

# Étude comparative des collocations à partir d'un corpus spécialisé en environnement et d'un corpus général

**Zhiwei HAN**

Xi'an Jiaotong-Liverpool University

## Résumé

Cet article explore la variation diaphasique à travers une analyse comparative des collocations dans deux corpus : l'un spécialisé sur la biodiversité et l'autre représentant l'usage général du français. Nous visons à examiner les similitudes et les disparités entre les collocations contenant 15 unités terminologiques (UT) repérées dans des discours de nature variée. Notre démarche méthodologique s'appuie sur la linguistique de corpus, l'approche lexico-sémantique de la terminologie et la Lexicologie explicative et combinatoire. En comparant les schémas collocationnels des UT dans les deux corpus, nous analysons l'association entre la polysémie et la variation diaphasique des collocations.

## Abstract

*This article explores diaphasic variation through a comparative analysis of collocations in two corpora: one focused on biodiversity and the other representing general usage of French language. The main objective is to scrutinize the semantic similarities and differences among collocations of 15 terminological units (TUs) identified across various discourse types. The methodology draws from corpus linguistics, the lexico-semantic approach to terminology, and Explanatory and Combinatorial Lexicology. By comparing collocational patterns in the two corpora, the study investigates the relationship between polysemy and variation in collocations across diverse discourse.*

## 1. Introduction

La notion de collocation est abordée sous différents angles, que ce soit en langue générale ou en langue de spécialité. En langue générale, cette notion a été introduite par Firth (1957) pour désigner l'association habituelle et récurrente des mots sémantiquement apparentés. Selon les travaux du contextualisme britannique qui sont dans la même lignée, la collocation se définit par la fréquence des groupements lexicaux (Halliday 1966 ; Sinclair 1987, 1991).

La collocation est également envisagée en termes de contraintes qu'elle exerce sur l'utilisation du lexique (Hausmann 1979, 1989 ; Mel'čuk *et al.* 1995). Elle résulte de la cooccurrence privilégiée de deux lexies liées par une affinité syntactico-sémantique : la « base », qui est sémantiquement autonome, maintient son sens au sein de la collocation et influence le choix du « collocatif ». Cette relation sémantique et syntaxique entre la base et le collocatif a fait l'objet d'études poussées en linguistique, notamment par Grossmann et Tutin (2003) et Tutin (2013).

En langue de spécialité, une perspective différente se dessine avec le concept de combinaison lexicale spécialisée (CLS), qui désigne le regroupement lexical faisant l'objet d'études en terminologie (L'Homme, 2000, 2003). Une CLS, envisagée comme un sous-type de collocation, associe un collocatif à un terme de base, dont le sens est défini au sein d'un domaine de spécialité, à savoir « un domaine de la connaissance humaine, souvent associé à une activité socioprofessionnelle » (L'Homme 2020 : 24). De ce fait, les usages des CLS sont envisagés par rapport au contexte d'utilisation et aux locuteurs d'une langue de spécialité, utilisé pour véhiculer des connaissances spécialisées. Un lien clair se dessine ainsi avec le phénomène de la variation linguistique, plus précisément de la variation diaphasique. Cette variabilité souligne comment les locuteurs adaptent leur langage en fonction de divers facteurs tels que leurs interlocuteurs et le sujet de la discussion.

Dans un sens plus large, les collocations comportant un élément polysémique peuvent présenter des variations de comportement selon le contexte discursif dans lequel elles sont utilisées. Un exemple de cette variation se manifeste dans l'utilisation du terme *prédateur*. La collocation *prédateur financier* recourt à une métaphore usuelle dans le langage courant. Elle désigne une personne ou une entreprise qui utilise des méthodes agressives pour s'approprier le contrôle d'une entreprise ou d'un actif financier. Dans ce cas, le terme emprunte la notion de « comportement d'un prédateur animal » pour l'appliquer au contexte financier. En contraste, *prédateur carnivore* se réfère à un animal se nourrissant d'autres animaux, ses proies, dans le contexte de la biodiversité. Dans cette acception, le terme dépeint une réalité biologique. Ces deux exemples illustrent comment un même item lexical peut revêtir des sens différents en fonction du contexte dans lequel il est employé.

Dans cette étude, nous explorons les collocations dans une perspective sémantique et cherchons à établir un lien explicite avec la variation diaphasique

qui se manifeste au niveau lexical. La suite de notre travail s'organise de la façon suivante. Dans la section 2, nous formulons l'objectif et les hypothèses de notre recherche. La section 3 aborde notre cadre théorique, qui se décline en trois parties : la linguistique de corpus (3.1), l'approche lexico-sémantique de la terminologie (3.2) et la Lexicologie explicative et combinatoire (LEC) (Mel'čuk *et al.* 1995) (3.3). Dans la section 4, nous présentons en détail la méthodologie de cette recherche. Guidée par le cadre théorique, notre démarche méthodologique s'est déroulée en trois étapes : la sélection d'un corpus général et la création d'un corpus spécialisé, l'extraction des unités terminologiques (UT) à partir des deux corpus, ainsi que l'identification des collocations et l'analyse comparative des données collectées. La section 5 est consacrée à l'analyse des résultats de notre recherche. Enfin, nous clôturons cette recherche par quelques remarques conclusives dans la section 6.

## 2. Objectif et hypothèses de recherche

Cet article propose une analyse comparative afin d'explorer la variation diaphasique des collocations contenant 15 UT relevées de manière systématique dans deux corpus distincts : un corpus spécialisé portant sur la biodiversité et un corpus général représentant l'usage courant du français. La description détaillée des corpus est fournie dans la section 4.1. À travers une extraction semi-automatique des UT et de leurs collocations, nous cherchons à examiner les similitudes et les disparités sémantiques entre ces UT. Nous explorons ensuite leurs schémas collocationnels dans les deux corpus afin de mettre en évidence un éventuel lien entre la polysémie et le comportement collocationnel des UT dans des contextes communicatifs de nature variée.

Notre étude s'inspire du travail de L'Homme et Azoulay (2020), qui ont effectué une comparaison des collocations de langue anglaise dans un corpus général et un corpus spécialisé dans le domaine de l'environnement. Les auteures ont eu recours à une méthode d'extraction automatique pour relever un échantillon de 15 items lexicaux et des listes de 50 candidats-collocatifs pour chaque item. Leur objectif était d'observer divers aspects des collocations, notamment la polysémie des items lexicaux et les caractéristiques des collocatifs (convergences, rangs et classes sémantiques des collocatifs), dans le but de sensibiliser les lexicographes et les terminologues aux particularités du comportement des collocations dans des corpus de nature diversifiée. Les résultats de leur recherche ont montré que la sélection des collocations pertinentes par les lexicographes et les terminologues peuvent être orientés et influencés par la nature des corpus, entre autres facteurs.

À la lumière des conclusions de L'Homme et Azoulay (2020), nous avons formulé deux hypothèses qui structurent notre propre recherche : 1) la polysémie s'observerait aussi bien au sein d'un même corpus qu'entre différents corpus, mais elle serait plus importante dans le corpus de langue générale ; 2) la polysémie

est associée au comportement combinatoire des UT dans différents contextes discursifs. Cette association peut être observée dans le choix des collocatifs et la structuration des liens sémantiques au sein des collocations, bien que son ampleur puisse varier selon les contextes.

### 3. Cadre conceptuel

Le cadre conceptuel de cette recherche se décline en trois volets qui guident notre démarche méthodologique à travers différentes étapes. La linguistique de corpus nous sert d'assise théorique pour fonder notre analyse de la variation diaphasique sur deux corpus de nature différente (3.1). Nous nous appuyons sur l'approche lexico-sémantique de la terminologie (3.2) pour extraire un échantillon d'UT et leurs collocations à partir des corpus assemblés. Nous examinons et comparons ensuite les caractéristiques sémantiques des collocations, en nous inspirant de la méthode d'analyse proposée par la Lexicologie explicative et combinatoire (LEC) (Mel'čuk *et al.* 1995) (3.3).

#### 3.1. Linguistique de corpus

Notre approche méthodologique repose sur les principes de la linguistique de corpus (Condamines *et al.* 2004; Condamines 2005), pour plusieurs raisons. Premièrement, la linguistique de corpus permet une exploration approfondie des phénomènes linguistiques, dont celui des collocations, dans des contextes d'utilisation authentiques. Cette méthode nous libère des hypothèses ou théories préconçues, puisqu'elle prend comme point de départ la langue telle qu'elle est réellement utilisée. Les travaux de Biber *et al.* (1998) soulignent l'importance de cette approche empirique pour une analyse linguistique rigoureuse.

Deuxièmement, la sélection des textes dans un corpus est guidée par des critères de représentativité, visant à refléter la langue utilisée dans un contexte défini et paramétré. L'analyse de corpus offre ainsi la possibilité d'explorer la variation linguistique à travers différents axes tels que le temps, l'espace et les groupes sociaux, ce qui peut être pertinent pour comprendre l'utilisation d'une langue dans des contextes variés.

En ce qui concerne l'application de la linguistique de corpus à la terminologie, Condamines et Dehaut (2011 : 268-270) soulignent trois éléments qui en démontrent la pertinence : la nature des corpus, la variation linguistique et le lexique.

Définir un corpus consiste à circonscrire un ensemble de textes représentant un usage spécifique de la langue dans un contexte particulier. Cette approche vise à donner une vue d'ensemble des termes dans un domaine de spécialité précis et peut s'avérer pertinente pour une étude empirique ancrée dans les discours spécialisés.

La linguistique de corpus permet également de prendre en compte la variation en terminologie, et ce, à l'aide de l'analyse comparative de différents corpus basés sur des critères extralinguistiques comme les genres textuels, les niveaux d'expertise des locuteurs et les objectifs de la communication. Cette analyse de la variation aide à comprendre comment l'utilisation des termes peut être adaptée à un contexte délimité, ce qui en offre une vision dynamique et contextuelle.

Du point de vue du lexique, la linguistique de corpus permet d'explorer divers phénomènes qui vont très souvent au-delà des mots et concernent des aspects syntaxiques, discursifs et sémantiques. Bien que la linguistique de corpus se focalise initialement sur les mots, ces derniers sont plutôt « considérés comme un accès à la “profondeur” des textes » (Condamines et Dehaut 2011 : 270). L'utilisation d'outils informatiques pour analyser de grands volumes de données textuelles rend cette approche encore plus efficace, en ce qu'elle permet une analyse approfondie des occurrences et des usages des termes dans différents contextes.

En résumé, la linguistique de corpus s'avère pertinente pour l'étude des phénomènes linguistiques dans des contextes d'utilisation authentiques. Cette pertinence justifie notre choix d'utiliser cette méthode pour analyser la variation des collocations en langue générale et en langue de spécialité. La méthode de la linguistique de corpus, couplée aux outils informatiques, nous permet de traiter une vaste collection de textes tout en maintenant la rigueur et la finesse de l'analyse des collocations dans leurs contextes.

### **3.2. Approche lexico-sémantique de la terminologie**

Notre étude du lexique spécialisé repose sur l'approche lexico-sémantique de la terminologie, qui propose un modèle de description terminologique à orientation linguistique. Inspirée de la sémantique lexicale (Cruse 1986 ; Polguère 2016), cette approche définit un terme comme une lexie, soit l'association d'une forme linguistique et d'un sens lexical. Un terme se distingue des autres lexies en ce que son sens relève d'un domaine de connaissances spécialisées. L'Homme (2020) précise pour sa part que les termes constituent un sous-ensemble du lexique d'une langue, ce qui implique que la description des termes doit rendre compte de leurs propriétés linguistiques. Cela a des conséquences méthodologiques sur différentes étapes du travail terminographique, notamment l'identification et l'analyse des termes dans leurs contextes d'utilisation. Ces propositions sont essentielles pour guider notre démarche de description du comportement combinatoire des termes dans les corpus général et spécialisé.

Contrairement à la terminologie classique, qui privilégie une démarche onomasiologique centrée sur le concept, l'approche lexico-sémantique prend le terme comme point de départ. Cette méthode est en adéquation avec notre approche basée sur l'analyse de corpus. De plus, la méthode retenue prend en compte la polysémie possible d'une même forme lexicale où les différents sens

exprimés doivent être décrits de façon distincte. Dans le cadre de la présente étude, nous avons donc appliqué des tests lexico-sémantiques pour vérifier si une UT est monosémique ou polysémique. Nous y reviendrons dans la section 4.3 consacrée à la distinction des sens associés à chaque UT.

En outre, la structuration des données descriptives des termes, inspirée de la LEC, met en évidence leurs relations lexicales. Ainsi, la description des collocations d'un terme rend compte de la relation sémantique qui relie le terme à ses collocatifs, ce qui constitue un autre principe guidant notre analyse comparative des collocations extraites de différents corpus.

L'approche lexico-sémantique est appliquée à la confection de dictionnaires spécialisés, entre autres, le DiCoEnviro (Observatoire de linguistique Sens-Texte, 2024), un dictionnaire multilingue portant sur les termes fondamentaux dans le domaine de l'environnement. Nous nous inspirons largement de cette ressource pour l'analyse sémantique des UT.

En résumé, l'approche lexico-sémantique retenue ici offre un cadre théorique pertinent pour guider le traitement et l'analyse des UT que nous nous proposons d'effectuer.

### 3.3. Lexicologie explicative et combinatoire

La Lexicologie explicative et combinatoire (LEC) (Mel'čuk *et al.* 1995) nous sert d'assise théorique pour explorer les relations sémantiques entre la base et le collocatif d'une collocation.

La LEC est une théorie lexicologique s'inscrivant dans le cadre de la Théorie Sens-Texte (TST) (Žolkovskij et Mel'čuk 1965), une théorie linguistique qui présuppose la correspondance entre le sens et l'ensemble des énoncés paraphrastiques qui expriment ce sens en langue naturelle, à savoir le texte. La TST vise la description de cette correspondance au moyen de modèles formels. La modélisation des phénomènes linguistiques en TST « part d'une représentation sémantique pour construire des arbres syntaxiques à l'aide du lexique » (Mel'čuk *et al.* 1995 : 17), d'où la place centrale occupée par la lexicologie dans cette approche.

La LEC vise la description formelle des propriétés linguistiques du lexique. Elle modélise les relations lexicales au moyen des fonctions lexicales (FL), un outil formel inspiré des fonctions mathématiques (Mel'čuk et Polguère 2021). Les FL représentent deux types de liens sémantiques récurrents en langue naturelle : 1) les liens paradigmatiques, qui associent une lexie à ses dérivés sémantiques, tels que les synonymes et les antonymes et 2) les liens syntagmatiques, qui décrivent les collocations contrôlées par une lexie. Les FL syntagmatiques décrivent les phénomènes collocationnels sur deux axes. Sur le plan sémantique, elles rendent compte du sens exprimé par le collocatif ; sur le plan syntaxique, elles rendent compte du rôle syntaxique que le collocatif joue auprès de la base (Polguère 2016 : 201).

Dans notre recherche, nous n'avons pas eu recours aux FL pour représenter les liens sémantiques des UT, étant donné que notre objectif n'était pas de décrire ces relations, mais de les utiliser comme un paramètre pour comparer les collocations contrôlées des UT dans divers contextes d'utilisation. Cependant, nous avons adopté le principe de regrouper les relations lexicales en liens paradigmatiques et syntagmatiques. Les liens paradigmatiques ont été particulièrement utiles pour distinguer les sens des UT. Il s'agit en fait d'un test lexico-sémantique essentiel pour identifier la polysémie, tel que proposé par l'approche lexico-sémantique de la terminologie (voir section 4.3).

Les liens syntagmatiques, quant à eux, sont au cœur de notre étude des collocations. La manière dont nous avons regroupé les collocatifs d'une UT s'inspire de la LEC. Cette approche propose un modèle de description lexicale axée sur la structuration des liens lexicaux. Le recours à ce modèle nous permet d'analyser comment la variation diaphasique des collocations exerce une influence sur leur structuration et éventuellement leur représentation dans une ressource lexicographique.

## 4. Méthodologie

Nous décrivons dans cette section notre démarche méthodologique qui s'est déployée en trois phases. La section 4.1 présente le choix d'un corpus général et la compilation d'un corpus spécialisé. La section 4.2 explique la méthode d'extraction semi-automatique des UT à partir des deux corpus. Enfin, la section 4.3 se concentre sur l'identification des collocatifs et sur l'analyse comparative des collocations relevées dans chaque corpus.

### 4.1. Création de corpus

La première étape de notre démarche consiste à sélectionner un corpus général (CG) existant et à assembler un corpus spécialisé (CS). Les deux corpus reflètent respectivement l'usage du français langue générale et l'usage du français dans le domaine de la biodiversité<sup>1</sup>. Pour le CG, nous avons eu recours au FRENCH WEB 2020 (FRTENTEN20), accessible via *Sketch Engine* (Kilgarriff *et al.* 2014).

---

1. La biodiversité désigne « l'ensemble des êtres vivants ainsi que les écosystèmes dans lesquels ils vivent (et) les interactions des espèces entre elles et avec leurs milieux » (Office français de la biodiversité s. d. : paragraphe 1). Il s'agit d'un domaine de spécialité qui englobe des connaissances scientifiques, notamment en écologie, biologie et géographie, ainsi que des activités socioprofessionnelles liées à la protection de la faune et de la flore, à la lutte contre le changement climatique et à la gestion des ressources naturelles. Cela permet d'aborder de manière intégrée les défis associés à la conservation et à la durabilité de « la variété de la vie sur Terre et (aux) modèles naturels qu'elle forme » (Nations Unies 2023 : paragraphe 1).

Cette plateforme de gestion et d'analyse de corpus offre des fonctionnalités avancées pour la compilation et l'exploration de corpus, que nous avons exploitées dans les phases ultérieures de notre analyse.

Le corpus FRTENTEN<sub>20</sub> est constitué de textes collectés sur des sites web français. Il contient environ 1,5 milliard de mots et couvre des sujets variés, dont l'actualité, la politique, la culture, la science et la technologie. Ce corpus fait partie de la famille des corpus TenTen, un ensemble de corpus web dont la taille cible est de plus de 10 milliards de mots. *Sketch Engine* donne actuellement accès aux corpus TenTen dans plus de 40 langues (Lexical Computing CZ s.r.o., 2024). Il est important de souligner que le corpus FRTENTEN<sub>20</sub> reflète une diversité de variétés du français, incluant le français européen, canadien et africain. Pour minimiser l'impact de cette variation régionale sur nos résultats nous avons veillé à sélectionner les textes spécialisés dans ces variantes géographiques du français, dans le but de garantir une représentativité équilibrée pour notre étude.

Ensuite, nous avons élaboré notre propre CS portant sur le thème de la biodiversité. Ce corpus contient 30 textes comportant au total 122 953 occurrences. Le comptage des mots a été réalisé à l'aide de *Sketch Engine*.

La différence de taille entre le CS et le CG s'explique par le fait que la langue de spécialité est un sous-ensemble de la langue générale. Utilisée dans des contextes de communication délimitées, la langue de spécialité est associée à un seul domaine précis ou à plusieurs disciplines connexes à un domaine spécifique, tels que l'économie et la finance. En revanche, la langue générale est utilisée par un grand nombre de locuteurs et couvre une variété étendue de sujets et de domaines. Par conséquent, notre corpus spécialisé est relativement restreint en contenu et en diversité de domaines, ce qui explique sa taille considérablement plus petite par rapport au corpus général.

Nous avons défini les critères suivants pour constituer le CS afin d'assurer la représentativité de la documentation assemblée : les thèmes abordés, le type de documents, les destinataires, le niveau de spécialité et la date de parution.

D'abord, les textes assemblés dans notre corpus spécialisé portent sur cinq sous-thèmes de la biodiversité : les écosystèmes, les espèces menacées, la pollution lumineuse, le corridor écologique ainsi que les stratégies et les plans d'action pour la conservation de la biodiversité. Ces sous-thèmes ont été sélectionnés à partir de sources fiables, notamment les sites web d'organisations internationales et régionales engagées dans la préservation de la biodiversité, telles que les Nations Unies, l'Office français de la biodiversité et le Centre de la science de la biodiversité du Québec. Nous nous sommes également appuyée sur les enjeux discutés lors de la 15e Conférence des Parties à la Convention sur la diversité biologique (COP 15). Ces sous-thèmes ont ciblés lors de la construction du CS afin de refléter les enjeux actuels en matière de biodiversité à l'échelle internationale.

Nous avons sélectionné quatre types de documents, à savoir les articles scientifiques, les rapports de recherche, les rapports officiels des organisations œuvrant dans la conservation de la biodiversité et les textes de vulgarisation scientifique.

Ces textes s'adressent à des lectorats ayant des niveaux d'expertise variés. Ils sont produits dans différents contextes de communication. Certains facilitent la communication entre experts, notamment les chercheurs universitaires et les professionnels exerçant dans l'un des sous-domaines présentés ci-dessus. D'autres sont rédigés par des experts du domaine pour transmettre des connaissances spécialisées aux non-experts (Bowker et Pearson 2022). Parmi ces textes, on trouve par exemple un article de vulgarisation scientifique visant à expliquer la pollution lumineuse et à sensibiliser le grand public aux conséquences de cette pollution sur la biodiversité.

Les textes spécialisés sélectionnés proviennent des sources que nous avons consultées lors du choix des sous-thèmes. Tous les textes sont rédigés en français.

La définition de ces critères assure que notre corpus spécialisé est représentatif de différents contextes d'usage et de niveaux de spécialité. La collecte des collocations basée sur ce corpus permet de rendre compte des phénomènes de variation dans des contextes d'usage authentiques.

La date de parution a été prise en compte pour garantir une cohérence temporelle avec notre corpus général. Parmi les types de documents retenus, seuls les articles scientifiques et les rapports de recherche sont susceptibles d'évoluer rapidement en raison des avancées scientifiques. Par conséquent, nous avons ciblé des publications de 2020. Cependant, nous avons également inclus d'autres types de textes datant de 2018 à 2022. Les rapports officiels des organisations dédiées à la conservation de la biodiversité affichent une stabilité notable, reposant sur des concepts bien établis et évoluant à un rythme mesuré, comme l'indiquent nos recherches documentaires sur les sites web des organisations internationales. Par ailleurs, les textes de vulgarisation scientifique ont pour objectif de rendre accessibles au grand public les connaissances fondamentales sur la biodiversité, plutôt que de transmettre des avancées scientifiques de pointe. En somme, bien que le corpus spécialisé couvre une période légèrement plus large que le corpus général, il demeure pertinent pour nos analyses comparatives.

## 4.2. Extraction d'UT à partir des corpus

La deuxième étape de notre méthodologie consiste à extraire 15 UT, définies comme des items lexicaux porteurs de sens en relation avec un domaine de spécialité. Une UT peut être monosémique, correspondant ainsi à un terme unique, comme *biodiversité*, qui résulte de l'association d'une forme lexicale et d'un seul sens spécialisé. À l'inverse, une UT peut être polysémique. Par exemple, *population* véhicule des sens distincts dans les collocations *population*

*active et population endémique*. Une analyse ancrée dans les contextes d'usage de *population endémique* révèle que cette collocation se réfère à une espèce animale ou végétale localisée dans un lieu spécifique. On peut donc en déduire que cette UT a deux sens et correspond ainsi à deux lexies distinctes (voir la section 3.2 pour la définition de lexie), dont l'une est un terme dans le domaine de la biodiversité.

L'extraction d'UT a été réalisée à l'aide de TermoStat (Drouin 2003), un outil d'extraction automatique de termes qui compare un corpus d'analyse, le plus souvent spécialisé, à un corpus de référence non spécialisé. Cet outil est accessible en ligne et prend en charge cinq langues alphabétiques, à savoir le français, l'anglais, l'espagnol, l'italien et le portugais.

Nous avons soumis notre corpus spécialisé à TermoStat Web 3.0 (Drouin 2010). Cet outil reçoit un fichier en entrée et produit une liste de candidats-termes (CT) dont la fréquence est proportionnellement plus élevée dans le corpus d'analyse par rapport au corpus de référence (Drouin 2003). Cette fréquence accrue indique le potentiel du CT pour être reconnu comme un terme du domaine concerné.

Le corpus de référence français intégré dans TermoStat réunit des articles du journal *Le Monde* publiés en 2002, totalisant environ 28 500 000 occurrences. Bien que les textes dans ce corpus soient parus il y a plus de 20 ans, la diversité des sujets qu'ils abordent permet une comparaison efficace pour identifier les termes liés à la biodiversité. Nous reconnaissons que la terminologie du domaine est en évolution constante depuis les 20 dernières années et soulignons la nécessité d'intégrer des sources plus récentes. Néanmoins, l'utilisation de ce corpus de référence demeure pertinente pour établir une base de comparaison en termes de spécificité de CT.

Les CT extraits par TermoStat sont initialement triés par leur score de spécificité, calculé en fonction de leur fréquence dans le corpus spécialisé par rapport au corpus de référence (Figure 1). Plus le score d'un CT est élevé, plus celui-ci est susceptible de correspondre à un terme. Un CT peut être une seule entité graphique (terme simple) ou une combinaison des entités graphiques (terme complexe). Pour notre étude, nous avons configuré TermoStat pour cibler spécifiquement les termes simples nominaux.

Corpus >> merge130 xmzxmw | Aide | Déconnexion

## Résultats

Liste des termes Nuage Statistiques Structuration **Bigrammes**

| Candidat de regroupement | Fréquence | Score (Spécificité) | Variantes orthographiques     | Matrice |
|--------------------------|-----------|---------------------|-------------------------------|---------|
| caribou                  | 958       | 422.92              | caribou<br>caribous           | Nom     |
| coll                     | 749       | 364.08              | coll                          | Nom     |
| biodiversité             | 426       | 250.62              | biodiversité                  | Nom     |
| habitat                  | 466       | 231.47              | habitat<br>habitats           | Nom     |
| écosystème               | 334       | 227.57              | écosystème<br>écosystèmes     | Nom     |
| espèce                   | 612       | 186.13              | espèce<br>espèces             | Nom     |
| perturbation             | 202       | 154.07              | perturbation<br>perturbations | Nom     |
| conservation             | 206       | 140.36              | conservation                  | Nom     |
| uvre                     | 95        | 133                 | uvre                          | Nom     |
| population               | 597       | 110.68              | population<br>populations     | Nom     |
| aire                     | 153       | 107.03              | aire<br>aires                 | Nom     |
| prédation                | 66        | 100.05              | prédation                     | Nom     |
| prédateur                | 105       | 98.54               | prédateur<br>prédateurs       | Nom     |
| dérangement              | 54        | 96.19               | dérangement                   | Nom     |
| connectivité             | 48        | 94.55               | connectivité                  | Nom     |
| faune                    | 117       | 94.27               | faune                         | Nom     |
| restauration             | 194       | 93.44               | restauration<br>restaurations | Nom     |
| diversité                | 236       | 92.75               | diversité                     | Nom     |
| durabilité               | 46        | 79.73               | durabilité                    | Nom     |
| lichen                   | 42        | 79.32               | lichen<br>lichens             | Nom     |
| forêt                    | 206       | 79.16               | forêt<br>forêts               | Nom     |
| herbier                  | 42        | 78.58               | herbier<br>herbiers           | Nom     |
| faon                     | 35        | 77.95               | faon<br>faons                 | Nom     |
| écosystémiques           | 31        | 76.33               | écosystémiques                | Nom     |
| changement               | 308       | 75.44               | changement<br>chanaements     | Nom     |

Figure 1. Liste des candidats-termes (CT) extraits par TermoStat

Une fois la liste des CT nominaux générée, nous avons procédé à une sélection manuelle des UT. Dans cette étude, nous avons écarté les noms à sens prédicatif<sup>2</sup> qui désignent un événement, une action, un état ou une relation (L'Homme 2020 : 70-71 ; Mel'čuk *et al.* 1995 : 76), car nous supposons qu'ils sont moins fréquemment utilisés comme bases dans les collocations par rapport aux noms non-prédicatifs. Un certain nombre de termes prédicatifs qui dénotent une entité, tels que *population* et *prédateur*, ont été retenus, du fait qu'ils peuvent aussi agir comme base. Il convient de souligner que la base d'une collocation est sémantiquement autonome et détermine le choix de son collocatif (Grossmann et Tutin, 2003 ; Tutin, 2013). Par exemple, dans la collocation *restauration d'un*

- Le sens prédicatif d'un item lexical est exprimé par un prédicat sémantique, qui dénote un fait et implique nécessairement des actants sémantiques qui participent à ce fait (L'Homme, 2020, p. 70-71; Mel'čuk *et al.*, 1995, p. 76). Par exemple, le terme *prédation* dénote une action qui fait appel à deux actants sémantiques : un agent qui fait l'action, réalisé par des termes tels que *prédateur*, ainsi qu'un patient qui subit cette action, réalisé par des termes tels que *proie*.

*écosystème*, le terme *écosystème* constitue la base et détermine son interaction avec les collocatifs, tels que *restauration*, qui lui sont associés par une relation sémantique. Par conséquent, il nous semble pertinent d'examiner le comportement combinatoire des UT dans les collocations qu'elles régissent.

Un autre critère de sélection est étroitement lié à notre objectif de recherche, qui consiste à comparer les comportements collocationnels dans le CG et le CS. Dans cette optique, nous avons cherché à sélectionner les données les plus significatives pour une analyse comparative visant à explorer les divergences entre les deux corpus. Pour ce faire, nous avons effectué une analyse contextualisée des CT générés au moyen de l'outil *Word Sketch* dans *Sketch Engine*, afin d'obtenir un aperçu rapide des collocations (voir la section 4.3 pour une présentation détaillée de l'outil). Cette démarche nous a permis d'exclure des CT très spécifiques dont l'utilisation ne varie pas considérablement entre les deux corpus. Par exemple, le CT *caribou* donne lieu à des collocations très similaires dans les deux corpus, ce qui limite sa pertinence en tant qu'item de comparaison dans cette étude. Ce critère a été appliqué d'abord aux 50 premiers CT ayant le score de spécificité le plus élevé. Nous avons ensuite élargi la liste des CT examinés à 100 items en raison du nombre de CT exclus au cours de ce processus ; cela illustre l'ampleur de notre analyse contextualisée pour identifier les UT qui sont statistiquement significative.

À l'issue de ce processus, nous avons extrait 15 UT classés par le score de spécificité. Le statut terminologique de ces UT a été validé à l'aide de ressources terminologiques du domaine abordé. Outre le DiCoEnviro (Observatoire de linguistique Sens-Texte 2024), nous avons également eu recours à la base de données Loterre (Institut de l'Information Scientifique et Technique s. d.). Développée par l'Institut de l'Information Scientifique et Technique (Inist-CNRS), Loterre est une plateforme de diffusion des terminologies scientifiques multidisciplinaires et multilingues. Cet outil donne accès à un thésaurus de biodiversité en français, qui nous a servi de référence pour nous assurer que les termes retenus étaient bien ancrés dans le domaine de biodiversité.

Pour chaque UT retenue, nous avons collecté des données comprenant sa fréquence dans le corpus spécialisé, son score de spécificité et ses variantes orthographiques. Le Tableau 1 illustre la liste des UT lemmatisées et les données d'extraction terminologique. Ces items ont par la suite été analysés pour l'extraction des collocations.

**Tableau 1.** Liste des UT identifiées à l'issue de l'analyse manuelle

|    | UT           | Fréquence | Score (Spécificité) | Variantes orthographiques |
|----|--------------|-----------|---------------------|---------------------------|
| 1  | biodiversité | 426       | 250.62              | biodiversité              |
| 2  | habitat      | 466       | 231.47              | habitat<br>habitats       |
| 3  | écosystème   | 334       | 227.57              | écosystème<br>écosystèmes |
| 4  | espèce       | 612       | 186.13              | espèce<br>espèces         |
| 5  | population   | 597       | 110.68              | population<br>populations |
| 6  | aire         | 153       | 107.03              | aire<br>aires             |
| 7  | prédateur    | 105       | 98.54               | prédateur<br>prédateurs   |
| 8  | faune        | 117       | 94.27               | faune                     |
| 9  | diversité    | 236       | 92.75               | diversité                 |
| 10 | forêt        | 206       | 79.16               | forêt<br>forêts           |
| 11 | impact       | 187       | 64.38               | impact<br>impacts         |
| 12 | eau          | 321       | 60.18               | eau<br>eaux               |
| 13 | ressource    | 176       | 53.86               | ressource<br>ressources   |
| 14 | milieu       | 238       | 53.23               | milieu<br>milieux         |
| 15 | zone         | 225       | 41.74               | zone<br>zones             |

### 4.3. Identification des collocatifs et analyse comparative des collocations

La dernière étape de notre démarche méthodologique consiste à identifier un échantillon de collocations contenant des UT retenues, respectivement dans le corpus général (CG) et le corpus spécialisé (CS), et à analyser de façon comparative le comportement combinatoire des UT dans ces deux corpus.

Pour ce faire, nous avons d'abord élaboré une liste de collocatifs pour chaque UT au moyen d'une méthode d'extraction semi-automatique. Plus précisément, nous avons utilisé deux fonctions de *Sketch Engine* : *Word Sketch* et *Concordance*.

La fonction *Word Sketch* nous a permis d'obtenir un aperçu du comportement combinatoire d'une UT. Cette fonction génère automatiquement une liste de candidats-collocatifs (CC), c'est-à-dire des mots qui apparaissent

fréquemment avec une UT et qui sont susceptibles d'être reliés à celle-ci par une relation syntaxico-sémantique.

Les résultats générés par *Word Sketch* sont classés selon la structure syntaxique des collocations formées. La figure 2 présente un extrait des CC de *biodiversité* relevés dans le CG. Les CC relevés sont regroupés en différentes catégories, parmi lesquelles nous nous intéressons particulièrement aux trois suivantes : 1) les verbes ayant *biodiversité* comme objet, ce qui donne lieu à des collocations telles que *préserver la biodiversité* ; 2) les verbes ayant *biodiversité* comme sujet, ce qui donne des collocations telles que *biodiversité s'érode* ; 3) les modificateurs de *biodiversité* dans les collocations nominales telles que *biodiversité marine*.

La catégorisation des CC selon la configuration syntaxique des collocations facilite l'analyse des relations syntaxico-sémantiques entre la base et ses collocatifs. En outre, les CC regroupés au sein d'une même catégorie peuvent être triés par leur fréquence ou leur score de typicité, mesuré par l'algorithme *logDice*. Un score de typicité élevé indique une forte association entre l'UT recherchée et un CC identifié. Ce critère est utilisé comme paramètre de triage par défaut pour les résultats générés. Les CC extraits sont généralement présentés sous forme de lemmes plutôt que sous forme fléchié. Il est également possible d'afficher des exemples d'usage des collocations, comme illustré dans la Figure 2.

**WORD SKETCH** French Web 2020 (frTenTen20) 🔍 ⓘ [Get more space](#)

biodiversité as noun 249,192x ...

| verbs with "biodiversité" as object                       | verbs with "biodiversité" as subject                       | modifieurs of "biodiversité"   |
|---|--|--|
| <b>préserver</b> 8.4 ...<br>préserver la biodiversité     | <b>éroder</b> 6.9 ...<br>la biodiversité s' érode          | <b>marin</b> 8.5 ...<br>de la biodiversité marine                        |
| <b>cultiver</b> 6.8 ...<br>de la biodiversité cultivée    | <b>effondrer</b> 6.8 ...<br>la biodiversité s' effondre    | <b>floristique</b> 7.1 ...<br>biodiversité floristique et faunistique    |
| <b>menacer</b> 6.7 ...<br>biodiversité menacée            | <b>appauvrir</b> 6.7 ...<br>la biodiversité s' appauvrit   | <b>végétal</b> 6.9 ...<br>de la biodiversité végétale                    |
| <b>abriter</b> 6.7 ...<br>abrite une biodiversité         | <b>régresser</b> 5.9 ...<br>la biodiversité régresse       | <b>aquatique</b> 6.7 ...<br>de la biodiversité aquatique                 |
| <b>favoriser</b> 6.5 ...<br>favoriser la biodiversité     | <b>disparaître</b> 5.6 ...<br>la biodiversité disparaît    | <b>microbien</b> 6.6 ...<br>de la biodiversité microbienne               |
| <b>sauvegarder</b> 6.4 ...<br>sauvegarder la biodiversité | <b>amenuiser</b> 5.2 ...<br>La biodiversité s' amenuise    | <b>ultramarin</b> 6.3 ...<br>de la biodiversité ultramarine              |
| <b>protéger</b> 6.2 ...<br>protéger la biodiversité       | <b>abriter</b> 5.1 ...<br>la biodiversité qu' ils abritent | <b>faunistique</b> 6.2 ...<br>la biodiversité faunistique et floristique |
| <b>appauvrir</b> 6.1 ...<br>appauvrit la biodiversité     | <b>diminuer</b> 5.0 ...<br>la biodiversité diminue         | <b>riche</b> 6.2 ...<br>une riche biodiversité                           |
| <b>restaurer</b> 5.6 ...<br>restaurer la biodiversité     | <b>décliner</b> 4.7 ...<br>la biodiversité décline         | <b>terrestre</b> 6.1 ...<br>de la biodiversité terrestre                 |
| <b>valoriser</b> 5.5 ...<br>et valoriser la biodiversité  | <b>englober</b> 4.7 ...<br>La biodiversité englobe         | <b>animal</b> 5.9 ...<br>la biodiversité animale et végétale             |

Figure 2. Liste des candidats-collocatifs de *biodiversité* relevés à partir du corpus général (CG)

Pour chaque UT, nous avons identifié 30 candidats-collocatifs dans chacun des deux corpus. L'outil *Word Sketch* s'avère utile pour repérer rapidement un ensemble de candidats-collocatifs pour une UT. Cependant, les résultats de l'extraction automatique nécessitent une validation, car *Word Sketch* présente des limites dans la distinction entre les différents sens associés à une même forme lexicale. Par exemple, pour une UT polysémique telle que *population*, la liste des résultats peut inclure des cooccurrents correspondant aux différentes acceptions de cette forme. Comme mentionné précédemment, cela peut entraîner un certain flou dans l'interprétation des résultats, ce qui rend indispensable une validation contextuelle plus poussée.

Partant de ce constat, nous avons procédé à une distinction sémantique des UT en analysant les contextes d'utilisation des collocations dans le CG et le CS, à l'aide de l'outil *Concordance*. Cet outil permet de localiser précisément une UT dans les énoncés où elle apparaît, ce qui facilite ainsi l'analyse sémantique.

Notre démarche de distinction sémantique est guidée par l'approche lexi-co-sémantique de la terminologie (voir section 3.2). Afin de vérifier si une UT est polysémique, nous avons appliqué des tests lexi-co-sémantiques (L'Homme 2020 : 76-80), inspirés de ceux mis au point en sémantique lexicale (Cruse 1986 : 55-57 ; Mel'čuk *et al.* 1995 : 64-68). Prenons l'exemple de l'UT *habitat*. Nous avons relevé un échantillon de collocations à l'aide de *Concordance*, dont *habitat participatif* et *habitat benthique* (Figure 3). Pour déterminer si l'UT conserve le même sens dans ces deux collocations, nous avons fait appel à deux tests lexi-co-sémantiques (L'Homme 2020 : 76-80) présentés ci-dessous.

**CONCORDANCE** French Web 2020 (frTenTen20)

CDL habitat + participatif • 4,781  
0.27 par million tokens • 0.000027%

|   | Left context       | KWIC   | Right context   |
|---|--------------------|--|---|
| 1 | recma.org          | 'Him souhaitant favoriser le développement de projet d'    | habitat participatif et coopératif -> Des groupes d'habitat         |
| 2 | agenda21paspyre... | saine et locale en Béarn, l'usage des logiciels libres, l' | habitat participatif, l'éco-construction, etc. Près de 1300 pers/c  |
| 3 | assemblee-natio... | mobilités », sont insérés les mots : « ou de sociétés d'   | habitat participatif + </> </> l'exercice des fonctions de          |
| 4 | assemblee-natio... | Créer de nouvelles formes d'accès au logement par l'       | habitat participatif -> Article 22 -> Au début du livre l           |
| 5 | assemblee-natio... | re préliminaire ainsi rédigé </> Art. L. 200-1. - L'       | habitat participatif est une démarche citoyenne qui permet à d      |
| 6 | assemblee-natio... | des politiques menées aux niveaux national et local, l'    | habitat participatif favorise la construction et la mise à disposit |
| 7 | assemblee-natio... | tres formes juridiques prévues par la loi, les sociétés d' | habitat participatif peuvent se constituer sous la forme de coo     |
| 8 | assemblee-natio... | dans cette démarche peuvent s'associer en sociétés d'      | habitat participatif, sous réserve, lorsqu'elles se constituent s   |

**CONCORDANCE** French Web 2020 (frTenTen20)

CDL benthique + habitat • 428  
0.02 par million tokens • 0.000001%

|   | Left context        | KWIC  | Right context |
|---|---------------------|---|---------------|
| 1 | ulaval.ca           | benthiques et leurs utilisations dans la cartographie des habitats benthiques de l'estuaire du Saint-Laurent->Étude             |               |
| 2 | metiers-quebec...   | pacts de l'exploitation sur les habitats, modélisation des habitats benthiques sensibles aux activités de pêche, évaluation     |               |
| 3 | dfo-mpo.gc.ca       | long terme pour la protection et la connaissance de ces habitats benthiques importants -> Une stratégie de conser               |               |
| 4 | dfo-mpo.gc.ca       | ons d'augmenter le financement pour la recherche sur l' habitat benthique -> C2 -> Surveiller et évaluer la s                   |               |
| 5 | stremfer.fr         | s au suivi de la biodiversité faunistique et floristique des habitats benthiques côtiers, afin de mettre à disposition des scie |               |
| 6 | science.gc.ca       | tagerie sous-marine pour mieux décrire et monitorer les habitats benthiques et les espèces qui s'y trouvent -> Coo              |               |
| 7 | ouroecean2017.or... | al Laurentien -> Le chenal Laurentien contient des habitats benthiques sensibles et présente l'une des concentrac               |               |

**Figure 3.** Contextes d'utilisation des collocations *habitat participatif* et *habitat benthique* relevés dans le corpus général (CG)

Le premier test de distinction sémantique examine la présence de liens paradigmatiques différentiels : ce critère a pour but de déterminer si des ensembles de liens paradigmatiques différentiels peuvent être dégagés pour des sens distincts véhiculés par une même UT.

Pour identifier des liens paradigmatiques concrets à partir des corpus, nous avons utilisé l'outil *Word Sketch* pour analyser le comportement d'une combinaison lexicale contenant l'UT en question (par exemple, *habitat participatif*) dans divers contextes. Cet outil repère des cooccurents qui peuvent servir d'indicateurs des paradigmes auxquels cette combinaison appartient. En observant les cooccurents suggérés pour *habitat participatif*, nous pouvons repérer *intergénérationnel* et *coopératif* (Figure 4). Afin de confirmer les liens paradigmatiques identifiés, nous avons extrait des contextes d'usage contenant *habitat participatif* et les cooccurents proposés à l'aide de l'outil *Concordance*. Ces extractions ont été effectuées en utilisant des recherches avancées formulées au moyen du *Corpus Query Language* (CQL), un langage de requête utilisé dans *Sketch Engine*.

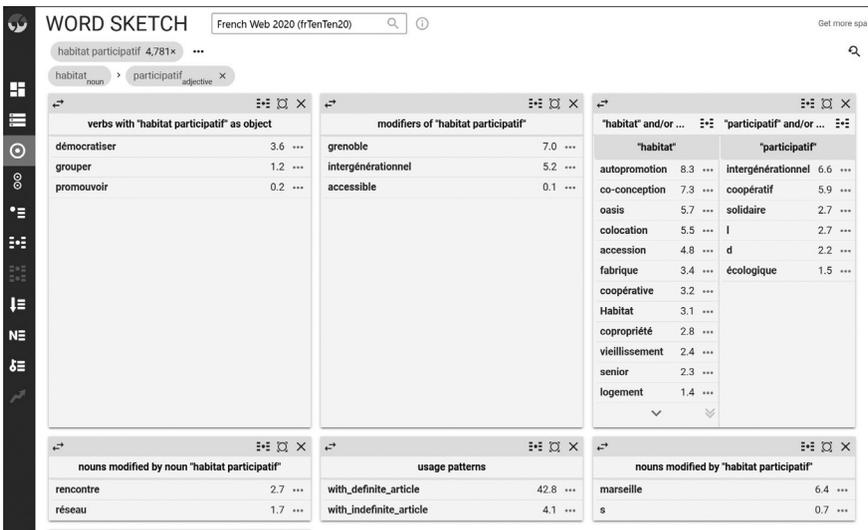


Figure 4. Liste des cooccurents de la combinaison lexicale *habitat participatif* relevés à partir du corpus général (CG)

Deux contextes illustrant l'emploi d'*habitat participatif* ainsi que les liens paradigmatiques extraits du corpus sont présentés dans les exemples (1) et (2) :

- (1) « Elle porte sur les constructions, sur leur programmation variée, mêlant des logements sur le marché libre, des logements en Location-Accession ou à prix maîtrisé, du logement social, du logement étudiant ou encore de l'*habitat participatif* et intergénérationnel. »  
Source : *Qu'est-ce qu'un écoquartier ?* Consulté le 7 octobre 2024, de <https://www.oppidea.fr/oppideainfos/quest-ce-quun-ecoquartier>

- (2) « Pour autant, les projets immobiliers d'*habitat participatif* ou coopératif restent encore marginaux en France. »

Source : *Coliving et coworking : ces nouveaux modèles qui font recette dans l'immobilier – Le Moniteur Emploi*. Consulté le 7 octobre 2024, de <https://emploi-btp.lemoniteur.fr/edito/article/coliving-et-coworking-ces-nouveaux-modeles-qui-font-recette-dans-l-immobilier-aea-11304>

The screenshot shows the CONCORDANCE software interface. At the top, there is a search bar with the query 'CQL [lemma="habitat"] [lemma="participatif"] [lemma="inte...']' and a search button. Below the search bar, there are several tabs: 'Details', 'Left context', 'KWIC', and 'Right context'. The 'KWIC' tab is selected, showing a list of search results. Each result consists of a checkbox, a source URL, and a snippet of text containing the search terms. The results are as follows:

| Source               | Left context   | KWIC   | Right context   |
|----------------------|--|--|---|
| 1 oppidies.fr        | maîtrisé, du logement social, du logement étudiant ou encore de l'     | habitat participatif et intergénérationnel     | /s-s-La programmation variée de logements permet d'accueillir             |
| 2 relie-toits.org    | nbazac (87), les Journées "Habitat et initiatives solidaires" </s>< s> | Habitat participatif, intergénérationnel       | , ou encore adapté, initiatives locales et citoyennes seront passés e     |
| 3 colbris-lafabr...  | veillance et l'empreinte écologique nous animent dans ce projet d'     | habitat participatif et intergénérationnel     | </s>< s>Lieu de vie convivial, créatif, et solidaire de 4 foyers en habit |
| 4 over-blog.fr       | es -> Réf. : 20927 /s-s-La Réole s'engage dans un programme d'         | habitat participatif social intergénérationnel | en cœur de ville – Gironde- 33/s-s>Engagée depuis 2010 dans un            |
| 5 reseau.relier.o... | res, 2002.</s>< s>Ainsi, cette édition vous permettra de découvrir l'  | habitat participatif, intergénérationnel       | , adapté, de repérer des initiatives locales et citoyennes, de rencon     |

Below the list, there is a preview of a document snippet showing the following text:

...  
 fait partie des autres principes d'aménagement d'un écoquartier. </s>< s> Elle porte sur les constructions, sur leur programmation variée, mêlant des logements sur le marché libre, des logements en Location-Accession ou à prix maîtrisé, du logement social, du logement étudiant ou encore de l'habitat participatif et intergénérationnel. </s>< p>< p>< s> La programmation variée de logements permet d'accueillir des habitants dont le profil sociologique et générationnel diffèrent - ce qu'on appelle la mixité sociale - favorise dans la durée les échanges, le lien social et le vivre ensemble, des valeurs et principes clés de tout écoquartier  
 ...

**Figure 5.** Liste des contextes contenant la combinaison lexicale *habitat participatif* et le cooccurrent *intergénérationnel* relevés à partir du corpus général (CG)

L'analyse des contextes d'usage (Figure 5) révèle qu'*habitat participatif* apparaît dans le même paradigme qu'*intergénérationnel* et *coopératif*. Ces résultats indiquent que l'UT utilisée dans cette collocation est liée aux activités humaines. Cela étant, nous avons appliqué la même méthode d'extraction de cooccurrents pour découvrir qu'*habitat benthique* se situe dans un paradigme différent, souvent en relation conceptuelle avec des termes décrivant le vivant non humain, tels qu'*espèces, faune et flore*. Nous en déduisons ainsi qu'*habitat* n'a pas le même sens dans ces deux collocations.

Le deuxième test, qui repose sur un lien paradigmatique spécifique, concerne la substitution par un synonyme. Une UT ne peut être remplacée par un même synonyme si elle véhicule plus d'un sens dans différents contextes. (Mel'čuk *et al.* 1995 : 64-68). L'application de ce test nous permet de confirmer que l'UT *habitat* véhicule deux sens distincts dans les exemples (3) et (4).

- (3) « Passés inaperçus dans la loi ALUR, de nouveaux statuts légaux sont donnés aux logements alternatifs : *habitat participatif*, logements mobiles et constructions démontables [...]. »

Source : *L'habitat alternatif de la loi ALUR*. Consulté le 7 octobre 2024, de <http://verts-region.org/spip.php?article1270>

- (4) « Les objectifs du projet REBENT sont de recueillir et mettre en forme les données relatives au suivi de la biodiversité faunistique et floristique des *habitats* benthiques côtiers, afin de mettre à disposition des scientifiques, des gestionnaires et du public un état des lieux pertinent et cohérent et d'en détecter les évolutions temporelles. »  
 Source : *Flore et faune benthiques*. Consulté le 7 octobre 2024, de [https://envlit.ifremer.fr/index.php/region/basse\\_normandie/qualite/flore\\_et\\_faune\\_benthiques](https://envlit.ifremer.fr/index.php/region/basse_normandie/qualite/flore_et_faune_benthiques)

Dans l'exemple (3), contrairement à l'exemple (4), il est possible de remplacer *habitat* par *logement*.

La dernière étape de notre méthodologie consiste à effectuer une analyse comparative pour mettre en évidence la convergence et la divergence des collocations extraites des deux corpus, ainsi que le regroupement des collocations en fonction des relations sémantiques qui relient les collocatifs à la base.

## 5. Polysémie et collocatif des UT : comparaison des usages dans le corpus spécialisé et dans le corpus de langue générale

Nous présentons, dans cette section, les résultats de nos analyses. La section 5.1 examine la polysémie des UT constatée dans le CG et dans le CS. Nous y mettons en évidence la polysémie des UT à la fois au sein de chaque corpus et entre les deux corpus. La section 5.2 explore le lien entre la polysémie et le choix des collocatifs des UT. Enfin, la section 5.3 analyse les relations sémantiques au sein des collocations relevées. Nous observons ici une association entre la polysémie des UT et leurs schémas collocationnels dans différents contextes discursifs. Ces résultats permettent de répondre à notre problématique de recherche et de valider les hypothèses formulées à la section 2.

### 5.1. Polysémie des UT

En ce qui concerne la polysémie des UT, nous avons constaté que 14 des 15 UT étudiées sont polysémiques et que le nombre de sens présents dans le CG est plus important par rapport au CS, comme l'illustre la Figure 6.

De plus, les sens spécialisés des UT polysémiques que nous avons identifiés dans le CS sont également présents dans le CG. Cela suggère que les termes peuvent subir des variations sémantiques lorsqu'ils sont employés en langue générale, ce qui nous a amenée à établir un lien avec le concept de déterminologisation (*de-terminologization*), telle que décrite par Meyer et Mackintosh (2000). Ce phénomène se manifeste lorsque des termes techniques entrent dans l'usage courant et acquièrent ainsi des sens généraux ou contextuels. Meyer et Mackintosh (2000) examinent les changements sémantiques, grammaticaux et pragmatiques susceptibles de survenir lorsque ces termes sont utilisés par des

non-spécialistes et sur leurs conséquences pour la description terminologique. Nous avons remarqué que certains termes de la biodiversité, tels qu'*écosystème*, subissent également la déterminologisation. Cette UT polysémique correspond à un terme qui désigne un « système dynamique de plantes, d'animaux et d'autres organismes ainsi que des éléments abiotiques du milieu fonctionnant comme une unité indépendante » (Nations Unies s. d. : Section « French Definition ») dans les contextes de la biodiversité, comme l'illustre l'exemple (5) extrait du CS. En langue générale, *écosystème* peut se référer à une « organisation structurée (d'un secteur d'activité par exemple) dans laquelle les différents acteurs (entreprises, fournisseurs, institutions, etc.) sont reliés par un maillage fort leur permettant d'interagir efficacement » (Larousse s. d. b : section « Économie »), illustré par l'exemple (6) extrait du CG. Ce cas particulier de polysémie suggère que la variation diaphasique est associée au phénomène de la déterminologisation.

- (5) « Les écosystèmes sont d'excellents thermomètres des effets du changement climatique, et leur gestion doit prendre en compte les évolutions constatées. »  
Source : *Qu'est-ce que la biodiversité ?* Consulté le 9 octobre 2024, de <https://www.ofb.gouv.fr/quest-ce-que-la-biodiversite>
- (6) « Cette reconnaissance place l'agglomération de Bordeaux et son écosystème entrepreneurial parmi les territoires leaders dans le domaine du numérique et de l'innovation, leviers de croissance pour toute l'économie. »  
Source : *Toute l'actualité de l'environnement normatif - Le blog Enjeux*. Consulté le 9 octobre 2024, de <http://www.enjeux.fr/index.php?d=08&m=12&y=14&category=4>

En somme, la polysémie observée dans les deux corpus, ainsi que la distribution des sens individuels, reflètent une dynamique dans laquelle les termes s'adaptent à des usages variés, ce qui confirme ainsi les observations des recherches antérieures selon lesquelles la terminologie constitue un sous-ensemble du lexique de la langue générale (Cabré 1998 ; L'Homme 2020).

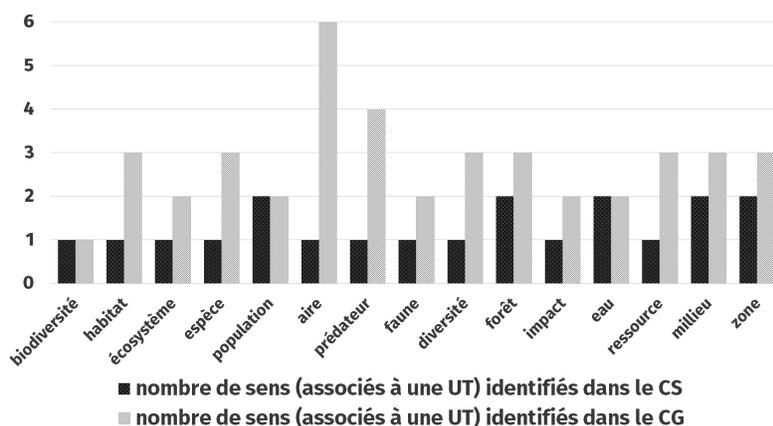


Figure 6. Comparaison inter-corpus de la polysémie des UT

Dans le CG, plusieurs des UT étudiées véhiculent des sens couramment utilisés en langue générale ou dans d'autres domaines de spécialité. Pour l'UT *aire*, à titre d'exemple, nous avons identifié plusieurs sens, tant généraux que spécialisés. Dans le domaine de la biodiversité, *aire* correspond à un terme qui désigne une zone géographique dans laquelle vit une espèce animale ou végétale, ce qui donne lieu à des collocations comme *aire de répartition*, *aire de mise bas* et *aire protégée*. Cette forme lexicale est également associée à d'autres lexies utilisées respectivement dans les domaines de la géométrie, de l'anatomie et du bâtiment. En langue générale, les collocations contenant *aire* apparaissent fréquemment dans des contextes de la vie quotidienne ou de la démographie, comme illustré dans le Tableau 2. Les sens identifiés ont été validés à l'aide du *Dictionnaire de français Larousse* (Larousse s. d. a : section « Définition »). Cette polysémie est liée à la variation des contextes discursifs et, plus précisément, aux thèmes abordés dans les textes.

**Tableau 2.** Polysémie de l'UT *aire* dans les collocations repérées à partir du CG et du CS

|        | Domaine                   | Sens identifié   | Exemple d'usage  |
|--------|---------------------------|--|--|
| Sens 1 | Écologie > biodiversité   | Zone géographique dans laquelle vit une espèce animale ou végétale               | aire de répartition, aire de mise bas, aire protégée                       |
| Sens 2 | Mathématiques > géométrie | Mesure de la surface d'une forme géométrique plane                               | aire d'un triangle, aire hachurée  |
| Sens 3 | Anatomie                  | Région correspondant à un découpage anatomique ou fonctionnel.                   | aire cérébrale, aire corticale, aire ganglionnaire, aire tegmentale        |
| Sens 4 | Bâtiment                  | Espace disponible dans une habitation réservée à une activité précise            | aire de détente, aire de jeu   |
| Sens 5 | Général                   | Surface délimité, aménagée et réservée à une activité, une fonction particulière | aire de pique-nique, aire de stationnement, aire de service, aire de repos |
| Sens 6 | Général                   | Espace occupé par certains êtres humains ou assignés à certains phénomènes       | aire linguistique, aire culturelle   |

Quand il s'agit de la variation sémantique au sein de chaque corpus, nous avons constaté que la polysémie des UT est plus marquée dans le CG par rapport à ce qui est observé dans le CS. Dans le CG, 14 des 15 UT étudiées présentent plus d'un sens, la seule exception étant *biodiversité* (Figure 6). En revanche, dans le CS, 10 des 15 UT (67 %) véhiculent uniquement un sens spécialisé, parmi lesquelles figurent *biodiversité*, *habitat*, *écosystème*, *espèce*, *aire*, *prédateur*, *faune*, *diversité*, *impact* et *ressource*. Seulement deux UT ont été identifiées comme ayant deux sens spécialisés (13 %), à savoir *population* et *forêt*. Enfin, trois UT (20 %), à savoir *aire*, *milieu* et *zone*, possèdent à la fois un sens spécialisé et un sens général dans les contextes d'utilisation en biodiversité. Ces résultats soulignent une différence importante en ce qui concerne la polysémie constatée dans les deux corpus étudiés.

En ce qui concerne la distribution des sens individuels des UT, comme nous l'avons mentionné, tous les sens spécialisés que nous avons identifiés dans le CS sont également présents dans le CG. Autrement dit, aucun des sens liés au domaine de la biodiversité n'est spécifique au CS. En revanche, 12 des 15 UT étudiées (*habitat, écosystème, espèce, aire, prédateur, faune, diversité, forêt, impact, ressource, milieu* et *zone*) présentent des sens exclusifs au CG, en plus de leur sens spécialisé (Figure 6). Ce constat s'explique par la diversité des thèmes abordés dans le CG, qui couvre une gamme étendue de domaines et de sujets. L'UT *forêt*, à titre d'exemple, présente des nuances sémantiques subtiles dans deux collocations que nous avons identifiées dans le CS : *aménager une forêt*<sub>1</sub> et *surexploiter la forêt*<sub>2</sub><sup>3</sup>. Dans la première collocation, *forêt*<sub>1</sub> désigne une étendue de terrain couverte d'arbres, tandis que dans la deuxième collocation, *forêt*<sub>2</sub> fait référence à un ensemble de *forêts*<sub>1</sub> en tant que ressource naturelle. Les deux acceptions sont reliées par une relation sémantique d'holonymie-méronymie. La distinction de ces sens a été validée par les tests de cooccurrence compatible et de cooccurrence différentielle, écrits dans la section 4.3. Dans le CG, *forêt* véhicule un sens figuré dans l'expression *forêt enchantée*, utilisée dans un contexte littéraire pour décrire un lieu magique et mystérieux, caractérisé par des éléments fantastiques, comme l'illustre l'exemple (7) extrait du CG :

- (7) « Or, le magicien Razandar est excédé par les jeux des jeunes dragons, qui s'amuse à cracher du feu dans la forêt enchantée, ce qui pourrait endommager sa superbe tour ! ». Source : *Terres de Jeux, Queen Games*. Consulté le 9 octobre 2024, de <http://www.terres-dejeux.fr/taxonomy/term/429%2C709/3?page=1>

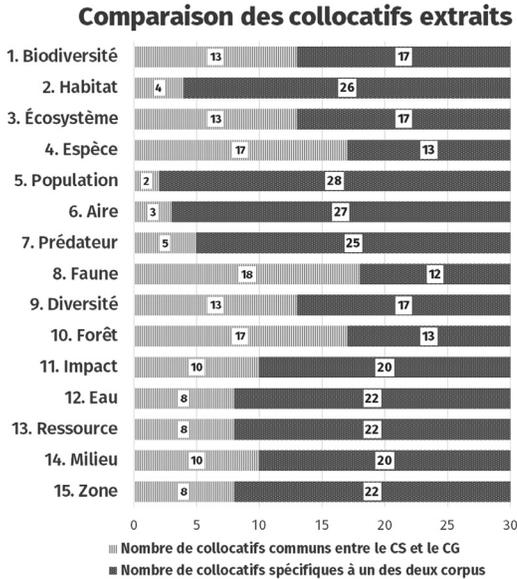
Dans ce cas, *forêt* se réfère à une « grande quantité de choses abstraites formant un ensemble complexe ou confus, inextricable. » (Centre National de Ressources Textuelles et Lexicales 2012 : Section B2).

## 5.2. Sélection des collocatifs pour les UT

La comparaison des listes des collocatifs extraits des deux corpus nous a permis de mettre en évidence les divergences notables du comportement collocationnel des UT dans les deux corpus (Figure 7).

---

3. Dans chaque collocation analysée, l'UT étudiée (*forêt*) est assortie d'un numéro afin de distinguer les différents sens associés à cette forme lexicale.



**Figure 7.** Comparaison des collocatifs communs dans les deux corpus et ceux qui sont uniques à chaque corpus

La figure 7 fait état du nombre de collocatifs des UT apparaissant dans les deux corpus et des collocatifs spécifiques à l'un des corpus. Des UT comme *population*, *aire* et *habitat* sont associées à des collocatifs différents dans les deux corpus, ce qui suggère une variation notable de leur comportement collocationnel.

En revanche, des UT monosémiques dans le CS comme *faune*, *espèce*, *biodiversité*, et *écosystème* partagent un grand nombre de collocatifs dans le CG et le CS ; cela indique une certaine stabilité dans leurs usages à travers divers contextes discursifs. Quant à l'UT *forêt*, bien qu'elle soit polysémique et véhicule un sens figuré qui n'apparaît que dans le CG, le nombre de collocatifs spécifiques à un corpus reste limité. Cette limitation s'explique par le fait que les autres sens associés à *forêt* sont présents dans une grande variété de contextes et donnent lieu à des collocatifs communs aux deux corpus.

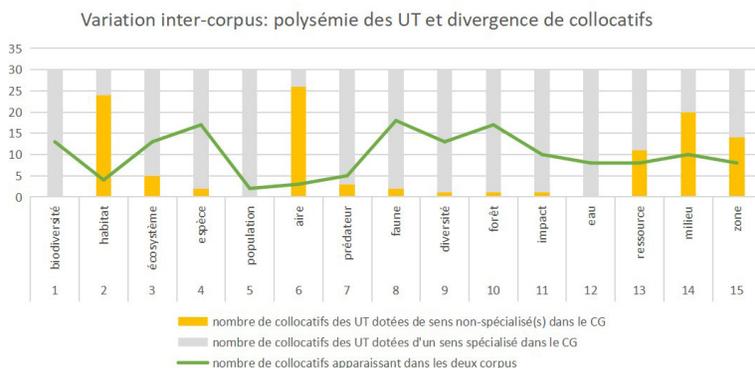
Une analyse approfondie des collocatifs communs révèle un certain chevauchement entre le CG et le CS en ce qui concerne le contexte discursif. Le Tableau 3 montre un extrait de ces collocatifs. Nous avons constaté que ces collocatifs sont principalement utilisés dans les contextes où la biodiversité est abordée, que soit dans une perspective de vulgarisation scientifique ou de sensibilisation du public à l'importance de la conservation de la biodiversité. Dans ces contextes discursifs, les UT étudiées ne montrent guère de variation des sens. La stabilité sémantique de ces UT semble être liée à leur comportement collocationnel similaire dans les contextes relevant de la biodiversité, ce qui est illustré par la convergence des collocatifs observée.

Du point de vue des thèmes abordés, la convergence des collocatifs identifiés révèle également des thèmes récurrents dans le CG et le CS. Par exemple, la conservation, la protection des ressources naturelles et les environnements marins sont des sujets fréquents dans les deux corpus. Cette convergence peut être attribuée à l'importance croissante de la biodiversité dans les discours scientifiques et généraux.

**Tableau 3.** Extrait de la liste des collocatifs communs entre le CS et le CG

| UT          | biodiversité | habitat   | écosystème  | espèce      | population | aire        | prédateur | faune         |
|-------------|--------------|-----------|-------------|-------------|------------|-------------|-----------|---------------|
| Collocatifs | marin        | perturber | terrestre   | menacer     | abriter    | protéger    | naturel   | sauvage       |
|             | communal     | essentiel | restaurer   | exotique    | augmenter  | marin       | marin     | terrestre     |
|             | préservé     | naturel   | dégrader    | envahissant |            | aménagement | chasser   | occuper       |
|             | abriter      | faunique  | sain        | vivre       |            |             | nourrir   | abriter       |
|             | conserver    |           | protéger    | faunique    |            |             | attaquer  | protéger      |
|             | faunistique  |           | marin       | végétal     |            |             |           | peupler       |
|             | nocturne     |           | côtier      | protéger    |            |             |           | préservé      |
|             | forestier    |           | aquatique   | animal      |            |             |           | endémique     |
|             | floristique  |           | naturel     | sauvage     |            |             |           | aquatique     |
|             | menacer      |           | fragile     | aquatique   |            |             |           | sous-marin    |
|             | endémique    |           | intact      | endémique   |            |             |           | riche         |
|             | végétal      |           | forestier   | marin       |            |             |           | ichtyologique |
|             | éroder       |           | littoral    | migrateur   |            |             |           | nocturne      |
|             |              |           | diversifier | peupler     |            |             |           | diversifier   |
|             |              |           |             | survivre    |            |             |           | varier        |
|             |              |           |             | éteindre    |            |             |           | coloniser     |
|             |              |           |             | végétal     |            |             |           | arctique      |
|             |              |           |             |             |            |             | indigène  |               |

En résumé, une polysémie plus marquée des unités terminologiques entre les deux corpus est associée à une divergence accrue de leurs collocatifs. La Figure 8 illustre cette relation en montrant visuellement la connexion entre la polysémie et le choix des collocatifs. Parmi les UT présentant le comportement combinatoire le plus varié, soulignons des UT polysémiques telles qu'*habitat*, *aire*, *milieu* et *zone*. Elles se distinguent par une divergence notable des collocatifs entre les corpus CG et CS.



**Figure 8.** Relation entre la polysémie des UT et la divergence de collocatifs dans le CG et le CS

Nous avons identifié d'autres facteurs qui contribuent à la variation diaphasique des collocations entre le CG et le CS, notamment la multidimensionnalité des termes (Bowker 1993). Cette notion permet de représenter le concept désigné par un terme sous divers angles dans un cadre de recherche terminologique thématique. En effet, un concept peut être classifié et organisé de manière différente dans divers domaines de connaissances selon les caractéristiques qui le distinguent des autres concepts dans un contexte d'utilisation. Le terme *biodiversité* en est un exemple. Il présente des comportements collocationnels différents dans le CS et le CG. Dans le cadre d'une description touristique, les collocatifs associés à *biodiversité* peuvent être spécifiques à une région, tels que *pyrénéenne* ou *réunionnaise*. En revanche, dans le contexte de la recherche scientifique et de la planification d'actions, les collocatifs se rapportent souvent à la dégradation de la biodiversité et à des actions de restauration, tels que *s'appauvrir* ou *restaurer la biodiversité*.

### 5.3. Analyse des relations sémantiques au sein des collocations relevées

Nous avons examiné les relations sémantiques au sein des collocations relevées, afin de valider notre deuxième hypothèse, selon laquelle la polysémie est associée au comportement combinatoire des UT dans différents contextes discursifs, notamment le choix des collocatifs et la structuration des liens sémantiques au sein des collocations.

Inspirée par le DiCoEnviro (Observatoire de linguistique Sens-Texte 2024), notre analyse a révélé que la variation diaphasique d'une UT peut influencer la classification de ses collocations selon les relations sémantiques. En effet, cette ressource terminologique traite uniquement le lexique de l'environnement. Pour une UT polysémique telle qu'*habitat*, le DiCoEnviro ne présente que la description du sens propre au domaine abordé. Les autres sens d'*habitat* relèvent de l'usage du français langue générale et ne font pas l'objet de la description terminologique dans cette ressource. De ce fait, nous avons eu recours à la méthodologie proposée par le DiCoEnviro pour structurer les relations sémantiques entre l'UT examiné et ses collocatifs, plutôt qu'à la description terminologique.

Reprenons l'exemple de l'UT *habitat*. Elle revêt trois sens selon notre analyse (voir aussi la section 4.3). La figure 9 illustre comment les collocatifs combinés à ces trois acceptions sont regroupés.

La première acception d'*habitat*, qui se rapporte à un environnement spécifique où vivent des êtres vivants, est associée à deux catégories de collocatifs : ceux qui dénotent des lieux spécifiques, tels qu'*habitat aquatique* ou *forestier*, et ceux qui décrivent des types spécifiques d'occupants, comme *faunistique* ou *benthique* dans des contextes spécialisés.

La deuxième acception de l'UT est liée à l'organisation et à la colonisation de l'environnement par les êtres humains. Elle se combine avec deux types de

collocatifs : ceux qui décrivent des lieux spécifiques, comme *urbain* ou *rural*, et ceux qui sont liés à des types spécifiques d'occupants, comme *individuel* ou *intergénérationnel* dans des contextes spécialisés ou généraux.

Enfin, la troisième acception de l'UT concerne les conditions de logement. Elle se combine avec des collocatifs tels que *propice* et *insalubre*, qui sont usuels dans des contextes généraux. Ces collocations sont classées en fonction de leur pertinence pour la vie humaine.

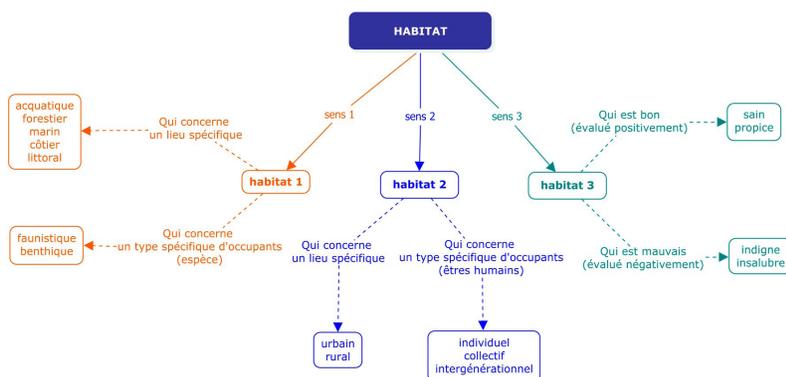


Figure 9. Regroupement des collocatifs associés aux trois acceptions d'*habitat*

## 6. Conclusion

Cette recherche a permis d'explorer la polysémie dans les corpus général (CG) et spécialisé (CS), et de mettre en lumière le lien entre cette polysémie et le comportement collocationnel des UT dans différents contextes. Les résultats montrent que la polysémie est plus marquée dans le CG que dans le CS.

Nous avons, en particulier, observé que certaines UT dans le CG véhiculent des sens courants ou issus d'autres domaines, ce qui entraîne une variation notable des collocations. Certaines UT, comme *population* et *forêt*, présentent des sens spécialisés distincts, tandis que d'autres UT dans le CG renvoient également à des réalités quotidiennes, en complément de leurs sens spécialisés. Ces observations appuient l'idée que la langue de spécialité est un sous-ensemble de la langue générale.

La comparaison des collocations dans les deux corpus a révélé à la fois des convergences et des divergences. D'une part, certains contextes discursifs communs, en particulier ceux liés à la conservation des ressources naturelles,

contribuent à des similitudes collocationnelles entre le CG et le CS. D'autre part, l'association entre la polysémie et les divergences collocationnelles suggère que plus la polysémie des UT est prononcée, plus la divergence dans les choix de collocations entre les corpus est notable. Par ailleurs, la diversité thématique et la multidimensionnalité des termes influencent également la sélection des collocatifs.

Enfin, notre analyse diaphasique des UT pourrait enrichir la structuration des collocations dans des ressources terminologique en fonction des relations sémantiques. Ces perspectives ouvrent la voie à de futures recherches visant à intégrer le contexte discursif dans la description des collocations en langue de spécialité à des fins pédagogiques. Il serait également intéressant d'examiner le niveau de spécialité et les genres textuels pour approfondir notre compréhension de la variation diaphasique des CLS.

## Bibliographie

- BIBER D., CONRAD S., REPPEN R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. New York : Cambridge University Press.
- BOWKER L. (1993). Multidimensional Classification of Concepts for Terminological Purposes. *Advances in Classification Research Online*, 39-56.
- BOWKER L., PEARSON J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora* (1st ed.). London : Routledge.
- CABRÉ M. T. (1998). *La terminologie : théorie, méthode et applications*, traduit du catalan et adapté par Monique C. Cormier et John Humbley. Ottawa & Paris : Les Presses de l'Université d'Ottawa/André Colin.
- CENTRE NATIONAL DE RESSOURCES TEXTUELLES ET LEXICALES. (2012). *Portail lexical : « forêt »*. Consulté le 31 octobre 2024, de <https://www.cnrtl.fr/definition/for%C3%AAt>
- CONDAMINES A. (2005). Linguistique de corpus et terminologie. *Langages* 157, 6-47.
- CONDAMINES A., DEHAUT N. (2011). Mise en œuvre des méthodes de la linguistique de corpus pour étudier les termes en situation d'innovation disciplinaire : le cas de l'exobiologie. *Méta : journal des traducteurs/Meta : Translators' Journal* 56(2), 266-283.
- CONDAMINES A., REBEYROLLE J., SOUBEILLE A. (2004). Variation de la terminologie dans le temps : Une méthode linguistique pour mesurer l'évolution de la connaissance en corpus. *EURALEX 2004 Proceedings*, 547-557.
- CRUSE D. A. (1986). *Lexical Semantics*. Cambridge & New York : Cambridge University Press.
- DROUIN P. (2003). Term Extraction Using Non-Technical Corpora as a Point of Leverage. *Terminology* 9(1), 99-115.

- DROUIN P. (2010). *Université de Montréal, OLST, TermoStat Web 3.0*. Consulté le 31 octobre 2024, de [https://termostat.ling.umontreal.ca/index.php?lang=en\\_CA](https://termostat.ling.umontreal.ca/index.php?lang=en_CA)
- FIRTH J. R. (1957). *Papers in Linguistics, 1934-1951*. London : Oxford University Press.
- GROSSMANN F., TUTIN A. (éds) (2003). *Les collocations : analyse et traitement*. Amsterdam : Éditions De Werelt.
- HALLIDAY M. A. K. (1966). Lexis as a Linguistic Level. In : C. E. Bazell, J. C. Catford & R. H. Robin (eds), *In memory of J.R. Firth*. London : Longmans, 148-162.
- HAUSMANN F. J. (1979). Un dictionnaire des collocations est-il possible ? *Travaux de Linguistique et de Littérature* 17(1), 187-195.
- HAUSMANN F. J. (1989). Le dictionnaire de collocations. In : F. J. Hausmann, O. Reichmann, H. E. Wiegand & L. Zgusta (eds), *Wörterbücher / Dictionaries / Dictionnaires, Ein internationales Handbuch zur Lexikografie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie*. Berlin & New York : De Gruyter, 1010-1019.
- IFREMER. (2024). *Flore et faune benthiques*. Consulté le 7 octobre 2024, de [https://envlit.ifremer.fr/index.php/region/basse\\_normandie/qualite/flore\\_et\\_faune\\_benthiques](https://envlit.ifremer.fr/index.php/region/basse_normandie/qualite/flore_et_faune_benthiques)
- INSTITUT DE L'INFORMATION SCIENTIFIQUE ET TECHNIQUE. (s.d.). *Biodiversité (thésaurus)*. Consulté le 31 octobre 2024, de <https://skosmos.loterre.fr/BLH/fr/>
- KILGARRIFF A., BAISA V., BUSTA J., JAKUBICEK M., KOVAR V., MICHELFEIT J., RYCHLÝ P., SUCHOMEL V. (2014). The Sketch Engine: Ten Years On. *Lexicography* 1, 7-36.
- LAROUSSE (s.d.a). *Dictionnaire de français Larousse : entrée « aire »*. Consulté le 31 octobre 2024, de <https://www.larousse.fr/dictionnaires/francais/aire/1945>
- LAROUSSE (s.d.b). *Dictionnaire de français Larousse : entrée « écosystème »*. Consulté le 31 octobre 2024, de <https://www.larousse.fr/dictionnaires/francais/%C3%A9cosyst%C3%A8me/27682>
- LE BLOG ENJEUX (2024). *Toute l'actualité de l'environnement normatif*. Consulté le 9 octobre 2024, de <http://www.enjeux.fr/index.php?d=08&m=12&y=14&category=4>
- LE MONITEUR EMPLOI. (2024). *Coliving et coworking : ces nouveaux modèles qui font recette dans l'immobilier*. Consulté le 7 octobre 2024, de <https://emploi-btp.lemoniteur.fr/edito/article/coliving-et-coworking-ces-nouveaux-modeles-qui-font-recette-dans-l-immobilier-aea-11304>
- LEXICAL COMPUTING CZ S.R.O. (2024, March 21). *frTenTen – French corpus from the web | Sketch Engine*. *Sketch Engine*. Consulté le 31 octobre 2024, de <https://www.sketchengine.eu/frtenten-french-corpus/>

- L'HOMME M.-C. (2000). Understanding Specialized Lexical Combinations. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 6(1), 89-109.
- L'HOMME M.-C. (2003). Les combinaisons lexicales spécialisées (CLS) : Description lexicographique et intégration aux banques de terminologie. In : F. Grossmann & A. Tutin (éds), *Les collocations : analyse et traitement*. Amsterdam : Éditions De Werelt.
- L'HOMME M.-C. (2020). *La terminologie : Principes et techniques* (2e éd.). Montréal : Les Presses de l'Université de Montréal.
- L'HOMME M.-C., AZOULAY D. (2020). A Comparative Analysis: Collecting Collocations from General and Specialised Corpora. In: G. Corpas Pastor, J.-P. Colson (eds), *Computational Phraseology*. Amsterdam & Philadelphia: John Benjamins Publishing Company, 151-176.
- MEL'ČUK I., CLAS A., POLGUÈRE A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.
- MEYER I., MACKINTOSH K. (2000). When Terms Move into Our Everyday Lives: An Overview of De-terminologization. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 6(1), 111-138.
- NATIONS UNIES (2023). *Qu'est-ce que la biodiversité ?* Consulté le 31 octobre 2024, de <https://www.un.org/sustainabledevelopment/fr/2023/08/01/explainer-what-is-biodiversity/>
- NATIONS UNIES (s.d.). *The United Nations Terminology Database : entrée « écosystème »*. Consulté le 31 octobre 2024, de <https://unterm.un.org/unterm2/en/view/3e914a85-6f04-49c5-9c50-f384e9fcca35>
- OBSERVATOIRE DE LINGUISTIQUE SENS-TEXTE. (2024). *DiCoEnviro. Dictionnaire fondamental de l'environnement*. Consulté le 31 octobre 2024, de [https://olst.ling.umontreal.ca/?page\\_id=214](https://olst.ling.umontreal.ca/?page_id=214)
- OFFICE FRANÇAIS DE LA BIODIVERSITÉ. (s.d.). *Qu'est-ce que la biodiversité ?* Consulté le 9 octobre 2024, de <https://www.ofb.gouv.fr/quest-ce-que-la-biodiversite>
- OPPIDEA. (2024). *Qu'est-ce qu'un écoquartier ?* Consulté le 7 octobre 2024, de <https://www.oppidea.fr/oppideainfos/quest-ce-quun-ecoquartier>
- POLGUÈRE A. (2016). *Lexicologie et sémantique lexicale : notions fondamentales* (3e éd.). Montréal : Les Presses de l'Université de Montréal.
- SINCLAIR J. (1987). Collocation : A Progress Report. In: T. Threadgold, R. Steel (eds), *Language Topics: Essays in Honour of Michael Halliday*. Amsterdam & Philadelphia: John Benjamins Publishing, 319-331.
- SINCLAIR J. (1991). *Corpus, concordance, collocation*. Oxford : Oxford University Press.
- TERRES DE JEUX. (2024). *Queen Games*. Consulté le 9 octobre 2024, de <http://www.terresdejeux.fr/taxonomy/term/429%2C709/3?page=1>

- TUTIN A. (2013). Les collocations lexicales : une relation essentiellement binaire définie par la relation prédicat-argument. *Langages* 189, 47-63.
- VERTS REGION. (2024). *L'habitat alternatif de la loi ALUR*. Consulté le 7 octobre 2024, de <http://vertsregion.org/spip.php?article1270>
- ŽOLKOVSKIJ A., MEL'ČUK I. (1965). O Vozmoznom Metode i Instrumentax Semanticeskogo Sinteza. [À propos d'une méthode possible pour la synthèse sémantique et des instruments correspondants]. *Naucno-TeXniceskaja Informacija* 5, 23-28.

