



DIFFICULTES RECURRENTES DANS L'APPRENTISSAGE DE L'ECRITURE EN FRANÇAIS LANGUE ETRANGERE : ANALYSE D'UN CORPUS D'APPRENANTS A L'AIDE DE L'IA

Eleni MAVROPOULOU

Département de langue et littérature françaises,

Université Aristote de Thessaloniki

emavropoulou@frl.auth.gr

Résumé : Cette étude examine les difficultés récurrentes rencontrées par les apprenants dans la production écrite en français langue étrangère (FLE), à partir d'une analyse statique d'un corpus de 50 copies manuscrites de 25 étudiants et d'un soutien méthodologique issu de l'intelligence artificielle (IA) et plus précisément du logiciel ChatGPT AI open. L'analyse, combinant traitement automatique et relecture humaine des erreurs, met en évidence les principales difficultés persistantes : orthographe, morphologie (conjugaison, accord) et syntaxe, suivies de loin par le vocabulaire et la cohérence du discours. Des exemples authentiques tirés des écrits des apprenants illustrent ces erreurs fréquentes. Les résultats montrent que, malgré un niveau intermédiaire en français, certaines erreurs fondamentales persistent chez les apprenants (par exemple, des erreurs récurrentes d'accord ou de conjugaison). L'article examine les causes possibles de ces difficultés (complexité du français écrit, influence de la langue maternelle, automatisation incomplète de certaines règles, etc.) et propose des approches didactiques concrètes pour y remédier. En conclusion, l'étude souligne l'intérêt d'utiliser des outils d'IA pour analyser de grands corpus d'apprenants afin d'affiner les pratiques pédagogiques et de mieux cibler les erreurs récurrentes dans l'enseignement du FLE.

Mots-clés : difficultés d'apprentissage ; production écrite ; FLE ; intelligence artificielle ; erreurs récurrentes.

RECURRENT DIFFICULTIES IN LEARNING WRITTEN FRENCH AS A FOREIGN LANGUAGE: ANALYSIS OF A LEARNER CORPUS WITH THE HELP OF AI

Abstract : This article analyses the recurrent difficulties encountered by university learners of French as a foreign language in written production. We draw on a corpus of 50 handwritten exam scripts produced by 25 students and digitised for analysis. The errors are first categorised into five main domains (orthography, morphology, syntax, lexis and textual coherence), then quantified in order to identify the most frequent patterns. A static corpus analysis is combined with the use of AI-assisted tools, which make it easier to locate recurrent sequences, to visualise distributions and to refine the qualitative examination of the learners' productions. The results show a clear predominance of orthographic and morphological errors, followed by problems relating to sentence structure, lexical choice and cohesion at discourse level. These findings confirm the persistence of certain difficulties already documented in the literature and highlight the specific challenges posed by written French to learners whose first language has different orthographic and

morphosyntactic systems. From a pedagogical perspective, the study argues for more systematic, data-informed work on feedback and remediation, and for a thoughtful integration of AI-assisted analysis into the teaching of writing in FFL.

Keywords: writing difficulties; French as a foreign language (FFL); error analysis; artificial intelligence; learner corpus

Introduction

La personne qui apprend une langue non maternelle, elle apprend une langue étrangère (Cuq & Gruca, 2017). Apprendre à écrire dans une langue étrangère est un processus complexe, souvent semé d'embûches. L'écriture requiert des compétences linguistiques et cognitives avancées, ce qui en fait l'une des compétences les plus difficiles à maîtriser pour les apprenants d'une langue seconde (Garcia-Debanç, 1995 ; Richards & Renandya, 2002 ; Raymond & Cornaire, 1999). En FLE, les apprenants rencontrent des difficultés récurrentes telles que des erreurs grammaticales (conjugaison, accord), des fautes d'orthographe et des structures syntaxiques maladroites. Ces erreurs, loin d'être aléatoires, reflètent souvent des règles intermédiaires que les apprenants construisent eux-mêmes, un phénomène connu sous le nom d'interlangue (Selinker, 1972). De plus, comme le souligne Astolfi (1997), « il n'y a pas d'apprentissage sans erreurs », car celles-ci font partie intégrante du processus d'acquisition.

Malgré les progrès réalisés grâce aux méthodologies communicatives et orientées vers l'action dans l'enseignement du FLE (Cuq & Gruca, 2003), de nombreuses études et observations en classe montrent que certains types d'erreurs ont tendance à persister dans le temps. On observe fréquemment que, d'un semestre à l'autre, les apprenants commettent les mêmes erreurs sur des points linguistiques qui ont été enseignés à plusieurs reprises. Cette observation soulève des questions : s'agit-il d'un manque d'assimilation de ces concepts, d'une fossilisation des erreurs (Selinker, 1972) ou d'une inadéquation des pratiques pédagogiques face à ces difficultés récurrentes ? Par exemple, dans la production écrite, de nombreux apprenants continuent d'omettre les marqueurs du pluriel ou de confondre les homophones grammaticaux, même après plusieurs années d'étude du français (Blanche-Benveniste & Chervel, 1969 ; Guyon, 2003).

L'article cherche essentiellement à répondre à la question suivante :

« Quelles sont les difficultés/erreurs récurrentes et persistantes que rencontrent des apprenants universitaires de français langue étrangère dans leur production écrite, et dans quelle mesure une analyse de corpus assistée par l'IA permet-elle de les repérer, les classer et les quantifier afin d'orienter la remédiation pédagogique ? »

Dans ce contexte, la présente étude vise à identifier et à analyser les difficultés les plus récurrentes dans l'apprentissage de l'écriture en FLE, en adoptant une perspective synchronique approfondie. En d'autres termes, nous examinons un large corpus de textes écrits par des apprenants à un stade donné de leur formation afin de déterminer quelles sont les erreurs les plus fréquentes et celles qui semblent persistantes. Pour ce faire, nous avons utilisé des outils d'IA capables de traiter le langage naturel afin de détecter et de classer les erreurs dans un large corpus de textes écrits par des apprenants. L'utilisation de l'IA dans l'analyse des travaux écrits offre un double avantage : un traitement plus rapide et la capacité de détecter des schémas d'erreurs récurrents dans un volume important de donnée, analyse peaufinée ensuite par l'enseignant-chercheur. Cette approche s'inscrit dans le cadre de recherches récentes explorant la contribution de l'IA à l'enseignement des langues, tout en restant centrée sur une question didactique classique en FLE : la gestion des erreurs (Corder, 2015).

Le défi est double. D'une part, il s'agit de mieux comprendre la nature et la fréquence des erreurs persistantes dans l'écriture en FLE, afin d'identifier les tendances générales et d'orienter les méthodes d'enseignement. D'autre part, nous devons évaluer dans quelle mesure un outil d'IA peut soutenir l'analyse didactique de ces erreurs et fournir des informations supplémentaires aux enseignants. Après avoir passé en revue le cadre théorique relatif aux erreurs dans l'enseignement des langues et les spécificités de la production écrite en FLE, nous présenterons la méthodologie de notre analyse assistée par l'IA. Nous présenterons ensuite les principaux résultats obtenus, illustrés par des exemples tirés des copies des apprenants, avant de discuter des implications de ces résultats pour l'enseignement du FLE et des perspectives offertes par cette approche combinant expertise humaine et outils numériques.

1. Cadre théorique

Dans le domaine de l'enseignement des langues, les erreurs ont longtemps été stigmatisées comme des échecs pédagogiques, avant d'être réhabilitées comme des indicateurs du processus d'apprentissage (Corder, 2015 ; Astolfi, 1997). Une erreur linguistique répétée par un apprenant n'est généralement pas aléatoire : elle reflète souvent l'état de son interlangue, c'est-à-dire le système intermédiaire de règles que l'apprenant a construit pour combler le fossé entre sa langue maternelle et la langue cible (Selinker, 1972). L'analyse des erreurs permet ainsi de mieux comprendre les processus cognitifs de l'apprenant et d'adapter l'enseignement en conséquence (Astolfi, 1997). Par exemple, des erreurs récurrentes dans la conjugaison du passé – telles que l'omission des verbes auxiliaires « être » ou « avoir » ou la formation incorrecte du participe passé – indiquent que l'apprenant n'a pas encore intériorisé la règle correspondante et

continue d'appliquer soit une généralisation incorrecte, soit l'influence de sa langue maternelle.

En ce qui concerne la production écrite, en particulier en FLE, plusieurs caractéristiques de la langue française contribuent à la difficulté d'apprentissage. Premièrement, le français a un système orthographique notoirement complexe et non phonétique : les nombreuses lettres muettes, les règles d'accord en genre et en nombre qui sont marquées à l'écrit mais pas à l'oral, les accents, etc. sont autant de pièges pour les apprenants (Blanche-Benveniste & Chervel, 1969 ; Guyon, 2003). La maîtrise de l'orthographe française est un défi cognitif à long terme (Guyon, 2003). Deuxièmement, la grammaire française possède des systèmes de conjugaison et d'accord relativement riches (morphologie verbale, accord du sujet et du verbe, accord du participe passé, accord des adjectifs, etc.) qui multiplient les points de difficulté. De nombreux apprenants du français langue étrangère commettent des erreurs d'accord (par exemple, ne pas marquer la forme plurielle ou féminine appropriée) ou de conjugaison (mauvais choix de temps, terminaisons incorrectes), qui peuvent être dues soit à une simplification de la part de l'apprenant, soit à une interférence de sa langue maternelle lorsque celle-ci fonctionne différemment (Odlin, 1989).

De plus, la syntaxe française peut être source de confusion pour les locuteurs non natifs. La place des pronoms, l'ordre obligatoire sujet-verbe-objet, l'utilisation des prépositions et la structure des phrases complexes sont autant de sources d'erreurs syntaxiques. Par exemple, un apprenant dont la langue maternelle n'utilise pas de verbe auxiliaire pour former le passé pourrait écrire « Il arrivé en retard » au lieu de « Il est arrivé en retard », en omettant le verbe auxiliaire être. De même, l'influence de constructions spécifiques à la langue maternelle peut donner lieu à des phrases françaises mal structurées telles que « Parce que il pleut, je suis resté à la maison », où l'ordre des mots ou l'utilisation d'un connecteur trahit une traduction littérale. Ces erreurs dites d'interférence sont bien documentées dans la littérature (Odlin, 1989) et font partie des difficultés récurrentes en FLE.

Enfin, au-delà des aspects formels, la capacité à organiser un texte de manière cohérente et cohésive constitue un autre niveau de difficulté. La cohérence textuelle, c'est-à-dire l'articulation des idées dans un ordre logique et la structuration d'un texte (introduction, développement, conclusion, etc.), est une compétence qui s'acquiert progressivement. Carter-Thomas (1999) souligne l'importance d'enseigner explicitement la cohérence textuelle dans la pédagogie de l'écriture, car de nombreux apprenants produisent des textes décousus ou manquant de progression claire. Dans cette étude, nous nous concentrons principalement sur les erreurs linguistiques (orthographe, morphologie, syntaxe) qui peuvent être quantifiées objectivement, tout en tenant compte des difficultés

discursives liées à la cohérence textuelle lorsqu'elles apparaissent dans le corpus. Il convient de noter que les difficultés à planifier et à réviser le texte peuvent également contribuer à la persistance de certaines erreurs : par exemple, un apprenant qui ne relit pas son travail ou qui ne sait pas encore s'autocorriger laissera dans son texte des erreurs qui auraient pu être évitées par une relecture (Cuq & Gruca, 2003).

En résumé, le cadre théorique de cette recherche repose sur l'idée que les erreurs ont un sens dans l'apprentissage des langues (Corder, 2015; Astolfi, 1997). Les erreurs récurrentes en français écrit langue étrangère (FLE) résultent d'une combinaison de facteurs : linguistiques (complexité de l'orthographe et de la grammaire françaises), cognitifs (stratégies d'apprentissage, stade d'acquisition du langage écrit) et sociolinguistiques (influence de la langue maternelle, conditions d'exposition au langage écrit, etc.). La compréhension de ces erreurs à travers l'analyse d'un corpus d'apprenants permet d'identifier celles qui ont tendance à persister malgré l'enseignement. La contribution de l'IA dans ce contexte est d'aider les chercheurs à détecter systématiquement ces phénomènes au sein d'un corpus important, afin d'identifier objectivement les erreurs les plus fréquentes et les schémas récurrents sur lesquels concentrer l'attention pédagogique.

2. Méthodologie

2.1. Participants et corpus

Cette recherche a été menée dans un cadre universitaire (Université d'Aristote de Thessalonique en Grèce, année académique 2024-2025) auprès d'un groupe d'apprenants adultes de français langue étrangère (25 au total) dont la langue maternelle n'était pas le français. Le groupe était hétérogène en termes de L1 (une majorité d'apprenants avaient le grec pour L1 et trois l'italien), mais relativement homogène en termes de niveau de français, évalué à environ le niveau intermédiaire B1 du CECR (Council of Europe, 2020 ; Council of Europe, 2001) au moment de la collecte des données. Nous avons compilé un corpus de 50 textes manuscrits produits par ces étudiants dans le cadre de leur cours de FLE. Il s'agissait principalement de travaux écrits formels d'une longueur de 300 – 400 mots (dissertations, essais argumentatifs, lettres, résumés, etc.) réalisés dans le cadre des évaluations et des exercices du cours. Chaque copie a été scannée afin de pouvoir être utilisée sous forme numérique.

2.2. Procédure d'analyse des erreurs

Pour analyser ce corpus, nous avons utilisé des outils d'IA spécialisés dans le traitement automatique du langage naturel (ChatGPT AI open). Tout d'abord, toutes les copies manuscrites ont été transcrites en texte numérique brut à l'aide

d'un logiciel de reconnaissance optique de caractères (OCR) afin de créer un corpus numérique exploitable. Nous avons ensuite utilisé un algorithme d'analyse linguistique capable de détecter les erreurs dans le texte d'un apprenant en comparant chaque phrase aux règles grammaticales et orthographiques de la langue française. Ce type d'outil, comparable à un correcteur grammatical intelligent, signale, par exemple, les verbes mal conjugués, les mots inconnus (potentiellement mal orthographiés) ou les accords manquants. Il convient de noter que l'IA utilisée ne remplace pas l'expertise humaine : elle apporte un soutien méthodologique en facilitant l'identification initiale des erreurs dans chaque texte. Les erreurs détectées automatiquement ont été vérifiées manuellement par le chercheur afin d'éliminer les faux positifs (par exemple, l'algorithme peut signaler à tort un mot inhabituel qui est en fait correct dans le contexte).

Nous avons ensuite classé les erreurs identifiées en plusieurs catégories prédéfinies, en fonction de leur nature linguistique : orthographe, morphologie (y compris les erreurs de conjugaison verbale et d'accord grammatical), syntaxe (erreurs de structure de phrase), lexicale (choix de mots inappropriés) et discours (cohérence et organisation du texte). Chaque erreur identifiée dans le corpus a ainsi été codée dans l'une de ces catégories. Nous avons également noté, lorsque cela était évident, si l'erreur pouvait résulter d'une interférence de la langue maternelle de l'apprenant (par exemple, un calque de structure ou un faux ami lexical), sans toutefois en faire une catégorie distincte, l'interférence étant ici considérée comme une cause possible d'erreurs dans les catégories ci-dessus.

Enfin, nous avons procédé à une analyse quantitative et qualitative des données relatives aux erreurs. Sur le plan quantitatif, nous avons compté le nombre total d'erreurs par catégorie dans l'ensemble du corpus afin d'établir la répartition des types d'erreurs et d'identifier les domaines linguistiques qui posent le plus de problèmes. Au total, près de 600 erreurs ont été identifiées dans les 50 textes, soit une moyenne d'une douzaine d'erreurs par texte (avec des variations liées à la longueur des textes et au profil de chaque apprenant). Parallèlement, nous avons examiné qualitativement la nature des erreurs récurrentes : si le même type d'erreur (par exemple, la confusion entre l'imparfait et le passé composé, ou l'omission systématique des accents) apparaissait dans plusieurs copies, que ce soit pour le même apprenant ou pour différents apprenants, nous l'avons noté comme un modèle d'erreur récurrent. L'IA nous a aidée dans cette étape en extrayant, par exemple, toutes les phrases du corpus contenant un certain type d'erreur (toutes les phrases où un participe passé est utilisé sans verbe auxiliaire, toutes les phrases contenant un faux ami lexical donné, etc.), facilitant ainsi l'identification de ces schémas.

Enfin, afin d'illustrer concrètement nos résultats, nous avons sélectionné quelques extraits significatifs des copies des élèves (après OCR et vérification) mettant en évidence les erreurs typiques observées. Ces exemples, reproduits tels qu'ils apparaissent avec les erreurs originales, sont présentés dans la section des résultats pour chaque catégorie de difficulté. Ils nous permettent d'ancrer l'analyse dans la réalité du travail des apprenants, tout en préservant leur anonymat (aucune information personnelle identifiable n'apparaît dans les extraits).

3. Résultats

3.1. Principales catégories de difficultés identifiées

L'analyse de notre corpus révèle que les erreurs des apprenants se concentrent principalement dans cinq domaines linguistiques : l'orthographe, la morphologie (conjugaison et accord grammatical), la syntaxe des phrases, le vocabulaire et le discours (cohérence textuelle). Les trois premières catégories représentent à elles seules la grande majorité des erreurs identifiées (environ 85 % du total). Les erreurs purement lexicales sont moins nombreuses (environ 10 %), et les erreurs liées à la cohérence du discours, bien que présentes chez certains apprenants, ne constituent qu'une part modeste d'environ 5 % du total. Les graphiques ci-dessous illustrent la répartition des erreurs par catégorie dans l'ensemble du corpus étudié.

Tableau 1. Répartition des erreurs par catégorie dans le corpus analysé (N = 600).

Catégorie d'erreurs	Nombre d'erreurs (N)	Pourcentage
Orthographe	240	40
Morphologie (conjugaison et accord)	180	30
Syntaxe	90	15
Glossaire	60	10
Discours (cohérence textuelle)	30	5

Figure 1. Répartition (%) des erreurs par catégorie dans le corpus analysé (N = 50 textes, N ≈ 600 erreurs).

Source : données de l'auteur (corpus analysé)

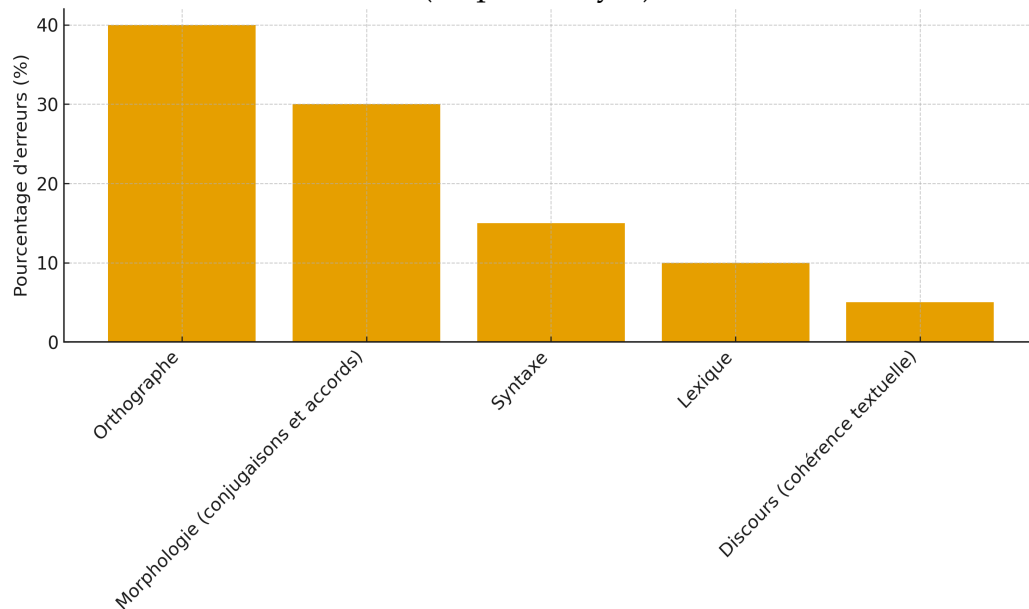
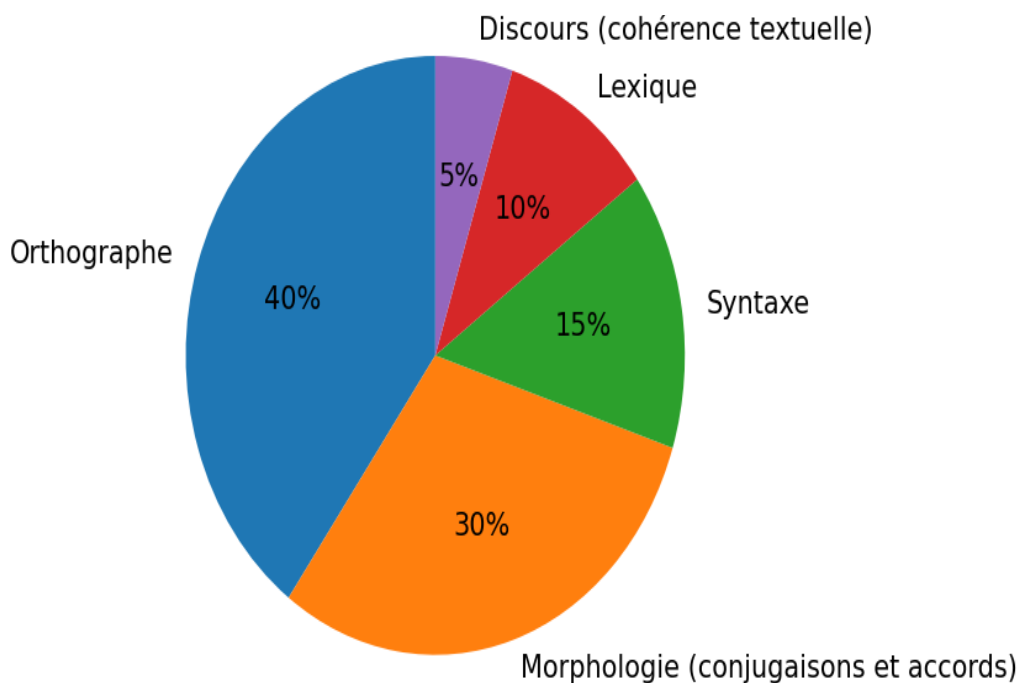


Figure 2. Diagramme circulaire montrant les proportions d'erreurs par catégorie (N = 600). Source : données de l'auteur (corpus analysé)



3.2. Orthographe

Il s'agit de la catégorie d'erreurs la plus courante, représentant environ 40 % de toutes les erreurs identifiées. Cette catégorie comprend à la fois les erreurs orthographiques lexicales (liées à l'orthographe des mots eux-mêmes) et les erreurs orthographiques grammaticales (liées à l'accord du genre/nombre ou à l'utilisation d'homophones grammaticaux). Par exemple, de nombreux apprenants omettent les lettres muettes à la fin des mots ou confondent des homophones tels que *a / à* ou *et / est*. Un exemple typique tiré du corpus est :

« Je suis aller au magasin hier. » (au lieu de « Je suis allé au magasin hier. »). Ici, l'apprenant a écrit le participe passé *aller* sans le *é* final, révélant soit une confusion entre l'infinitif *aller* et le participe passé *allé*, soit une méconnaissance de la règle d'accord du participe passé avec être, confusion qui apparaît souvent.

Dans un autre extrait, on peut lire : « Il y a beaucoup des erreurs dans mon texte. » – l'apprenant a écrit *beaucoup des* au lieu de **beaucoup d**erreurs, commettant une erreur grammaticale courante (formation incorrecte de l'expression *beaucoup de*). En général, les fautes d'orthographe récurrentes observées concernent principalement : les accents (par exemple, *interêt* au lieu de *intérêt*), les doubles consonnes (par exemple, *apellation* au lieu de *appellation*) et les accords silencieux en nombre ou en genre (par exemple, *les enfant* au lieu de « les enfants »). Ces fautes d'orthographe s'avèrent persistantes chez la plupart des apprenants, ce qui corrobore l'idée que la maîtrise de l'orthographe française nécessite beaucoup de temps et d'exposition (Guyon, 2003).

3.3. Morphologie (conjugaisons et accords)

Le deuxième type de difficulté majeur concerne la morphologie verbale et nominale, en particulier la conjugaison des verbes et l'accord grammatical (accord du sujet et du verbe, accord des adjectifs, accord du participe passé, accord des adjectifs, etc.). Cette catégorie représente environ 30 % des erreurs identifiées. De nombreux apprenants commettent des erreurs dans les temps verbaux (par exemple, en utilisant systématiquement l'infinitif au lieu du passé dans un récit au passé) et dans la forme des verbes (personne incorrecte ou terminaison incorrecte). Un exemple tiré d'une épreuve d'examen illustre bien ces problèmes :

« Hier, il prendre le bus et descendre à la mauvaise station. » (au lieu de « Hier, il a pris le bus et est descendu à la mauvaise station. »).

Cette phrase contient deux erreurs morphosyntaxiques : l'utilisation des verbes à l'infinitif au lieu du passé composé et l'utilisation des verbes sans l'auxiliaire nécessaire pour former le passé, ce qui donne « *il descendre* » au lieu de *il est descendu* et « *prendre* » au lieu de *il a pris*. Nous observons ici une simplification de la part de l'apprenant, qui utilise l'infinitif pour plus de facilité

et oublie la structure composée du passé. Un autre exemple d'erreur morphologique récurrente est la maîtrise imparfaite de l'accord du participe passé (par exemple, elles sont arrivé au lieu de : elles sont arrivées). Dans l'ensemble, nos données montrent que la plupart des apprenants ont du mal à appliquer correctement toutes les règles complexes de l'accord en français. Plus de 60 % des étudiants de notre échantillon ont encore des difficultés avec les accords grammaticaux avancés (par exemple, l'accord du participe passé), ce qui indique que ce type d'erreurs subtiles persiste largement dans leurs écrits.

3.4. *Syntaxe*

Les erreurs de syntaxe phrastique (au sens strict, c'est-à-dire à l'exclusion des erreurs purement morphologiques) représentent environ 15 % des erreurs identifiées. Il s'agit principalement d'omissions de mots obligatoires (par exemple, oublier la préposition « de » dans des constructions telles que « j'essaie Ø comprendre » au lieu de « j'essaie de comprendre »), d'erreurs dans l'ordre des mots et de calques de la langue maternelle. Par exemple, un extrait de la dissertation d'un élève dont la L1 est le grec contenait la phrase suivante : « Parce qu'il pleuvait, je suis resté à la maison ». L'apprenant a utilisé « parce que » au début de la phrase, copiant une structure de sa langue maternelle, alors qu'en français standard écrit, on s'attendrait à « Comme il pleuvait, je suis resté à la maison ».

Bien que grammaticalement acceptable, cette construction révèle une influence de la L1 dans la construction de la phrase française (interférence ; cf. Odlin, 1989). De manière générale, nous constatons que les apprenants les plus avancés de notre corpus font moins d'erreurs syntaxiques élémentaires (par exemple, ils n'oublient plus les prépositions courantes et respectent mieux l'ordre sujet-verbe), mais que certains schémas incorrects persistent chez d'autres, en particulier lorsque la construction française diffère sensiblement de celle de la L1. En d'autres termes, dès que la formulation française requise s'écarte des habitudes syntaxiques de leur langue maternelle, certains apprenants continuent à produire des structures incorrectes, démontrant une difficulté à inhiber l'influence de la L1 dans ces contextes.

3.5. *Lexique*

Les erreurs lexicales représentent environ 10 % de toutes les erreurs identifiées. Elles se manifestent sous la forme d'un choix de mots inappropriés ou d'une confusion lexicale, souvent due à des faux amis ou à des traductions littérales de la langue maternelle. Ces erreurs purement lexicales sont relativement moins fréquentes, car les apprenants utilisent généralement des mots qu'ils connaissent bien et peuvent éviter les termes incertains (ou consulter

un dictionnaire). Néanmoins, lorsqu'elles se produisent, elles peuvent altérer le sens du texte.

Un exemple typique est l'utilisation d'un mot français calqué sur l'anglais ou une autre L1 sans tenir compte de sa signification réelle en français. Par exemple, un étudiant anglophone (les étudiants grecs étant presque bilingues, apprenant l'anglais dès la maternelle) a écrit : « Je suis confuse avec cette explication » pour exprimer qu'il ne comprenait pas l'explication. En français, être confus/confuse signifie être embarrassé ou désolé, et non être perplexe ; ici, l'apprenant a traduit littéralement l'anglais « I am confused », commettant une erreur lexicale de faux ami. La formulation attendue aurait été, par exemple, « Cette explication me rend perplexe » ou « Je ne comprends pas bien cette explication ».

Un autre exemple d'erreur lexicale due à l'interférence est : « Il cherche pour une solution » (calque de « he is looking for a solution » en anglais). En français, le verbe chercher s'utilise sans préposition dans ce sens : il faut dire « il cherche une solution ». Ici, l'ajout incorrect de « pour » est directement emprunté à la structure anglaise. Ce type d'erreur indique que l'apprenant applique au français des collocations spécifiques à sa langue maternelle. Dans l'ensemble, les erreurs lexicales observées proviennent soit d'une compréhension incomplète des nuances de sens (choix d'un mot inapproprié), soit d'une interférence lexicale de la L1. Bien que moins nombreuses que les erreurs grammaticales, elles soulignent l'importance de travailler le vocabulaire en contexte pour éviter les pièges des faux amis et des traductions littérales.

3.6. *Discours (cohérence textuelle)*

Les problèmes liés au discours et à la cohérence textuelle représentent environ 5 % des difficultés relevées, ce qui en fait la catégorie la moins représentée en termes de quantité. Ces difficultés concernent la structure générale du texte plutôt que des erreurs linguistiques spécifiques, ce qui les rend plus subjectives à identifier. Cependant, elles apparaissent dans les écrits de certains apprenants, généralement sous la forme d'un manque de liens logiques ou d'une organisation maladroite des idées.

Par exemple, un devoir contenait une série de phrases sans liens ni fil conducteur clair : « Je suis allé en voyage en France l'année dernière. J'aime apprendre de nouvelles langues. La pollution est un problème mondial. » Ici, l'élève juxtapose des informations disparates (une expérience personnelle, son goût pour les langues, une déclaration sur l'environnement) sans transition ni explication, ce qui rend le discours décousu. D'autres apprenants ont des difficultés à structurer leur texte de la manière attendue : absence d'introduction ou de conclusion dans un essai argumentatif, idées présentées dans un ordre

confus, utilisation inappropriée ou inexistante de connecteurs logiques (donc, cependant, d'autre part, etc.).

Bien que ces problèmes de cohérence ne soient pas omniprésents dans le corpus, leur présence dans certains cas indique une maîtrise insuffisante des conventions du texte structuré en français. Cela suggère que ces apprenants n'ont pas encore pleinement intégré les principes d'organisation du discours en FLE, un aspect dont l'enseignement de l'écriture doit également tenir compte, au-delà de la simple correction grammaticale. Il convient de noter que cette constatation fait écho aux observations de Carter-Thomas (1999) sur la nécessité d'enseigner aux apprenants comment structurer un texte de manière cohérente afin d'éviter une écriture décousue.

4. Discussion

Les résultats de cette étude mettent en lumière la nature des difficultés persistantes dans la production écrite en FLE et suggèrent des axes d'amélioration des pratiques pédagogiques actuelles, ainsi que des perspectives sur la contribution des technologies d'IA dans ce domaine.

Tout d'abord, la prévalence des erreurs orthographiques et morphosyntaxiques confirme que ces composantes de l'écriture française sont particulièrement difficiles pour les apprenants non natifs, même à un niveau intermédiaire. Cette observation n'est pas nouvelle en soi : de nombreuses études ont déjà souligné que l'orthographe française représente un obstacle majeur, même pour les apprenants avancés (Guyon, 2003). Notre contribution a été de documenter quantitativement cette réalité à l'aide d'un corpus d'apprenants et de montrer que, malgré un certain niveau de maîtrise du français, ces difficultés restent largement d'actualité. En d'autres termes, même après des centaines d'heures d'apprentissage, les apprenants continuent de commettre des erreurs fondamentales d'orthographe et de grammaire. Cela suggère que les approches pédagogiques actuelles, bien qu'efficaces pour assurer des progrès globaux, ne sont pas suffisantes pour éliminer complètement ces erreurs récurrentes.

Ansi, malgré l'enseignement formel de la règle d'accord du participe passé (souvent abordée au niveau B1/B2), une majorité d'apprenants de notre corpus continuent à faire des erreurs dans l'application de cette règle à l'écrit. Ce phénomène peut être le signe d'une surcharge cognitive : face à la complexité de la tâche d'écriture, les apprenants se concentrent sur le contenu à exprimer et négligent le contrôle de certaines formes, ou bien ils n'ont pas encore automatisé ces règles complexes. Il est également possible que les corrections reçues en classe ne soient pas suffisamment précises ou systématiques pour éliminer ces erreurs. En résumé, nos résultats confirment l'idée que certaines difficultés, notamment en orthographe et en morphologie, ont tendance à persister chez les apprenants

en FLE si les occasions de renforcement et de consolidation ne sont pas multipliées.

De plus, l'observation de quelques erreurs syntaxiques et discursives persistantes, bien que dans une moindre mesure, soulève des questions sur la manière d'enseigner la production écrite dans son ensemble. Le fait que certains apprenants produisent encore des textes incohérents ou mal organisés suggère que l'enseignement de la structure du texte et de la cohérence interne mérite une plus grande attention. Souvent, la pédagogie du FLE met l'accent sur la structure correcte des phrases (grammaire, vocabulaire), mais peut-être pas suffisamment sur la manière de structurer un paragraphe ou un texte entier. Cependant, comme le suggère Carter-Thomas (1999), la cohérence textuelle est une compétence à part entière, qui doit être enseignée explicitement afin d'éviter que les apprenants ne produisent qu'une série de phrases juxtaposées. Nos données, bien qu'exploratoires sur ce point, indiquent que pour un petit nombre d'apprenants, c'est la gestion du discours (plan, progression thématique, séquences logiques) qui fait défaut. Cela implique que les enseignants de FLE ne doivent pas seulement corriger les erreurs linguistiques, mais aussi guider les apprenants dans l'articulation de leurs idées (par exemple en travaillant sur l'utilisation de connecteurs logiques, en rédigeant des plans détaillés, etc.).

4.1. Implications pédagogiques

À la lumière de ces résultats, plusieurs approches pédagogiques concrètes peuvent être envisagées pour mieux répondre aux difficultés récurrentes identifiées. Il s'agit, par exemple, de mettre davantage l'accent sur la relecture et la conscience orthographique à partir des niveaux intermédiaires. Encourager systématiquement les apprenants à relire leurs textes à haute voix ou à les échanger pour une correction croisée peut aider à identifier des fautes d'orthographe qui passeraient autrement inaperçues. En outre, proposer des exercices ciblant les erreurs les plus courantes (par exemple, des activités axées sur les homophones grammaticaux, les terminaisons verbales difficiles ou l'accord du participe passé) permettra de renforcer ces points spécifiques. Ces exercices peuvent prendre la forme de jeux, de quiz ou d'ateliers d'écriture corrective afin de maintenir l'intérêt des apprenants.

Il serait également utile de personnaliser davantage les commentaires en fonction du profil d'erreurs de chaque apprenant. Étant donné que chaque élève a son propre « profil d'erreurs » (certains auront plus de difficultés avec l'orthographe, d'autres avec la syntaxe, d'autres encore avec le vocabulaire, etc.), l'enseignant aurait tout intérêt à adapter ses commentaires afin de cibler les faiblesses spécifiques. Par exemple, si un apprenant fait régulièrement des erreurs dans l'accord du participe passé, lui fournir une fiche de révision

personnalisée et des exercices supplémentaires sur ce point spécifique pourrait l'aider à surmonter cette difficulté. De même, un apprenant dont les textes manquent de cohésion pourrait bénéficier d'une session de tutorat axée sur la structuration du discours et l'utilisation de connecteurs.

Par ailleurs, la collaboration entre les professeurs de langues et les informaticiens pourrait donner naissance à de nouveaux outils pédagogiques basés sur l'IA, spécialement conçus pour les apprenants de français langue étrangère. On pourrait par exemple imaginer un module de formation en ligne qui, après avoir analysé les erreurs d'un apprenant, générerait automatiquement des exercices sur mesure pour travailler ses points faibles (un peu comme un « coach » personnalisé en orthographe/grammaire). De tels outils, s'ils étaient développés, offriraient aux apprenants une formation supplémentaire, autonome et ludique sur les aspects du français qui leur posent le plus de difficultés. Nos résultats indiquent précisément quels seraient les contenus prioritaires pour ces outils (orthographe, accord, etc.). L'IA pourrait également être utilisée en classe pour aider les enseignants : par exemple, en traitant rapidement les travaux écrits de tout un groupe, un logiciel d'IA pourrait fournir à l'enseignant une carte des erreurs les plus fréquentes dans la classe, lui permettant ainsi de cibler plus efficacement ses cours de rattrapage.

Conclusion

En conclusion, notre analyse des erreurs des apprenants permet de mieux comprendre les domaines dans lesquels l'apprentissage du français écrit présente des difficultés persistantes. L'orthographe, la conjugaison et l'accord grammatical restent, pour la plupart des apprenants, des obstacles majeurs qui ne peuvent être surmontés qu'à force de pratique intensive et d'un soutien pédagogique soutenu. Comprendre pourquoi et comment ces erreurs persistent – qu'elles soient dues à la complexité inhérente à la langue, à l'influence de la langue maternelle ou à des lacunes dans le processus d'automatisation – est une étape cruciale pour adapter nos méthodes d'enseignement. Il faut imaginer des façons nouvelles d'enseigner l'orthographe (Fayol & Jaffré, 2014). Les approches évoquées, qu'il s'agisse de renforcer la relecture, de cibler les erreurs fréquentes ou d'utiliser des outils numériques innovants, ont toutes le même objectif : apporter aux apprenants le meilleur soutien possible dans leur progression vers la maîtrise du français écrit.

En outre, cette étude souligne l'intérêt de combiner l'expertise pédagogique humaine avec les possibilités offertes par l'IA. En automatisant l'analyse de grands corpus d'erreurs, l'IA s'est révélée être un allié précieux pour identifier des tendances objectives et libérer les enseignants et les chercheurs de tâches fastidieuses d'analyse de données. Néanmoins, le rôle de l'enseignant reste

irremplaçable lorsqu'il s'agit d'interpréter en détail les erreurs et d'élaborer des stratégies pédagogiques appropriées. C'est donc dans la complémentarité des outils numériques et des enseignants que réside une voie prometteuse : l'IA pour détecter et quantifier les difficultés récurrentes, et les enseignants pour concevoir et mettre en œuvre des remèdes pédagogiques sur mesure. En fin de compte, un meilleur ciblage des erreurs récurrentes à l'aide de ces outils devrait contribuer à améliorer l'efficacité de l'enseignement du FLE en permettant de concentrer les efforts sur ce qui entrave réellement l'apprentissage durable du français écrit.

Bibliographie

- Astolfi, J. P. (1997). L'erreur, un outil pour enseigner.
- Blanche-Benveniste, C., & Chervel, A. (1969). *L'orthographe*. FeniXX.
- Carter-Thomas, S. (2000). La cohérence textuelle: pour une nouvelle pédagogie de l'écrit.
- Corder, S. P. (2015). The significance of learners' errors. In *Error analysis* (pp. 19-27). Routledge.
- Council of Europe (2020), Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume, Council of Europe Publishing, Strasbourg, available at www.coe.int/lang-cefr.
- Council of Europe (2001). Council for Cultural Co-operation. Education Committee. Modern Languages Division. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Cuq, J. P. (2003). Dictionnaire de didactique du français langue étrangère et seconde.
- Cuq, J. P., & Gruca, I. (2017). Cours de didactique du français langue étrangère et seconde : 4e édition, Presses Universitaires de Grenoble.
- Fayol, M., & Jaffré, J. P. (2014). *L'orthographe (Que sais-je ?)*, Presses Universitaires de France - PUF, Société, 1^{re} édition,
- Garcia-Debanc, C. (1990). *L'élève et la production d'écrits*. Centre d'analyse syntaxique de l'université de Metz.
- Guyon, O. (2003). *L'orthographe, une construction cognitive et sociale*. J. P. Jaffré, & J. Fijalkow (Eds.). Presses Universitaires du Mirail.
- Odlin, T. (1989). *Language transfer* (Vol. 27). Cambridge: Cambridge University Press.
- Cornaire, C., & Raymond, P. M. (1995). La production écrite, édi. Clé international, Paris, 1999. *Courtillon Janine, élaborer un cours de fle*, Ed. Hachette, Paris.
- Richards, J. C., & Renandya, W. A. (Eds.). (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge university press.

Selinker, L. (1972). Interlanguage. *IRAL: International Review of Applied Linguistics in Language Teaching*, 10, 209-231.
<https://doi.org/10.1515/iral.1972.10.1-4.209>