

Évaluation des compétences collocationnelles dans le discours académique chez les masterisants natifs et non-natifs

Sous la direction d'Agnès Tutin et de Tatiana Aleksandrova

Azadeh PIROOZ

Doctorante au laboratoire LIDILEM - UGA

Journée « Outillage de corpus pour l'acquisition des langues secondes »

16 décembre 2020

Contexte général

Ma thèse s'inscrit dans la phraséologie transdisciplinaire et en particulière, je m'intéresse aux collocations transdisciplinaire dans l'écrit académique.

- Les collocations transdisciplinaires sont des unités lexicales intermédiaire entre la langue générale et la terminologie. Elles n'appartiennent pas à une discipline précise, mais elles peuvent être utilisées dans les recherches scientifiques (Tutin, 2007)
- Elles sont «partagées par les productions scientifiques [...] un lexique de genre qui renvoie aux procédures, démarches et objets scientifiques » (Jacques et Tutin, 2018, p.6)

Les collocations étudiées sont les :

Collocations de type V-N : *émettre une hypothèse, présenter des résultats*

Collocations de type N-adj : *analyse statistique, étude empirique*

Objectif et question de recherche

L'évaluation des compétences collocationnelles en langue générale montrent que la production de collocations pose problème dans l'apprentissage d'une L2 même à un niveau avancé.



Mais les études existantes sur l'analyse des compétences en collocations transdisciplinaires en langue de spécialité sont rares.



Quels sont les problèmes rencontrés par les masterisants N et NN dans l'utilisation des collocations transdisciplinaires V-N et N-Adj ?



L'écart d'utilisation des collocations transdisciplinaires entre les N et les NN est moins évident que ce que montrent Nesselhauf (2003) et Bolly (2008) en langue générale.

Méthodes utilisées pour recueillir différents types de données :

- **Une méthode par analyse des productions écrites**

- 60 productions écrites d'un public étudiant N et NN en master de Sciences du langage à l'Université Grenoble-Alpes
- Les NN ont un niveau C1 selon le CECRL
- Textes argumentatifs avec trois parties principales : le cadrage théorique, la méthodologie et l'analyse

- **Une méthode par enquête**

- Sélection de 20 collocations de type N+adj et V+N dans la liste de LST¹, développée par Tutin dans le cadre du projet TermITH, selon le critère de fréquence

Collocations de type V+N : ***émettre une hypothèse, présenter des résultats***

Collocations de type N+adj : ***analyse statistique, étude empirique***

- Mise en contexte des collocations sélectionnées

Pour évaluer et annoter les associations transdisciplinaires repérées, en s'inspirant des différents travaux sur les erreurs lexicales comme Anctil (2011), Yan (2017) et Li (2019) en langue générale, une grille d'erreurs a été créée.

¹. <http://lst.demarre-shs.fr/collocations/all>

GRAMMATICAL

- **MORPHOSYNTAXIQUE**

Erreur de conjugaison et d'accord

Erreur d'orthographe

Erreur de genre

Erreur de nombre

- **SYNTAXIQUE**

Erreur de préposition

Erreur d'ordre des mots

Erreur de déterminant

Erreur de construction

LEXICAL

- **SÉMANTIQUE**

Association existante
Collocatif inadapté sens inadapté

Association existante
Base inadaptée
sens inadapté

Association inexistante
Collocatif inadapté mais sens adapté

CONTEXTUEL

- **PRAGMATIQUE**

Emploi d'une unité lexicale apparentant à la langue générale

LEXICALE (LEX)	SÉMANTIQUE (SEM)	Association existante Collocatif inadapté sens inadapté (CI)	La langue ergative est un cas spécial par rapport à la langue accusative (cas <u>particulier</u>) Les pronoms personnels en français se divisent en deux formes principales : les formes disjointes et les formes conjointes (présentent deux formes)
		Association existante Base inadaptée sens inadapté (BI)	L'accusatif est un système simple qui ne marque dans la phrase que le sujet – y compris l'agent et le sujet unique – et l'objet. (<u>cas simple</u>) En français la place du mot dans la phrase indique son rôle . (indique sa <u>fonction</u>)
		Association inexistante Collocatif inadapté mais sens adapté (AI)	Les trois exemples visibles au-dessus sont les seuls exemples qu'on trouve dans le sous-corpus OpenSubtitles2016. (exemples <u>cités</u>) Les formules caractéristiques venues de la littérature orale qui en garantissent la mémorisation (chanson de geste, etc.), mais également une série des syntagmes récurrents avec les fonctions difficilement déTECTABLES (fonction <u>identifiable</u>)
CONTEXTE (CONT)	PRAGMATIQUE (PRA)	Emploi d'une unité lexicale apparentant à la langue générale (REG)	Nos observations seront guidées par une quantité de données trouvée dans le corpus numérique Ruscorpora. (données <u>observées</u>) La production écrite, soit en L1 soit en L2, est une activité effectivement compliquée (activité <u>complexe</u>)

En ce qui concerne

- **le recueil des données**
- **l'annotation des erreurs**
 - l'annotation faite par des locuteurs experts, natives de la communauté scientifique qui possèdent une bonne maîtrise de la langue de spécialité.
 - l'utilisation la grille par les juges pour annoter les productions nous permet d'améliorer et de valider le fonctionnement de la grille.