



HAL
open science

Portrait des erreurs collocationnelles dans les écrits universitaires de quelques étudiants non francophones

Azadeh Pirooz

► To cite this version:

Azadeh Pirooz. Portrait des erreurs collocationnelles dans les écrits universitaires de quelques étudiants non francophones. *LIDIL - Revue de linguistique et de didactique des langues*, 2024, 70, 10.4000/12lma . hal-04782468

HAL Id: hal-04782468

<https://hal.science/hal-04782468v1>

Submitted on 14 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



Lidil

Revue de linguistique et de didactique des langues

70 | 2024

**L'écrit en langue seconde en contexte universitaire :
l'apport des recherches basées sur les corpus
d'apprenants**

Portrait des erreurs collocationnelles dans les écrits universitaires de quelques étudiants non francophones

Portrait of Collocational Errors in the Academic Writings of Some Non-Native Students

Azadeh Pirooz



Édition électronique

URL : <https://journals.openedition.org/lidil/13165>

DOI : 10.4000/12lma

ISSN : 1960-6052

Éditeur

UGA Éditions/Université Grenoble Alpes

Édition imprimée

ISBN : 978-2-37747-509-4

ISSN : 1146-6480

Référence électronique

Azadeh Pirooz, « Portrait des erreurs collocationnelles dans les écrits universitaires de quelques étudiants non francophones », *Lidil* [En ligne], 70 | 2024, mis en ligne le 31 octobre 2024, consulté le 31 octobre 2024. URL : <http://journals.openedition.org/lidil/13165> ; DOI : <https://doi.org/10.4000/12lma>

Ce document a été généré automatiquement le 31 octobre 2024.



Le texte seul est utilisable sous licence CC BY-SA 4.0. Les autres éléments (illustrations, fichiers annexes importés) sont « Tous droits réservés », sauf mention contraire.

Portrait des erreurs collocationnelles dans les écrits universitaires de quelques étudiants non francophones

Portrait of Collocational Errors in the Academic Writings of Some Non-Native Students

Azadeh Pirooz

1. Introduction

- 1 À l'université, l'écrit est l'un des outils de communication et d'évaluation des processus d'apprentissage et d'acquisition de connaissances scientifiques. L'écrit universitaire obéit à des normes précises touchant à la fois le fond et la forme (Rinck, 2010) et exige des compétences particulières. Parmi ces compétences scripturales figurent les collocations, cooccurrences privilégiées de deux unités lexicales (*la base* et le *collocatif*) liées par une relation syntaxique. Les travaux portant sur l'évaluation et la maîtrise des compétences collocationnelles mettent en évidence les véritables difficultés rencontrées par les étudiants dans leurs écrits universitaires, que ce soit au niveau formel, sémantique ou du registre de la langue (Boch & Rinck, 2015 ; Cavalla, 2008).
- 2 Cette étude de corpus se concentre sur les collocations dans les écrits, en excluant l'oral et en se limitant aux compétences linguistiques. Le terme « écrit universitaire » désigne ici les travaux des étudiants, excluant les mémoires et les thèses, considérés comme des écrits de recherche (Pollet, 2014, p. 43). Le corpus comprend des résumés et des synthèses d'articles scientifiques rédigés par un échantillon d'étudiants allophones. Notre objectif est d'évaluer leurs compétences collocationnelles dans les écrits universitaires et d'identifier les difficultés qu'ils rencontrent. Pour ce faire, nous avons

élaboré une grille d'erreurs permettant une identification fine des erreurs collocationnelles.

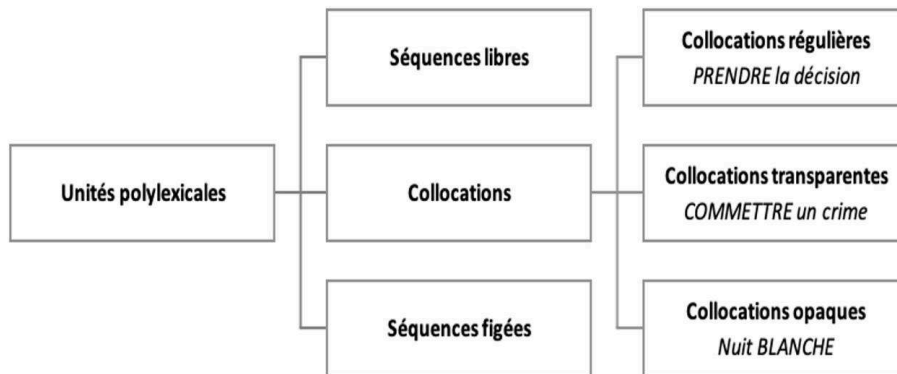
- 3 Cette étude s'inscrit dans le cadre de l'internationalisation de l'enseignement supérieur. Avec une augmentation du nombre d'étudiants non francophones choisissant la France pour leurs études (voir Campus France)¹, il est crucial de prendre en compte les besoins linguistiques des étudiants étrangers afin de mieux les soutenir et de les intégrer dans le système éducatif français.
- 4 Cet article explore les collocations et leurs implications, mettant en lumière leur importance dans les écrits universitaires et les défis qu'elles posent aux étudiants non natifs. La méthodologie donne des informations sur les participants, la collecte de données, la grille d'erreurs, et le traitement des données. Les résultats fournissent un aperçu des erreurs collocationnelles, accompagné d'une analyse des erreurs grammaticales, lexicales et contextuelles. La conclusion résume les principales découvertes de l'étude.

2. Cadre théorique

2.1. Les collocations et les collocations transdisciplinaires : définitions

- 5 Les collocations font partie des unités polylexicales définies comme des « séquence[s] de plusieurs mots ayant une existence autonome » (Gross, 1996, p. 9). Dans cette étude, la définition adoptée de la collocation s'inscrit dans une approche fonctionnelle (Hausmann, 1989 ; Mel'čuk, 1998 ; Tutin & Grossman, 2002). Sémantiquement, une collocation est une unité lexicale compositionnelle, dont le sens résulte de deux constituants. La *base* conserve sa fonction sémantique habituelle, tandis que le *collocatif*, moins prédictible, est sélectionné pour son sens en cooccurrence avec la base (Tutin & Grossmann, 2002, p. 12). Par exemple, dans *poser une question*, le verbe *poser* est collocatif, c'est-à-dire qu'il modifie son sens usuel, tandis que *question* est la base dont le sens usuel reste le même. Selon le degré d'opacité du collocatif, Tutin et Grossmann (*ibid.*) distinguent plusieurs types de collocations : *collocations régulières* comme *partir en voyage*, *collocations transparentes* telles que *commettre un crime* et *collocations opaques* comme *nuit blanche* (voir fig. 1)². Syntaxiquement, les éléments de la collocation entretiennent des relations grammaticales usuelles et sont formés selon les règles syntaxiques habituelles. En ce qui concerne l'usage, les collocations apparaissent fréquemment et sont mémorisées par les locuteurs, contribuant ainsi à la cohésion textuelle et sont souvent associées à un genre textuel particulier.
- 6 Les collocations se situent entre les *séquences libres* et les *séquences figées*. Contrairement aux collocations, une *séquence libre* associe des mots qui maintiennent leur sens usuel. Ce type d'association est compositionnel et transparent sur le plan sémantique, facilitant ainsi la prédictibilité du sens pour les apprenants.

Figure 1. – Trois types de collocations selon Tutin et Grossmann (2002).



- 7 Cette étude se concentre sur les collocations transdisciplinaires, un sous-ensemble du lexique scientifique transdisciplinaire (LST)³. Ces collocations « renvoient [...] non seulement aux procédures, démarches, objets scientifiques, mais aussi aux éléments d’argumentation, d’évaluation et de structuration du discours » (Jacques & Tutin, 2018, p. 6). L’utilisation des collocations transdisciplinaires dans le discours scientifique est cruciale pour assurer la clarté et la cohérence des idées. Elles facilitent l’articulation des arguments, permettant aux auteurs de relier efficacement les concepts et de structurer leur propos de manière logique. Par exemple, des collocations telles qu’*avancer une hypothèse*, *aborder un sujet* clarifient le processus de pensée et guident le lecteur. De plus, l’utilisation des collocations transdisciplinaires influence le registre du discours, établissant un niveau de formalité approprié au contexte académique.
- 8 La spécificité et la transdisciplinarité sont les deux caractéristiques principales des collocations transdisciplinaires. Le critère de spécificité, basé sur la comparaison des fréquences d’utilisation dans différents corpus, permet de contraster les genres et de faire émerger un lexique associé à une pratique et à un type de texte particuliers. La transdisciplinarité signifie que les collocations transdisciplinaires apparaissent dans l’ensemble des disciplines. Elles ne relèvent pas de la terminologie spécifique à un domaine particulier, mais sont utilisées de façon transversale dans les écrits scientifiques (Jacques & Tutin, 2018).
- 9 Les études sur le lexique du discours scientifique confirment que ce type de productions contient des *collocations générales*, des *collocations terminologiques* et des *collocations transdisciplinaires*, selon Mroue (2016, p. 130), qui les décrit ainsi :
- les collocations générales comprennent des combinaisons présentes aussi bien dans la langue générale que dans la langue de spécialité comme *jouer un rôle* ;
 - les collocations terminologiques regroupent des associations propres aux discours spécialisés, comme le discours médical avec *administrer un médicament* ou le discours juridique avec *abroger une loi* ;
 - enfin, le dernier type d’associations concerne les collocations transdisciplinaires comme *valider une hypothèse*, *adopter une approche*, qui sont partagées et comprises par la communauté scientifique à travers diverses disciplines, facilitant ainsi la communication et la collaboration interdisciplinaire.
- 10 Nous étudions ici ce dernier type afin de mettre en lumière les difficultés rencontrées par les étudiants allophones.

2.2. Importance des collocations dans les écrits universitaires

- 11 L'écriture, en tant que vecteur discursif, joue un rôle central dans le travail universitaire, servant d'objet d'apprentissage et de moyen d'acculturation dans l'enseignement supérieur (Scheepers, 2021). Cela est particulièrement vrai pour les étudiants étrangers qui doivent s'appropriier les genres de l'écriture académique et la langue seconde. Cette pratique complexe nécessite des compétences méthodologiques et langagières. Cavalla (2007) définit quatre dimensions essentielles pour les écrits universitaires : « la dimension scientifique (le savoir disciplinaire), la méthodologie (l'architecture), la terminologie (le lexique) et la linguistique (les structures qui contribuent au sens) » (p. 3-4). D'autres éléments, tels que la gestion de la polyphonie discursive et la conduite argumentative, sont également importants. Dans cette étude, nous nous concentrons sur la maîtrise des connaissances linguistiques, en privilégiant les collocations transdisciplinaires plutôt que la terminologie. Pecman (2012) souligne que les étudiants non natifs éprouvent un besoin particulier d'apprendre les expressions du discours scientifique, car la phraséologie spécialisée est souvent plus difficile à maîtriser que la terminologie. Contrairement à cette dernière, qui est spécifique à chaque discipline, les collocations transdisciplinaires forment un lexique commun à toute la communauté scientifique (Jacques & Tutin, 2018). Par exemple, des expressions comme *adopter une stratégie* ou *évoquer la question* sont utilisées dans divers domaines scientifiques, de la biologie aux sciences sociales.
- 12 Les travaux de Cavalla (2010), Tran (2014) et Yan (2017) diagnostiquent les besoins langagiers des étudiants dans l'utilisation du lexique scientifique, soulignant que l'utilisation adéquate de collocations renforce la crédibilité des travaux universitaires. Une maîtrise de ce lexique contribue à établir la légitimité des arguments et à favoriser la réussite académique en facilitant l'intégration dans la communauté scientifique.

2.3. Les erreurs collocationnelles dans les productions des non-natifs

- 13 La définition de l'erreur adoptée est celle de Hamel et Milicévić (2007), englobant les écarts morphosyntaxiques, sémantiques et contextuels par rapport à l'usage d'un locuteur expert. Nous nous inspirons aussi des travaux d'Anctil (2011) sur les maladresses lexicales, qui concernent l'utilisation d'un vocabulaire jugé douteux sans être catégoriquement considéré comme une erreur.
- 14 Les études sur l'acquisition des collocations identifient plusieurs sources d'erreurs. L'influence de la langue première (L1) est une cause majeure, notamment en matière de congruence, c'est-à-dire la présence ou l'absence d'un équivalent littéral en L1 (Nesselhauf, 2005 ; Peters, 2016). Le manque de connaissances sur les propriétés sémantiques et syntaxiques des collocations contribue également aux erreurs, se traduisant par un suremploi, un sous-emploi ou un mésusage de certaines constructions, ainsi qu'une incompréhension du registre attendu dans le contexte approprié (Bolly, 2011). Bolly montre que les apprenants anglophones et néerlandophones de français langue seconde (L2) suremploient des associations libres comme V+N avec deux verbes fréquents (*prendre* et *donner*). Ce suremploi peut résulter

de la similitude avec des équivalents dans la langue première ou de la fréquence dans la langue cible, reflétant le développement de l'interlangue des apprenants.

- 15 Howarth (1998) identifie trois aspects clés de l'utilisation des collocations : *figuratif*, *délexical* et *technique*. Les locuteurs non natifs (LNN) peuvent rencontrer des difficultés avec l'aspect figuratif, à leur tour les collocations délexicales et les termes spécialisés posent également problème en raison d'un manque de connaissances dans des contextes techniques spécifiques.
- 16 Les erreurs syntaxiques des LNN se répartissent généralement en deux catégories : les écarts phraséologiques, souvent dus à une méconnaissance des restrictions morphosyntaxiques, et les erreurs grammaticales, comme la mauvaise formation du pluriel, souvent liées à des interférences de la langue première. Les collocations utilisant des prototypes syntaxiques fréquents sont moins sujettes aux déviations que celles utilisant des prototypes moins courants (Nesselhauf, 2005).

3. Méthodologie

- 17 Pour évaluer les compétences collocationnelles d'un échantillon d'étudiants et identifier les difficultés rencontrées avec les collocations transdisciplinaires, nous avons adopté une approche basée sur l'analyse de productions écrites. Nous présentons ici l'échantillon, la méthode de collecte et de traitement des données, ainsi que la construction d'une grille d'erreurs et l'annotation du corpus.

3.1. Procédure de collecte de données et profil des participants

- 18 Pour la collecte de données, nous avons consulté le « corpus apprenants LIDILEM », élaboré par Tutin et ses collègues (2015), accessible sur la plateforme Alfresco de l'Université Grenoble Alpes. Nous avons sélectionné 30 textes de Master 1 (M1) et Master 2 (M2) en sciences du langage et en didactique des langues, avec 15 textes de chaque niveau, produits entre 2015 et 2020. Les scripteurs, d'origines linguistiques variées, ont un niveau avancé en français, évalué à C1 selon le *Cadre européen commun de référence pour les langues* (CECRL). Il est important de noter que les textes produits par les étudiants en M2 ne proviennent pas des mêmes étudiants ayant produit des textes en M1 l'année précédente.
- 19 Les textes, constitués de synthèses et résumés d'articles scientifiques, visent à répondre aux exigences académiques pour valider leur parcours universitaire. Les étudiants n'ont pas reçu de formation spécifique pour cet exercice. Bien qu'ils aient été rédigés à la maison, nous ne pouvons garantir s'ils ont utilisé des aides (dictionnaires, traducteurs automatiques, relecture). Cependant, nous confirmons que les textes ont été produits avant l'émergence des intelligences artificielles (IA) et ne proviennent pas de ChatGPT par exemple.
- 20 Les étudiants ont suivi la consigne suivante : « Résumez l'article en utilisant vos propres mots tout en préservant les idées principales et les informations essentielles. » Le produit final attendu doit démontrer leur capacité à synthétiser l'information tout en respectant la fidélité des idées et en intégrant cela dans leur propre expression. La longueur des textes varie de 3 à 6 pages et comprend une partie théorique, une méthodologie, ainsi qu'une analyse et des résultats.

3.2. Construction d'une grille d'erreurs

- 21 Parmi les recherches sur l'évaluation des compétences collocationnelles, certaines analysent les erreurs lexicales à l'aide de grilles d'évaluation. Par exemple, Anctil (2011) propose une grille pour analyser les productions en L1, tandis que Yan (2017) travaille sur des grilles adaptées pour la L2. L'approche d'Anctil se concentre sur les erreurs lexicales des francophones québécois, classées en trois catégories : la forme, le sens et la combinatoire restreinte. En revanche, la grille de Yan se spécialise dans les erreurs verbales dans l'écrit scientifique des apprenants sinophones, les classant en quatre catégories : 1) morphologiques, 2) sémantiques et cooccurrences lexicales, 3) grammaticales et syntaxiques, et 4) liées au genre et au contexte.
- 22 À notre connaissance, aucune grille spécifique n'évalue les compétences collocationnelles. Pour pallier cette lacune, nous avons développé une grille d'erreurs inspirée des travaux d'Anctil (2011) et de Yan (2017), axée exclusivement sur les erreurs de collocations. Notre grille s'appuie sur la sous-partie « Combinatoire restreinte » de la grille d'Anctil et sur la section « Cooccurrences lexicales » de la grille de Yan. Elle évalue les erreurs collocationnelles au niveau grammatical, lexical et contextuel (voir le tableau 1).
- 23 La partie grammaticale se divise en deux sous-parties, incluant des erreurs morphologiques et syntaxiques, comme celles liées au genre, au nombre, à la conjugaison, aux prépositions, aux déterminants, à la construction et à l'ordre des mots. La partie lexicale regroupe des erreurs de sens, comprenant trois types : *collocatif inadapté / sens inadapté*, *base inadaptée / sens inadapté*, et *collocatif inadapté / sens adapté*. Enfin, la dernière partie concerne les erreurs contextuelles liées au registre de langue. Nous expliquerons chaque type d'erreur avec des exemples tirés du corpus dans la section des résultats de cet article.

Tableau 1. – Grille des erreurs de collocations inspirée d'Anctil (2011) et Yan (2017).

GRAMMATICALE	MORPHOLOGIQUE	Erreur de conjugaison et d'accord (CA)	On obesrvons les résultats mentionnés (on observe les résultats)
		Erreur d'orthographe (ORT)	Pour étudier les problèmes mentionnés, nous observerons les différences et les solutions utilisées dans les traductions de ces deux langues à partir d'un corpus recueille sur le site internet <i>Opus Corpus</i> . (corpus recueilli)
		Erreur de genre (GEN)	Nous avons construit les tableaux comparatives de système de la flexion nominale du polonais. (tableaux comparatifs)
		Erreur de nombre (NB)	Nous faisons les liens vers les sites d'apprentissage du français (faire le lien)
	SYNTAXIQUE	Erreur de préposition (PRE)	L'ergativité reste donc un sujet complexe comportant des zones d'ombre que les linguistes s'efforcent d'éclairer afin de donner une définition précise et reconnue à ce phénomène linguistique. (reconnue de ce phénomène)
		Erreur d'ordre des mots (ORD)	Faire une description brève d'un élément ou d'une activité. (brève description)
		Erreur de déterminant (DET)	Dans cette étude, nous voulons comparer le système casuel en polonais et ses équivalents français et présenter des difficultés rencontrées (présenter les difficultés)
	Erreur de construction (CON)	Pour l'acquisition guidée, le processus se démarre d'une manière systématique par l'enseignement de langue (démarre d'une manière)	
LEXICALE	SÉMANTIQUE	Collocatif inadapté sens inadapté (CI)	Les théories évoquées ci-dessus sollicitent l'auteur à faire son travail de recherche abordé dans ce texte. (incitent l'auteur)
		Collocatif inadapté sens adapté (AI)	Pendant l'analyse des exemples en français, nous avons remarqué une présence de la construction aspectuelle GN+Vasp+GV dans laquelle on trouve des verbes aspectuels. (observé la présence)
		Base inadaptée sens inadapté (BI)	Nous parlerons ensuite de la rhétorique contrastive avant de souligner des facteurs socio-discursifs du français et du chinois en rapport à la rédaction de textes argumentés (souligner les aspects)
CONTEXTE	PRAGMATIQUE	Emploi d'une unité lexicale appartenant à la langue générale (REG)	La production écrite, soit en L1 soit en L2, est une activité effectivement compliquée (activité complexe)

3.3. Repérage et codage des associations

- 24 D'abord, les associations transdisciplinaires de type V+N sont repérées manuellement dans chaque texte. Pour réduire la subjectivité, nous avons utilisé des références de la norme lexicale, notamment la liste du Lexique scientifique transdisciplinaire (LST) de Jacques et Tutin (2018) et le logiciel Antidote (version 10).
- 25 Ensuite, en nous inspirant de la méthodologie de Nesselhauf (2005), les associations identifiées ont été évaluées par trois enseignants-chercheurs en sciences du langage, locuteurs natifs et spécialistes du lexique disciplinaire. Après que les juges ont identifié et annoté les erreurs, nous avons transféré les données dans Excel pour faciliter l'évaluation et comparer les annotations. Un certain consensus a été constaté entre les annotateurs. Une analyse quantitative a ensuite été réalisée pour calculer les fréquences relatives et la distribution des associations dans chaque groupe, et pour comparer les M1 et M2. Cette analyse visait aussi à identifier les tendances des types d'erreurs entre les deux groupes.

4. Résultats

4.1. L'aperçu général des erreurs collocationnelles

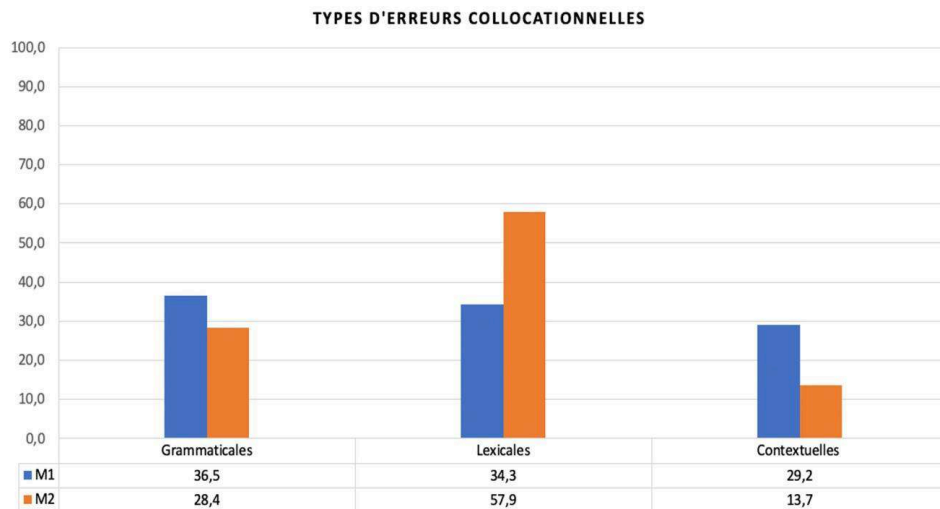
- 26 Le tableau 2 présente un aperçu des associations, des erreurs, des collocations correctes et du pourcentage d'erreurs dans les textes des deux groupes d'étudiants.

Tableau 2. – Récapitulatif des associations et erreurs par groupes d'étudiants.

Groupe	Total associations	Total erreurs	Collocations correctes	Pourcentage d'erreurs
Master 1	558	96	462	17,2 %
Master 2	730	95	645	13,0 %
Total	1 288	191	1 107	14,8 %

27 Parmi les 30 textes examinés, avec un total de 58 860 *tokens*⁴, nous avons identifié 1 288 associations. La répartition montre 558 associations dans les textes de M1 et 730 pour ceux de M2. Parmi ces associations, 191 sont erronées : 96 pour les M1 et 95 pour les M2. La figure 2 détaille la répartition des types d'erreurs (grammaticales, lexicales ou contextuelles). Notons que les textes de M2 présentent un plus grand nombre d'associations V+N, avec un taux d'erreurs légèrement inférieur. Nous allons maintenant vérifier si cette constatation s'applique à toutes les catégories d'erreurs.

Figure 2. – Répartition des types d'erreurs collocationnelles du corpus.

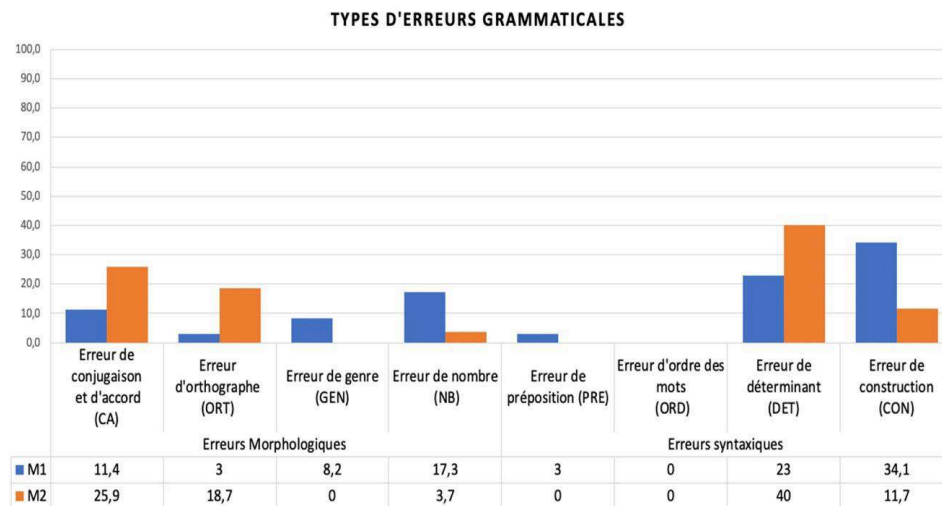


- 28 Les étudiants de M1 affichent un taux d'erreurs grammaticales de 36,5 %, tandis que celui des étudiants de M2 est légèrement inférieur, s'élevant à 28,4 %. En ce qui concerne les erreurs au niveau lexical, celles chez les étudiants de M1 représentent 34,3 % des associations, contre 57,9 % chez les M2. Quant aux erreurs contextuelles, les étudiants de M1 présentent un taux de 29,2 %, tandis que celles de M2 affichent un taux notablement plus bas, à 13,7 %.
- 29 La prochaine section examinera chaque type d'erreur, en commençant par les erreurs grammaticales, suivies des erreurs lexicales, puis des erreurs contextuelles.

4.2. Les erreurs grammaticales

- 30 La partie grammaticale de notre grille se divise en deux sous-parties et évalue les erreurs au niveau *morphologique* et *syntactique*. La figure 3 donne un aperçu de la répartition des erreurs grammaticales dans l'ensemble de notre corpus en fonction de leur appartenance à l'une des huit catégories d'erreurs formelles : les *erreurs de conjugaison* (CA), *d'orthographe* (ORT), *de genre* (GEN), *de nombre* (NB), *de préposition* (PRE), *d'ordre des mots* (ORD), *de déterminant* (DET) et *de construction* (CON).

Figure 3. – Répartition des types d'erreurs grammaticales du corpus.



- 31 L'observation des erreurs grammaticales chez les M1 et M2 révèle des améliorations dans certaines sous-catégories d'erreurs grammaticales et des augmentations dans d'autres. Les erreurs de construction montrent une réduction significative entre le M1 et le M2, indiquant des progrès dans la maîtrise de la complexité syntaxique du français. Cependant, cette progression n'est pas uniforme, car des problèmes persistent chez les M2 : les erreurs de déterminant, de conjugaison et d'orthographe, comme illustré dans les exemples 1 et 2.

(1)

[LNN-M1]⁵ **Nous appuyons** sur l'élément de la langue orale. (nous nous appuyons)⁶

(2)

[LNN-M2] Dans cette étude, nous voulons comparer le système casuel en polonais et ses équivalents français et **présenter des difficultés** rencontrées. (présenter les difficultés)

- 32 Dans l'exemple 1, une erreur de construction est identifiable. Une erreur de construction se produit lorsque la structure syntaxique ou grammaticale d'une phrase est incorrecte, ce qui peut inclure des phrases fragmentées, inversées, des erreurs de concordance de temps ou de voix, des erreurs de subordination ou de coordination, etc. En général, une erreur de construction rend la phrase confuse ou difficile à comprendre pour le lecteur. L'erreur de l'exemple 1 réside dans l'utilisation de la forme réfléchie *nous nous appuyons*.
- 33 Concernant l'exemple 2, une inadéquation dans le choix du déterminant se manifeste. Une erreur de déterminant peut découler d'une utilisation incorrecte, d'une méconnaissance du genre du nom, ou de l'absence d'un déterminant devant un nom.

Dans cet exemple, l'utilisation de l'article indéfini *des* devant le mot *difficultés* est erronée. Pour spécifier les difficultés dont il est question dans le cadre de l'étude, il est plus approprié d'utiliser l'article défini *les* au lieu de *des*. Cela permet de mieux identifier le référent.

- 34 Les sources de ces erreurs peuvent être diverses, comme la complexité des structures de la langue française et l'influence de la L1 des participants (Nesselhauf, 2005).

4.3. Les erreurs lexicales

- 35 Cette catégorie d'erreurs regroupe les erreurs de sens. Trois types d'erreurs de cette catégorie sont identifiées dans le corpus étudié : *collocatif inadapté / sens inadapté* (CI), *collocatif inadapté / sens adapté* (AI) et *base inadaptée / sens inadapté* (BI). Dans ce qui suit, nous expliquons chaque type d'erreur en détail.

4.3.1. Collocatif inadapté / sens inadapté

- 36 Les associations *collocatif inadapté / sens inadapté* (CI) regroupent les erreurs où l'association se produit lorsqu'on utilise un collocatif qui n'est pas couramment utilisé avec la base en question, mais qui a néanmoins un sens qui peut être compris dans le contexte donné. Un locuteur expert dans le même contexte ne produit pas spontanément ce type de combinaisons.

(3)

[LNN-M1] Les théories évoquées ci-dessus **sollicitent l'auteur** à faire son travail de recherche abordé dans ce texte. (incitent l'auteur)

- 37 Dans l'exemple 3, *sollicitent l'auteur* n'est pas incorrect en soi, mais dans ce contexte, le collocatif *inciter* est plus approprié, car ce verbe exprime mieux l'idée de motiver ou d'encourager quelqu'un à entreprendre une action. *Solliciter* signifie principalement demander quelque chose de manière pressante ou formelle, ce qui ne correspond pas vraiment à l'idée de stimuler ou d'encourager l'auteur à mener des recherches.

4.3.2. Collocatif inadapté / sens adapté

- 38 L'association *collocatif inadapté / sens adapté* (AI) se produit lorsque l'on utilise un collocatif qui n'est pas habituellement associé à la base en question, mais qui a cependant un sens qui peut être compris dans le contexte donné.

(4)

[LNN-M2] Les trois **exemples visibles** au-dessus sont les seuls exemples qu'on trouve dans le sous-corpus OpenSubtitles2016. (exemples cités)

(5)

[LNN-M1] Pendant l'analyse des exemples en français, nous avons **remarqué une présence** de la construction aspectuelle GN+Vasp+GV dans laquelle on trouve des verbes aspectuels. (observé la présence)

- 39 L'association *exemple visible* dans l'exemple 4 est classée comme inexistante, car le collocatif *visible* n'est pas associé à la base *exemple*. Cependant, malgré cette inadéquation, le sens demeure compréhensible.

- 40 Nous observons ce type d'erreurs également dans l'exemple 5. Le choix du collocatif *remarqué* est inapproprié avec la base *présence*. Bien que cela constitue une association

inexistante, le sens général de l'association est prévisible, car le collocatif erroné est un synonyme du collocatif approprié.

- 41 Ce type d'erreur, qui témoigne d'une certaine créativité, résulte d'une méconnaissance des collocations appropriées ou d'une compréhension limitée des nuances sémantiques. Cela souligne une difficulté potentielle des étudiants à sélectionner avec précision les collocatifs adéquats en fonction du contexte, influençant ainsi la clarté et la pertinence des associations (Anctil, 2011).

4.3.3. Base inadaptée / sens inadapté

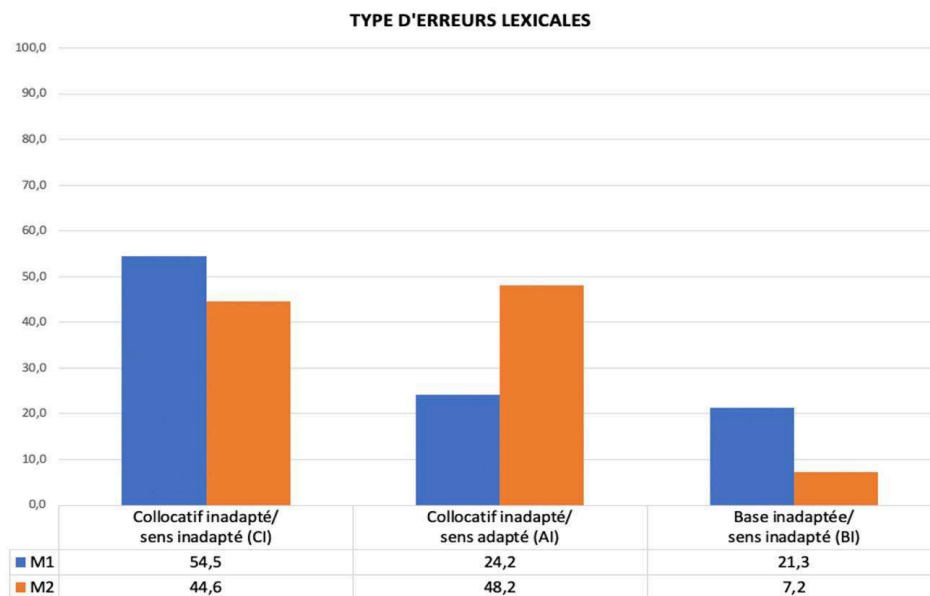
- 42 L'erreur de *base inadaptée / sens inadapté* (BI) se produit lorsqu'une base est associée à un collocatif qui n'est pas approprié pour exprimer une idée ou un concept donné. Dans ce cas, le sens devient inadapté, car la base utilisée ne convient pas à la signification que l'on souhaite transmettre.

(6)

[LNN-M1] Nous parlerons ensuite de la rhétorique contrastive avant de **souligner des facteurs** socio-discursifs du français et du chinois en rapport à la rédaction de textes argumentés (souligner les aspects).

- 43 L'exemple 6 présente une erreur de *base inadaptée / sens inadapté* où la base *facteurs* semble inappropriée pour exprimer l'idée souhaitée. Dans ce contexte, *aspects* est préférable à *facteurs*, car il désigne plus précisément des éléments ou dimensions particulières du phénomène discuté, ce qui est plus adapté pour décrire les subtilités socio-discursives liées à la rédaction.
- 44 La figure 4 présente la répartition de ces types d'erreurs.

Figure 4. – Répartition des types d'erreurs lexicales du corpus.



- 45 Dans la catégorie des erreurs de type CI, on observe une différence importante entre les étudiants en M1 et en M2, avec un taux d'erreurs de 54,5 % pour les M1. Bien que les étudiants en M2 présentent un meilleur score avec 44,6 % d'erreurs, ils rencontrent toujours des problèmes dans le choix des collocatifs.

- 46 Concernant les erreurs de type AI, les étudiants de M1 présentent un taux d'erreur de 24,2 %, tandis que ce taux atteint 48,2 % chez les étudiants de M2. Cela indique qu'ils rencontrent toujours des problèmes dans le choix des collocatifs, ce qui peut engendrer des associations erronées dues à l'abstraction des concepts et à leur maîtrise imparfaite.
- 47 Les erreurs de type BI sont moins fréquentes que les erreurs CI et AI, et nous observons une nette diminution de ce type d'erreurs chez les M2, avec un taux de 7,2 %. Le choix de la base dans les associations semble être une compétence relativement maîtrisée par les deux groupes. En revanche, le choix du collocatif s'avère plus problématique. La variété des collocatifs possibles, ainsi que les nuances de sens, peuvent rendre la sélection délicate. Ce constat est corroboré par Cavalla (2015), qui souligne que le type d'erreur le plus fréquent concerne la structure V+N, avec plus d'erreurs dans le choix du collocatif (le verbe) que dans celui de la base (le nom), en raison des déviations de sens possibles. Par exemple, au lieu d'utiliser la collocation *mener une étude*, un étudiant pourrait écrire *faire une étude*, ce qui manque de précision et de naturel linguistique attendu dans les écrits académiques.

4.4. Les erreurs contextuelles

- 48 L'analyse des erreurs contextuelles liées au registre inapproprié révèle une disparité significative : les M1 totalisent 27 erreurs contextuelles (28,7 %), tandis que les M2 présentent une nette diminution, avec seulement 13 erreurs (13,9 %). Cette réduction observée pourrait refléter une prise de conscience accrue du registre de langue chez les M2.

(7)

[LNN-M1] Les théories évoquées sollicitent l'auteur à **faire son travail de recherche** abordé dans ce texte (réaliser une recherche).

- 49 Dans l'exemple 7, la maladresse réside dans l'utilisation de la collocation *faire son travail de recherche*, qui appartient à un registre de langue plus général et informel. Pour respecter le registre académique, il serait préférable d'utiliser des expressions comme *mener une recherche*, *réaliser une recherche*, qui reflètent une formalité et une précision appropriées.
- 50 Nous avons observé un suremploi de certains verbes à haute fréquence, tels que *faire*, *avoir*, *voir*, *donner* et *trouver*, lorsqu'ils sont associés à des noms. Cette tendance peut s'expliquer par un répertoire lexical restreint et une maîtrise imparfaite de la langue. En d'autres termes, face à un vocabulaire limité et à un manque de confort avec la langue, les étudiants privilégient ces verbes courants. Cela se traduit, comme l'a souligné Kaszubski (2000), par une inclination des étudiants à se tourner vers des *safe options* pour éviter les erreurs linguistiques.

5. Conclusion

- 51 Les résultats de l'analyse des collocations dans les écrits des étudiants de M1 et M2 confirment que les collocations verbo-nominales peuvent constituer des obstacles à la norme native. Trois types de déviations ont été identifiés : la forme, le sens et l'usage.

- 52 Concernant la forme, les erreurs sont liées aux propriétés morphologiques ou syntaxiques des combinaisons lexicales. Bien que ce type d'erreur ait diminué chez les M2, des erreurs persistantes sur les déterminants et la conjugaison demeurent problématiques, indiquant que certains aspects de la grammaire française restent difficiles à maîtriser. La réduction des erreurs s'explique par le fait que les collocations suivent des relations grammaticales courantes, facilitant leur acquisition.
- 53 Concernant le sens, une part importante des erreurs résulte de l'utilisation de collocations erronées, dont le sens se rapproche de celui des unités lexicales visées, souvent sous forme de (quasi-)synonymes (Anctil, 2011 ; Cavalla, 2015). Ces erreurs sémantiques révèlent un déficit métasémantique (Anctil, 2011), qui entrave la précision du discours académique. Nos résultats soulignent l'importance de cette problématique, particulièrement chez les étudiants de M2. Enfin, en ce qui concerne l'usage, les étudiants choisissent parfois des collocations qui, bien que valides dans l'usage général, ne correspondent pas au contexte. Bien que les erreurs contextuelles aient significativement diminué chez les M2, des difficultés persistent dans l'appropriation complète du registre académique.
- 54 En raison de la taille limitée de notre échantillon et du corpus, les résultats ne peuvent pas être généralisés. Notre analyse se base sur les productions de 30 étudiants en sciences du langage et en didactique des langues d'un seul établissement, ce qui restreint nos conclusions.
- 55 Pour soutenir l'apprentissage des collocations transdisciplinaires et surmonter les difficultés rencontrées, il est crucial de sensibiliser les étudiants à l'importance de repérer et comprendre ces structures semi-figées (Cavalla & Legallois, 2020). Il est également essentiel de leur expliquer le rôle clé des collocations dans une communication académique efficace. L'utilisation d'exemples concrets et d'activités ciblées est nécessaire pour aider les étudiants à comprendre les aspects syntaxiques, sémantiques et contextuels des collocations. Une « exposition fréquente et prolongée » aux collocations est recommandée (Falkert, 2013, p. 158). Impliquer les étudiants dans des activités pratiques comme des discussions dirigées (présentations, débats) et des travaux écrits (résumés, synthèses) favorisent la manipulation active des collocations et renforce la maîtrise du LST.
- 56 Enfin, des recherches futures pourraient comparer les performances des natifs et des non-natifs et mener une analyse qualitative plus approfondie pour mieux comprendre les mécanismes sous-jacents aux erreurs observées.

BIBLIOGRAPHIE

ANCTIL, Dominic. (2011). *L'erreur lexicale au secondaire : analyse d'erreurs lexicales d'élèves de 3^e secondaire et description du rapport à l'erreur lexicale d'enseignement de français* (Thèse de doctorat, Université de Montréal, Montréal). <https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/5077/Anctil_Dominic_2011_these.pdf?sequence=2&isAllowed=y>.

- BOCH, Françoise & RINCK, Fanny. (2015). Anaphores démonstratives dans les écrits d'étudiants de Master : comparaison avec les pratiques expertes. *Linx*, 72, 131-150. <<https://doi.org/10.4000/linx.1631>>.
- BOLLY, Catherine. (2011). *Phraséologie et collocations : approche sur corpus français L1 et L2*. Peter Lang.
- CAVALLA, Cristelle. (2007). Réflexion pour l'aide à l'écrit universitaire auprès des étudiants étrangers entrants en Master et Doctorat. Dans J. Goes & J.-M. Mangiante (dir.), *L'accueil des étudiants étrangers dans les universités francophones* (p. 37-48). Artois Presses Université. <<https://hal.science/hal-00380482/document>>.
- CAVALLA, Cristelle. (2008). Les collocations dans les écrits universitaires : un lexique spécifique pour les apprenants étrangers. Dans O. Bertrand & I. Schaffner (dir.), *Apprendre une langue de spécialité : enjeux culturels et linguistiques* (p. 93-104). Éditions École polytechnique. <<https://hal.science/hal-00397684/document>>.
- CAVALLA, Cristelle. (2010). Propositions didactiques pour l'enseignement d'éléments phraséologiques en FLE. Dans O. Galatanu, M. Pierrard, D. Van Raemdonck, M.-È. Damar, N. Kemps & E. Schoonheere (dir.), *Enseigner les structures langagières en FLE* (p. 147-158). Peter Lang.
- CAVALLA, Cristelle. (2015). Collocations transdisciplinaires dans les écrits de doctorants FLS/FLE. *Linx*, 72, 111-130. <<https://doi.org/10.4000/linx.1624>>.
- CAVALLA, Cristelle & LEGALLOIS, Dominique. (2020). Caractériser et identifier les unités phraséologiques pour leur enseignement. *Action didactique*, 6, 12-30. <<https://hal.archives-ouvertes.fr/hal-03171264>>.
- FALKERT, Anika. (2013). À propos de l'existence d'une compétence phraséologique chez l'apprenant de FLE. Une étude expérimentale. Dans C. Garcia-Debanç, C. Masseron & C. Ronveaux (dir.), *Enseigner le lexique* (p. 153-172). Presses universitaires de Namur.
- GROSS, Gaston. (1996). *Les expressions figées en français : noms composés et autres locutions*. Ophrys.
- HAMEL, Marie-Josée & MILICEVIC, Jasmina. (2007). Analyse d'erreurs lexicales d'apprenants du FLS : démarche empirique pour l'élaboration d'un dictionnaire d'apprentissage. *Canadian Journal of Applied Linguistics*, 10(1), 25-45. <<https://journals.lib.unb.ca/index.php/CJAL/article/view/19733>>.
- HAUSMANN, Franz-Josef. (1989). Le dictionnaire de collocations. Dans F. J. Hausmann et coll. (dir.), *Wörterbücher: Ein internationales Handbuch zur Lexikographie / Dictionaries: An International Encyclopedia of Lexicography* (p. 1010-1019). De Gruyter Mouton.
- JACQUES, Marie-Paule & TUTIN, Agnès. (2018). Le lexique scientifique transdisciplinaire : une introduction. Dans M.-P. Jacques & A. Tutin (dir.), *Lexique transversal et formules discursives des sciences humaines* (p. 1-26). ISTE Éditions.
- KASZUBSKI, Przemyslaw. (2000). *Selected Aspects of Lexicon, Phraseology and Style in the Writing of Polish Advanced Learners of English: A Contrastive, Corpus-Based Approach* (Thèse de doctorat non publiée). Adam Mickiewicz University, Poznań.
- MEL'ČUK, Igor. (1998). Collocations and Lexical Functions. Dans A. P. Cowie (dir.), *Phraseology, Theory, Analysis and Applications* (p. 23-53). Clarendon Press.
- MROUE, Mariam. (2014). *Écrit de recherche universitaire : éléments pour une sensibilisation au positionnement scientifique à travers la phraséologie transdisciplinaire* (Thèse de doctorat, Université Grenoble Alpes, Grenoble). <<https://theses.hal.science/tel-01321057>>.

- NESSELHAUF, Nadja. (2005). *Collocations in a Learner Corpus*. John Benjamins.
- PECMAN, Mojca. (2012). Étude lexicographique et discursive des collocations en vue de leur intégration dans une base de données terminologiques. *The Journal of Specialised Translation*, 18, 113-138. <<https://hal.science/hal-01232654/>>.
- PETERS, Elke. (2016). The Learning Burden of Collocations: The Role of Interlexical and Intralexical Factors. *Language Teaching Research*, 20(1), 113-138. <<https://doi.org/10.1177/1362168814568131>>.
- POLLET, Marie-Christine. (2014). *L'écrit scientifique à l'aune des littéracies universitaires. Approches théoriques et pratiques*. Presses universitaires de Namur.
- RINCK, Fanny. (2010). L'analyse linguistique des enjeux de connaissance dans le discours scientifique. *Revue d'anthropologie des connaissances*, 4(3), 427-450. <<https://www.cairn.info/revue-anthropologie-des-connaissances-2010-3-page-427.htm>>.
- SCHEEPERS, Caroline. (2021). *Former à l'écrit, former par l'écrit dans l'enseignement supérieur*. De Boeck Supérieur.
- TRAN, Thi Thu Hoai. (2014). *Description de la phraséologie transdisciplinaire des écrits scientifiques et réflexions didactiques pour l'enseignement à des étudiants non natifs : application aux marqueurs discursifs* (Thèse de doctorat, Université Grenoble Alpes, Grenoble). <<https://theses.hal.science/tel-01330952>>.
- TUTIN, Agnès & GROSSMANN, Francis. (2002). Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif. *Revue française de linguistique appliquée*, 7, 7-25. <www.cairn.info/revue-francaise-de-linguistique-appliquee-2002-1-page-7.htm>.
- TUTIN, Agnès et coll. (2015). *Corpus apprenants LIDILEM*. Université Grenoble Alpes, Alfresco.
- YAN, Rui. (2017). *Étude des constructions verbales scientifiques dans une perspective didactique : utilisation des corpus dans le diagnostic des besoins langagiers en FLE à l'aide des techniques de TAL* (Thèse de doctorat, Université Grenoble Alpes, Grenoble). <<https://tel.archives-ouvertes.fr/tel-01691923>>.

NOTES

1. <www.campusfrance.org/fr>.
2. Dans la figure 1, les collocatifs sont en majuscules.
3. La liste du lexique scientifique transdisciplinaire (LST), élaborée par Tutin et ses collègues dans le cadre du projet TermITH (2013-2016), est disponible à l'adresse suivante : <<http://lst.demarreshs.fr>>. Voir aussi Jacques et Tutin (2018).
4. Un *token* désigne une unité de texte, généralement un mot.
5. Chaque exemple est suivi, entre crochets, du type de locuteur : Locuteur Non Natif Master 1 (LNN-M1) ou Locuteur Non Natif Master 2 (LNN-M2).
6. L'association erronée est en gras, et la correction proposée est entre parenthèses. Cette correction n'est pas exclusive ; d'autres propositions peuvent aussi être valides.

RÉSUMÉS

Cet article traite les erreurs collocationnelles présentes dans les écrits universitaires des étudiants allophones ayant un niveau avancé en français. En analysant 30 textes (des résumés et des synthèses d'articles) à l'aide d'une grille dédiée aux erreurs de collocations, nous avons identifié et évalué des collocations transdisciplinaires V+N comme *mener une étude*, mettant en lumière les erreurs fréquentes et les tendances observées. Les résultats montrent un processus d'apprentissage en cours. En comparant les productions des étudiants en M1 avec celles des M2, on constate que les erreurs grammaticales diminuent, ce qui révèle une certaine amélioration dans la maîtrise des aspects formels chez les étudiants de M2. En revanche, les erreurs lexicales persistent en M2, ce qui souligne une complexification au niveau lexical. Les erreurs contextuelles se réduisent considérablement, ce qui montre une sensibilisation accrue au registre scientifique, notamment en M2. Cette analyse souligne l'importance d'une acculturation aux collocations transdisciplinaires pour améliorer la compétence linguistique des étudiants dans un contexte académique exigeant.

This article examines the collocational errors present in the academic writings of advanced-level non-native French-speaking students. By analyzing 30 texts (summaries and syntheses of articles) using a grid dedicated to collocation errors, we identified and evaluated transdisciplinary collocations such as *mener une étude*, thereby highlighting common errors and observed trends. The results indicate an ongoing learning process. Comparing the work of M1 students with that of M2 students, we observe a decrease in grammatical errors, revealing an improvement in the mastery of formal aspects among M2 students. However, lexical errors persist in M2, indicating increased complexity at the lexical level. Contextual errors decrease significantly, reflecting greater awareness of the scientific register, particularly in M2. This analysis underscores the importance of familiarizing students with transdisciplinary collocations to enhance their linguistic competence in a demanding academic context.

INDEX

Mots-clés : collocations transdisciplinaires, erreurs collocationnelles, écrits universitaires, étudiants non natifs du français, acculturation

Keywords : transdisciplinary collocations, collocational errors, academic writing, non-native French-speaking students, acculturation

AUTEUR

AZADEH PIROOZ

Univ. Grenoble Alpes, LIDILEM, 38000 Grenoble, France
azadeh.pirooz@univ-grenoble-alpes.fr