des subtilités de l'écrit. Utiliser l'IA seule altérerait indéniablement la qualité globale du jugement, en négligeant les dimensions créatives, stylistiques et intentionnelles fondamentales en production écrite. La complémentarité pédagogique entre les deux approches est donc évidente : l'IA excelle sur

|Conclusion Générale

le plan formel et structurel de bas niveau, tandis que l'humain reste indispensable pour l'appréciation qualitative de haut niveau.

Nos hypothèses initiales se trouvent ainsi largement confirmées. L'évaluation humaine demeure indispensable pour juger la cohérence profonde, l'originalité et l'intention communicative. L'intelligence artificielle constitue un soutien efficace pour garantir une meilleure objectivité sur certains critères et accélérer le processus, mais elle ne peut se substituer au rôle irremplaçable de l'enseignant.

Ces constats appellent à des implications pédagogiques claires. L'avenir de l'évaluation des productions écrites en FLE réside très probablement dans une approche hybride, qui tire parti des forces de chaque modalité. Il devient crucial de former les enseignants non seulement à l'utilisation technique de ces outils, mais surtout à leur intégration critique et réfléchie dans leurs pratiques. Savoir quand et comment utiliser l'IA, comment interpréter ses résultats, comment les articuler avec son propre jugement devient une compétence essentielle. Parallèlement, il est impératif de former les apprenants à interagir avec ces outils de manière autonome mais critique, en comprenant la nature du feedback reçu, ses limites, et en ne déléguant pas aveuglément leur processus de révision. Des scénarios d'intégration pourraient voir l'IA utilisée pour un premier feedback formatif sur les aspects formels, l'enseignant se concentrant ensuite sur le fond, la structure argumentative, la créativité, et l'évaluation sommative finale, où son jugement expert reste prépondérant.

Nous sommes conscients des limites de notre propre recherche. La taille réduite de l'échantillon, la concentration sur un seul outil IA et une seule transcription d'évaluation enseignante, ainsi que l'absence d'analyse de la réception du feedback par les élèves invitent à la prudence quant à la généralisation des résultats. Ces limites ouvrent cependant des perspectives stimulantes pour des recherches futures. Des études comparatives élargies, incluant davantage d'outils, d'élèves, de niveaux et de contextes d'apprentissage, sont nécessaires. Des recherches longitudinales permettraient de mesurer l'impact réel de l'utilisation de ces outils sur la progression effective des compétences scripturales. Il serait également pertinent d'étudier plus finement la perception et l'utilisation du feedback généré par l'IA, tant par les élèves que par les enseignants, et d'explorer les méthodologies d'évaluation hybrides les plus efficaces. Enfin, une vigilance constante s'impose quant aux biais potentiels véhiculés par ces algorithmes dans le contexte spécifique du FLE.

|Conclusion Générale

En conclusion, l'intelligence artificielle transforme indéniablement le paysage de l'évaluation en éducation, et l'enseignement du FLE n'y échappe pas. Si les outils comme ChatGPT offrent des possibilités fascinantes pour renouveler et améliorer certaines facettes de l'évaluation des productions écrites, leur intégration ne doit pas se faire sans discernement. L'enjeu est de les apprivoiser, de les comprendre et de les utiliser comme des assistants intelligents au service d'une pédagogie enrichie, sans jamais perdre de vue la finalité première de l'évaluation : favoriser l'apprentissage et le développement de compétences complexes chez l'apprenant. L'intelligence humaine de l'enseignant, sa capacité d'empathie, de contextualisation et de jugement nuancé, demeure et demeurera, à notre sens, au cœur de ce processus complexe et profondément humain qu'est l'évaluation.

Références Bibliographiques

Références bibliographiques

Articles scientifiques

- Adam, J.-M. (2011). Les textes : types et prototypes. Paris : Armand Colin.
- Alfter, D., Wilkens, R., Wang, X., Pintard, A., Tack, A., Yancey, K. & François, T. (2022). FABRA: French Aggregator-Based Readability Assessment toolkit. In Proceedings of LREC 2022.
- Astolfi, J.-P. (1997). L'erreur, un outil pour enseigner. Paris: ESF.
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? Trends in Cognitive Sciences, 4(11), 417-423.
- Barbier, M.-L. (2003). Écrire en L2 : bilan et perspectives des recherches. Arob@se, 7(1-2), 5-21.
- Beacco, J.-C. (2007). L'approche par compétences dans l'enseignement des langues. Paris: Didier.
- Bereiter, C., & Scardamalia, M. (1987). The psychology of written composition. Hillsdale, NJ : Lawrence Erlbaum.
- Bibal, A., Cardon, R., Alfter, D., Wilkens, R., Wang, X., François, T. & Watrin, P. (2022). Is Attention Explanation? An Introduction to the Debate. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 3889-3900.
- Boniface, C. (1992). Les ateliers d'écriture. Paris : Retz.
- Corder, S. P. (1980). La sollicitation de données d'interlangue. Langages, 57, 29-38.
- Cornaire, C., & Raymond, P. M. (1999). La production écrite. Paris : CLE International.

Références bibliographiques

- Cornillie, F., Desmet, P., & Baten, K. (2023). L'évaluation automatisée des productions écrites en langues étrangères : état des lieux et perspectives. *Revue française de linguistique appliquée*, 28(1), 65-80.
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, 49(2), 222-251.
- Dejean-Thircuir, C., & Mangenot, F. (2006). Tâches et scénarios de communication dans les classes virtuelles. *Les Cahiers de l'ASDIFLE*, 17, 310-321.
- Durel, P. (2006). Utilisation de l'assistant grammatical Antidote dans le cadre d'activités de révision. *Apprentissage des langues et systèmes d'information et de communication*, 9(1), 33-60.
- Ferris, D. (2002). *Treatment of error in second language student writing*. Ann Arbor: University of Michigan Press.
- François, T. (2022). Automated essay scoring: where do you stand and where are we going? Communication présentée à la 57ème conférence ALTE, Paris.
- Germain, C. (1993). *Évolution de l'enseignement des langues : 5000 ans d'histoire*. Paris : CLE International.
- Guernine, A. A., & Kacem, I. (2024). L'exploitation de l'erreur dans la production écrite chez les apprenants de 4ème année moyenne. Mémoire de Master, Université Ibn Khaldoun-Tiaret.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ : Lawrence Erlbaum.
- Huver, E., & Springer, C. (2011). *L'évaluation en langues*. Paris: Didier.
- Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57-71). Mahwah, NJ: Lawrence Erlbaum.

Références bibliographiques

- Little, D. (2007). Language learner autonomy: Some fundamental considerations revisited. *Innovation in Language Learning and Teaching*, 1(1), 14-29.
- Lukogho Vagheni, G. (2019). La production écrite par mail en didactique du français langue étrangère : cas de l'Institut français de Goma. *Synergies Afrique des Grands Lacs*, 8, 115-129.
- Lukogho Vagheni, K. (2020). Le mail comme outil didactique pour la production écrite en classe de FLE. *Synergies Afrique des Grands Lacs*, 9, 73-85.
- Makhoulouf, M. & Legros, D. (2024). Analyse des pratiques d'évaluation en FLE : enseignant et IA face aux productions écrites. ResearchGate, Publication 386101735.
- Manchón, R. M. (2009). *Writing in foreign language contexts: Learning, teaching, and research*. Bristol : Multilingual Matters.
- Mangenot, F. (2000). L'intégration des TIC dans une perspective systémique. *Les Langues modernes*, 3, 38-44.
- Mangenot, F., & Soubrié, T. (2014). Le web social au service de tâches d'écriture. *Recherches*, 60, 89-109.
- Marquilló Larruy, M. (2003). *L'interprétation de l'erreur*. Paris : CLE International.
- Moirand, S. (1979). *Situations d'écrit*. Paris : CLE International.
- Moussouri, E. (2023). Argumenter à l'écrit avec un modèle d'IA. Pourquoi pas ? Oui, mais comment. ResearchGate, Publication 387220006.
- Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge : Cambridge University Press.
- Ollivier, C. (2012). Approche interactionnelle et didactique invisible – Deux concepts pour la conception et la mise en œuvre de tâches sur le web social. *Alsic*, 15(1).

Références bibliographiques

- Oriol-Boyer, C. (2013). *Ateliers d'écriture*. Grenoble : CEDITEL.
- Puren, C. (1988). *Histoire des méthodologies de l'enseignement des langues*. Paris: CLE International.
- Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal*, 59(1), 23-30.
- Rosen, É. (2009). La perspective actionnelle et l'approche par les tâches en classe de langue. *Le français dans le monde, recherches et applications*, 45, 487-498.
- Schneuwly, B. (1994). Genres et types de discours : considérations psychologiques et ontogénétiques. In Y. Reuter (Ed.), *Les interactions lecture-écriture* (pp. 155-173). Berne : Peter Lang.
- Tagliante, C. (2005). *L'évaluation et le Cadre européen commun*. Paris : CLE International.
- Tréville, M.-C. (2000). *Vocabulaire et apprentissage d'une langue seconde*. Montréal : Éditions Logiques.
- Veltcheff, C., & Hilton, S. (2003). *L'évaluation en FLE*. Paris : Hachette.
- Viau, R. (1994). *La motivation en contexte scolaire*. Bruxelles : De Boeck.
- Vigner, G. (2001). *Enseigner le français comme langue seconde*. Paris : CLE International.
- Vygotski, L. (1997). *Pensée et langage*. Paris : La Dispute.
- Wilkens, R., Seibert, D., Wang, X. & François, T. (2022). MWE for Essay Scoring English as a Foreign Language. In *Proceedings of the READI 2022 workshop*.
- Yamaguchi, N., Alfter, D., Sugiyama, K. & François, T. (2023). Towards a Verb Profile: distribution of verbal tenses in FFL textbooks and in learner productions. *Proceedings of NLP4CALL 2022*.

Références bibliographiques

Ouvrages de référence

- Barbot, M.-J., & Pugibet, V. (2002). Apprentissages des langues et technologies : usages en émergence. Le français dans le monde, recherches et applications.
- Conseil de l'Europe. (2001). Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer. Paris : Didier.
- Cuq, J.-P., & Gruca, I. (2005). Cours de didactique du français langue étrangère et seconde. Grenoble : Presses Universitaires de Grenoble.

Ressources en ligne

- Éducation simplifiée. (2024). Comment développer l'expression écrite des élèves. [Vidéo]. YouTube.
- Fulep, J. (2023, 8 mai). Comment travailler la production écrite en classe de FLE ? [Vidéo]. YouTube. https://www.youtube.com/watch?v=49giAn_wtro
- Google. (2024). Gemini : Modèle d'IA pour l'ère agentive. Récupéré de <https://blog.google/intl/fr-fr/nouvelles-de-lentreprise/gemini-2-0/>
- Google. (2024). Questions fréquentes sur les applications Gemini. Récupéré de <https://gemini.google.com/faq?hl=fr>
- Le français avec Nouma. (2022). Didactique du français fiche de production de l'écrit. [Vidéo]. YouTube.
- Mahfoud, Z. (2024). DIDACTIQUE DE L'ÉCRIT : COURS 8 - LA PRODUCTION. [Vidéo]. YouTube.

Ressources institutionnelles

Références bibliographiques

- France Éducation International. (2024). FIDELIA : la solution d'aide à la correction automatisée. Récupéré de <https://www.france-education-international.fr/article/fidelia-la-solution-daide-la-correction-automatisee>
- Réseau Canopé. (2023). Comment évaluer une production d'écrits assistée par une IA ? Récupéré de <https://www.canotech.fr/a/37903/comment-evaluer-une-production-decrits-assistee-par-une-ia>
- Université catholique de Louvain, CENTAL. (2024). CECR-FR Rator. Récupéré de <https://www.uclouvain.be/fr/instituts-recherche/ilc/cental/cecr-fr-rator>

Ressources pédagogiques

- Le Point du FLE. (s.d.). Expression écrite Fiches pédagogiques. Récupéré de <https://www.lepointdufle.net/penseigner/expression-ecrite-fiches-pedagogiques.htm>
- Le Point du FLE. (s.d.). Production écrite Fiches pédagogiques. Récupéré de <https://www.lepointdufle.net/penseigner/production-ecrite-fiches-pedagogiques.htm>
- TV5MONDE. (s.d.). Enseigner le français FLE avec TV5MONDE. Récupéré de <https://enseigner.tv5monde.com/>

Thèses et mémoires

- Benramdane, A. & Benramdane, S. (2024). L'exploitation de l'erreur dans la production écrite en classe de FLE. Cas des apprenants de 3ème année secondaire. Université Ibn Khaldoun Tiaret.
- Boukhannouche, L. (2023). L'évaluation de la production écrite dans l'enseignement du FLE. Université de Guelma.

Annexes

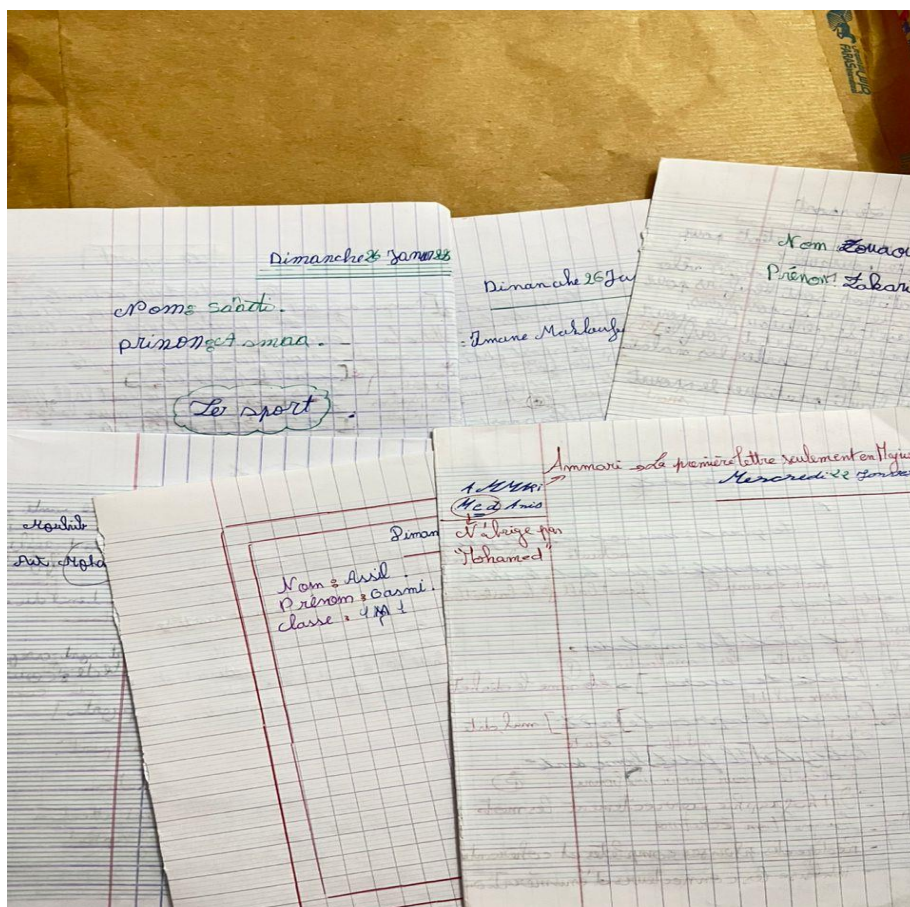
Annexe N°1 : Photos des élèves lors de l'activité d'écriture

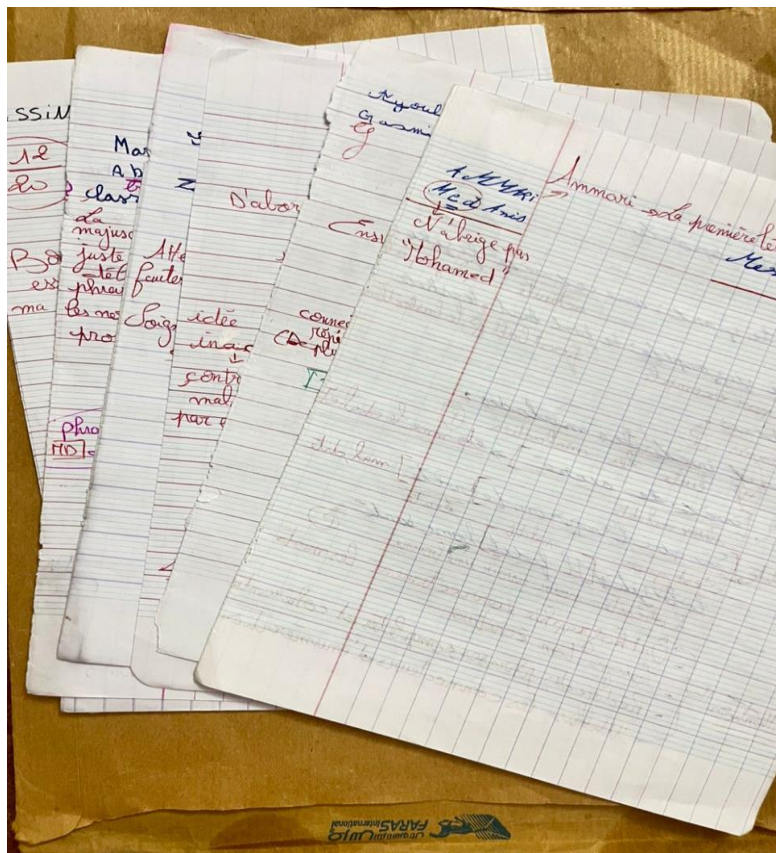
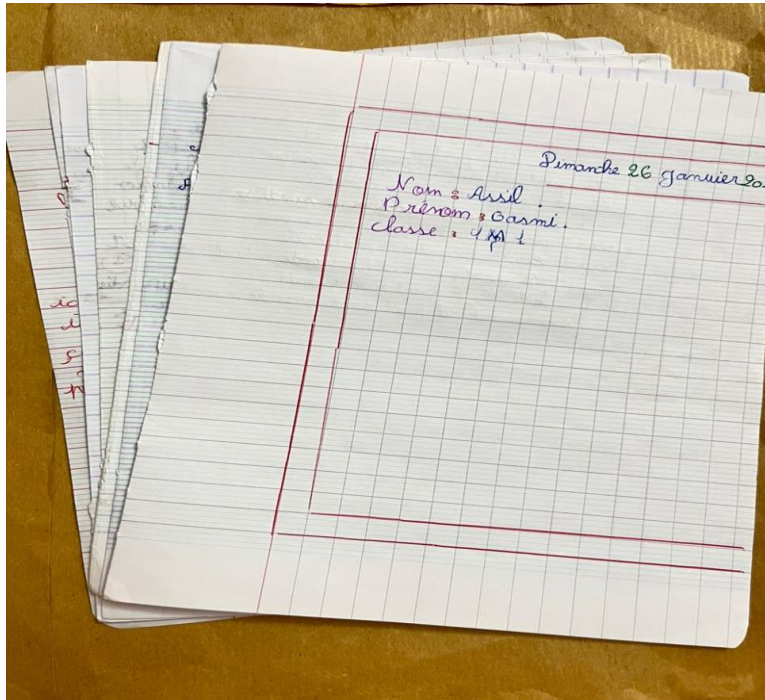


Annexes



Annexe N°2 : Corpus





Résumé

Ce mémoire de master explore l'évaluation des productions écrites en Français Langue Étrangère (FLE) à l'ère de l'Intelligence Artificielle (IA). Face aux défis de l'évaluation humaine traditionnelle (subjectivité potentielle, charge de travail) et aux promesses des outils d'évaluation automatisée, notamment les IA génératives comme ChatGPT.

Une étude comparative a été menée, confrontant l'évaluation d'un enseignant (via transcription) à celle de ChatGPT sur un corpus de huit productions écrites d'élèves au Collège, sur le thème du sport. L'analyse a révélé des convergences notables dans l'identification des erreurs factuelles de langue (orthographe, grammaire de base) et la compréhension globale du sujet par les élèves. Toutefois, des divergences significatives sont apparues concernant la sévérité de la notation (l'enseignant étant souvent plus indulgent), la finesse de l'analyse stylistique et syntaxique (ChatGPT se montrant plus sensible aux répétitions et maladresses), et surtout la nature et le format du feedback (codes d'annotation pour l'enseignant versus texte rédigé, corrections complètes et exercices de remédiation pour ChatGPT).

Les résultats confirment que l'IA peut constituer un complément utile à l'évaluation humaine, notamment pour le feedback formatif sur les aspects linguistiques de surface et la rapidité du retour. Elle peut améliorer certains aspects mais ne saurait remplacer le jugement expert de l'enseignant, indispensable pour évaluer les dimensions qualitatives profondes de l'écrit. Ces constats plaident pour une approche hybride de l'évaluation, combinant les forces des deux modalités. Ils soulignent également l'importance cruciale de former les enseignants et les apprenants à une utilisation critique et éclairée de ces outils, tout en réaffirmant la centralité du jugement humain dans le processus évaluatif final.

Mots-clés : IA, évaluation humaine, productions écrites, évaluation automatisée, français langue étrangère (FLE).

Summary

This master's thesis explores the evaluation of written productions in French as a Foreign Language (FLE) in the era of Artificial Intelligence (AI). Faced with the challenges of traditional human evaluation (potential subjectivity, workload) and the promises of automated evaluation tools, particularly generative AI like ChatGPT.

A comparative study was conducted, confronting the evaluation of a teacher (via transcription) with that of ChatGPT on a corpus of eight written productions by middle school students on the theme of sports. The analysis revealed notable convergences in identifying factual language errors (spelling, basic grammar) and the overall understanding of the topic by the students. However, significant divergences emerged concerning the severity of grading (the teacher often being more lenient), the finesse of stylistic and syntactic analysis (ChatGPT proving more sensitive to repetitions and awkward phrasing), and especially the nature and format of the feedback (annotation codes for the teacher versus written text, complete corrections, and remedial exercises for ChatGPT).

The results confirm that AI can constitute a useful complement to human evaluation, particularly for formative feedback on surface linguistic aspects and the speed of return. It can improve certain aspects but cannot replace the expert judgment of the teacher, which is essential for evaluating the deep qualitative dimensions of writing. These findings argue for a hybrid approach to evaluation, combining the strengths of both modalities. They also underscore the crucial importance of training teachers and learners in the critical and informed use of these tools, while reaffirming the centrality of human judgment in the final evaluation process.

Key-words: AI, human evaluation, written productions, automated evaluation, French as a Foreign Language (FLE).

الملخص

يستكشف هذا البحث لنيل شهادة الماستر موضوع تقييم الإنتاجات الكتابية في مادة الفرنسية كلغة أجنبية (FLE) في عصر الذكاء الاصطناعي (IA) ويأتي هذا في ظل التحديات التي تفرضها طرق التقييم البشري التقليدي، مثل احتمال الذاتية وكثرة الأعباء، مقابل الوعود التي تقدمها أدوات التقييم المؤتم، ولا سيما الذكاء الاصطناعي التوليدي مثل ChatGPT.

تم إجراء دراسة مقارنة بين تقييم أستاذ (عن طريق قراءة النسخ الأصلية) وتقييم ChatGPT لثمانين إنتاجات كتابية لتلاميذ في الطور المتوسط (CEM) حول موضوع الرياضة. وقد كشفت التحليلات عن توافق ملحوظ بين الطرفين في تحديد الأخطاء اللغوية السطحية (كالإملاء وقواعد اللغة الأساسية) وفهم التلاميذ العام للموضوع. ومع ذلك، ظهرت اختلافات كبيرة من حيث درجة الصرامة في التقييم (حيث أظهر الأستاذ تساهلاً أكبر)، ومن حيث دقة التحليل الأسلوبية والتركيبي) حيث كان ChatGPT أكثر حساسية للتكرار والركاكة، (والأهم من ذلك، في طبيعة وتنسيق التغذية الراجعة) الأستاذ يعتمد رموزاً توضيحية، بينما يقدم ChatGPT نصاً محرراً، وتصحيحات كاملة، وتمارين علاجية. تؤكد النتائج أن الذكاء الاصطناعي يمكن أن يكون أداة مكملة مفيدة للتقييم البشري، خاصة في التغذية الراجعة التكوينية المتعلقة بجوانب اللغة السطحية وسرعة الرد. إلا أنه لا يمكنه أن يحل محل الحكم الخبير للأستاذ، الذي يبقى ضرورياً لتقييم الجوانب العميقة والنوعية في الكتابة. وتدعو هذه المعطيات إلى تبني مقاربة هجينة تجمع بين مزايا الطريقتين، كما تؤكد على أهمية تكوين الأساتذة والمتعلمين على استخدام هذه الأدوات بطريقة نقدية وواعية، مع التأكيد في الوقت نفسه على مركزية التقدير البشري في عملية التقييم النهائية.

الكلمات المفتاحية: الذكاء الاصطناعي /التقييم البشري / الإنتاجات الكتابية / التقييم المؤتم /الفرنسية لغة أجنبية.