

## LEXIQUES ET CORPUS AU SERVICE DE LA LITTÉRATIE OCÉANIQUE : PROPRIÉTÉS ET RELATIONS LEXICALES DANS LE DOMAINE DE LA FAUNE MARINE

Silvia Domenica ZOLLO<sup>1</sup>

---

*Article history: Received 06 October 2023; Revised 15 December 2023; Accepted 31 January 2024; Available online 27 March 2024; Available print 27 March 2024.*

©2024 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

---

**ABSTRACT.** *Lexicons and Corpora Serving Ocean Literacy: Properties and Lexical Relations in the Field of Marine Fauna.* In this article, we present the first steps of the *Ocean Literacy* project, which aims to promote the creation and processing of a comparable medium-sized specialised corpus (French-Italian) in the field of marine fauna as well as the creation of a lexicographic resource to describe the lexical relations between the terminological units (TUs) of the field in question, using the theoretical framework of Explanatory and Combinatorial Lexicology. After describing the methodological and theoretical choices made when compiling and processing the *ZooCor* (fr-it) corpus, we will proceed to the analysis of the French term candidates resulting from semi-automatic extraction using *TermoStat* software. We will then describe the general architecture of the lexicographic resource *ZooTerm* which is currently under development, and explain, with the help of a terminographical data model and concrete examples, why we chose to use lexical functions to systematically encode the relations between terms associated with the marine fauna domain. Finally, we will discuss the results and the main theoretical issues raised from this exploratory work.

**Keywords:** *lexicon, corpus, ocean literacy, marine fauna, lexical relations*

---

<sup>1</sup> **Silvia Domenica ZOLLO** (PhD *Doctor Europæus*) est enseignante-chercheuse en Linguistique française à l'Université de Naples Parthenope. Ses recherches portent sur la linguistique de corpus et la sémantique lexicale appliquées aux études lexicographiques et terminographiques (domaines : biologie marine, économie bleue, droit environnemental). Elle s'intéresse également à la néologie dans les discours militants contemporains sur la cause animale, l'environnement et le *climat change*. Adresse : [silvia.zollo@uniparthenope.it](mailto:silvia.zollo@uniparthenope.it).

**ABSTRACT. Lexicoane și corpusuri în slujba alfabetizării oceanice: proprietăți și relații lexicale în domeniul faunei marine.** În acest articol, prezentăm primele etape ale proiectului *Ocean Literacy*, care are ca scop promovarea creării și prelucrării unui corpus specializat comparabil de dimensiuni medii (franceză-italiană) în domeniul faunei marine și crearea unei resurse lexicografice pentru a descrie relațiile lexicale dintre unitățile terminologice (UT) din domeniul în cauză, utilizând cadrul teoretic al lexicologiei explicative și combinatorii. După ce vom descrie alegerile metodologice și teoretice făcute la compilarea și prelucrarea corpusului *ZooCor* (fr-it), vom trece la analiza termenilor francezi candidați rezultați în urma extragerii semiautomate cu ajutorul programului *TermoStat*. Vom detalia apoi arhitectura generală a resursei lexicografice *ZooTerm* în curs de dezvoltare și vom explica, cu ajutorul unui model de fișă terminografică și a unor exemple concrete, motivele care ne-au determinat să folosim funcțiile lexicale pentru a codifica în mod sistematic relațiile dintre termenii asociați cu domeniul faunei marine. În cele din urmă, vom discuta rezultatele și principalele probleme teoretice care decurg din această activitate exploratorie.

**Cuvinte-cheie :** *lexicon, corpus, alfabetizare oceanică, faună marină, relații lexicale*

## Introduction

Cette contribution s'insère dans le cadre du projet de recherche *Littérature océanique : observations linguistiques, données terminologiques et modélisations lexicographiques basées sur corpus (français-italien)* en cours de réalisation auprès de l'Université de Naples Parthenope<sup>2</sup>. Ce projet a un double objectif :

(1) collecter, classer et analyser les lexiques et les discours technoscientifiques autour du domaine de la biologie marine par l'expérimentation des plus récentes approches théoriques et méthodologiques adoptées en linguistique de corpus, en terminologie et en lexicographie spécialisée ;

---

<sup>2</sup> Ce projet a été promu et financé par le Programme PNR (2021-2027) du Ministère de l'Université et de la Recherche (Italie) « Area d'intervento 5.6.5 Conoscenza, innovazione tecnologica e gestione sostenibile degli ecosistemi marini » au sein du laboratoire CLab de l'Université de Naples Parthenope (P.I. : Prof. A. Garofalo). En 2022, le projet a été également financé par le Conseil de recherche de l'Université de Naples Parthenope dans le cadre de l'appel à projet « Fonds de Développement de la Recherche Universitaire 2022-2025 » (HUMAENVI – P.I. : Dr. S. D. Zollo). Nous remercions les relecteurs anonymes pour leurs critiques et leurs conseils. Nous aimerions également remercier Alain Polguère pour les échanges très constructifs et pour ses précieuses suggestions qui ont permis de clarifier certains points.

(2) valoriser le patrimoine naturel marin (flore et faune, gestion des aires marines protégées, etc.) à travers la création de ressources linguistiques multilingues conçues pour la vulgarisation des savoirs scientifiques dans une perspective de science ouverte.

Nous présentons les premières phases de réalisation de ce projet qui vise à promouvoir la constitution et le traitement d'un corpus spécialisé bilingue (français-italien) de taille moyenne du domaine de la faune marine. Ce projet a un autre but très pratique : la conception d'une ressource lexicographique qui a pour objet de décrire la grande variété des relations sémantico-lexicales entre les unités terminologiques (UT) du domaine en question à l'aide d'une expérimentation basée sur le cadre théorique de la Lexicologie explicative et combinatoire (LEC) (Mel'čuk *et al.* 1995 ; Wanner 1996 ; L'Homme 2020b).

Il est largement reconnu que les termes entretiennent des relations sémantico-lexicales très variées avec d'autres unités lexicales (UL) : elles peuvent être hiérarchiques, comme celles qui s'établissent entre un hyperonyme et ses hyponymes (*cétacé* → *baleine, cachalot, épaulard*) ou non hiérarchiques, à savoir les relations d'implications sémantiques, de cause effet (*menace climatique* → *extinction des espèces*), etc. Les méthodologies adoptées par la théorie classique de la terminologie – qui placent le concept au centre de l'analyse – ont conduit les terminologues à se focaliser sur les relations taxonomiques et méronymiques et à négliger un ensemble d'autres relations jugées pertinentes et adéquates pour la représentation des langues de spécialité. En particulier, au cours des dernières décennies, les terminologues ont remis en question la validité de ces méthodologies et se sont tournés vers des modèles descriptifs alternatifs, tels que ceux fournis par la lexicologie et la lexicographie générale et spécialisée (Binon *et al.* 2000 ; Dancette *et al.* 2004 ; L'Homme 2008 ; Zanola 2018, 2023).

Or, pour ce qui est du domaine de la faune marine, comment envisager les relations entre les termes ? Faut-il essayer de décrire l'organisation des connaissances, en utilisant les termes comme vecteurs linguistiques de ces savoirs ou faut-il tenir compte des UL ayant des sens spécifiques ? L'approche conceptuelle promue par la théorie classique de la terminologie conduirait à considérer un ensemble limité de relations, principalement logiques ou hiérarchiques. En revanche, l'accent mis sur les sens spécifiques des UL nous amène à adopter l'approche lexico-sémantique et à envisager les multiples relations sémantico-lexicales dans lesquelles s'inscrivent les termes. Bien qu'expérimentale, cette étude tentera de montrer qu'il est possible de combiner les approches conceptuelle et lexico-sémantique pour l'étude du lexique de la faune marine. Après avoir décrit les choix méthodologiques et théoriques pour la constitution et le traitement du corpus *ZooCor* (§ 1.), nous analyserons un échantillon de termes candidats issus d'une extraction semi-automatique (§ 2.)

réalisée via le logiciel *TermoStat*. Ensuite, nous détaillerons l'architecture générale de la ressource lexicographique *ZooTerm* en cours d'élaboration (§ 3.) et expliquerons, au moyen d'un modèle de fiche terminographique et par des exemples concrets, les raisons qui nous ont poussés à utiliser les fonctions lexicales (Mel'čuk *et al.* 2021) pour encoder systématiquement les liens entre les termes associés à ce domaine (§ 4.). Nous verrons comment ce travail exploratoire fait apparaître plusieurs questions théoriques fondamentales sur l'articulation entre les classes de termes de la faune marine, leur nature compositionnelle/non compositionnelle et sur la modélisation des relations actancielles et circonstancielles. Pour terminer (§ 5.), nous discuterons les résultats obtenus avant de conclure sur les limites et les potentialités offertes par les nouvelles pratiques lexicographiques émergentes en terminologie en vue du partage des savoirs et des connaissances scientifiques.

## 1. Cadre théorique et prémisses méthodologiques

La production et la compréhension de textes spécialisés supposent l'acquisition de pratiques discursives représentatives des genres textuels de la communauté de discours de référence, en fonction également de l'intentionnalité de l'émetteur et de l'acceptabilité qui en découle de la part des destinataires. Dans ce contexte, les recherches menées en terminologie textuelle et discursive, en linguistique de corpus et en sémantique du genre (Branca-Rosoff 1999 ; Bowker *et al.* 2002 ; Rakotonoelina 2014 ; Altmanova *et al.* 2018) revêtent une importance particulière, tant pour les études descriptives que pour les applications didactiques en terminologie et en lexicographie spécialisée basées sur corpus. En dévoilant la manière dont les textes reflètent les communautés de discours qui les utilisent, ces études représentent un outil puissant aussi bien pour l'analyse des besoins communicatifs du public cible que pour le découpage d'une langue en différents sous-domaines. Bien que le discours autour de la faune marine se situe *a priori* à l'intersection de plusieurs domaines<sup>3</sup>, lors de sa délimitation on se heurte au même type de difficulté que celui qu'on rencontre avec toute langue de spécialité : les frontières qui séparent les différents sous-domaines ne sont pas nettes, étant donné le principe du *continuum* des connaissances. Néanmoins, dès qu'on adopte une approche linguistique fondée sur les corpus, on est amené à définir les contours de l'échantillon considéré.

---

<sup>3</sup> Le discours sur la faune marine se situe à la charnière de plusieurs sous-discours allant de la taxonomie à l'écologie, en passant par la biologie de la conservation et les sciences naturelles (Sandulli *et al.* 2011).

Compte tenu des descripteurs linguistiques employés pour la catégorisation des textes en domaines et en genres (Poudat *et al.* 2006) et des études sur l'écrit scientifique entre lexiques et discours (Tutin *et al.* 2014), dans la phase d'identification des sous-domaines, nous avons choisi de substituer les notions de langue de spécialité et sous-langage (Lerat 2016) par la notion de « communauté de discours » (Swales 1990, 24). S'appuyant sur des paramètres sociodiscursifs, Swales définit la communauté de discours comme un groupement d'individus qui partage des dispositifs communicatifs (actes de colloques, presse professionnelle et scientifique, etc.) et des buts communs explicites (la production de biens matériels ou de connaissances scientifiques) et qui emploie, de manière préférentielle, des genres discursifs propres et un lexique spécifique répondant aux attentes des membres locuteurs.

Dans le cadre de la communauté de discours que nous nous apprêtons à étudier, les buts communicatifs gravitent autour de la promotion de la recherche et de la découverte scientifique dans le domaine de la faune marine. Les matériaux récoltés pour la constitution du corpus se basent sur des typologies textuelles appartenant à la communication spécialisée – et plus proprement – aux genres du discours scientifique. Nous considérons ce dernier comme un discours englobant différents genres et textes produits par des experts et des chercheurs pour la diffusion de leurs travaux au sein de la communauté scientifique et à l'extérieur pour la transmission des savoirs scientifiques, qui varient selon le domaine, le niveau de technicité et le contexte. Deux typologies de variation peuvent ainsi être identifiées dans notre corpus : la première est liée aux contenus et aux sous-domaines ; la deuxième concerne la variété des genres textuels relevant des discours scientifique, professionnel et de vulgarisation, et le changement de registre qui a lieu en fonction du degré de spécialisation des termes et du contexte discursif.

### ***1.1. Pleins feux sur la faune marine : quelle place pour le corpus ZooCor ?***

Depuis une dizaine d'années, les biologistes marins, les zoologistes et les océanographes sont de plus en plus amenés à faire face à des thématiques d'intérêt public, comme les impacts du réchauffement climatique sur les océans, la surexploitation de ses ressources, la dégradation des écosystèmes marins et la présence d'espèces marines menacées, vulnérables et envahissantes dans nos mers. Être capable de divulguer ces savoirs, dans le contexte scientifique comme vis-à-vis du grand public, demande la connaissance d'un langage technique qui évolue constamment et qui s'adapte aux découvertes et aux changements de notre planète.

Les experts sont ainsi de plus en plus conscients de l'importance d'une approche non isolationniste de la terminologie de ce domaine et de la nécessité d'améliorer les stratégies d'enseignement et de vulgarisation de celle-ci. Le lexique fondamental de la faune marine, c'est-à-dire l'ensemble des termes qui forment le noyau notionnel du domaine en question, est celui qui pose beaucoup de problèmes en ce qui concerne les interactions avec la langue générale, étant donné la façon d'examiner les liens entre noms, concepts, termes et objets taxonomiques (Tillier 2005). Malgré les efforts faits par la Commission internationale de nomenclature zoologique pour créer un nouveau système nomenclatural plus analogue aux théories modernes de l'évolution, le lexique spécialisé de la faune marine ou – plus largement de la zoologie – reste vaguement normalisé et peu accessible à un large public en raison de la variété des méthodes employées pour relier les noms et les concepts en taxonomie<sup>4</sup>. Il existe également un manque de perspective multilingue car la plupart des articles scientifiques sont rédigés en anglais, ce qui entraîne des problèmes dans la communication multilingue et ce, sans tenir compte du fait que la nomenclature taxonomique utilise le latin pour classer et nommer l'ordre du monde vivant.

Or, comment trouver les termes adéquats pour parler de faune marine ? Comment faire dialoguer biologistes, zoologistes marins et linguistes ? Comment se comprennent-ils lorsqu'ils cherchent à communiquer à propos de leurs disciplines ? Il apparaît crucial de promouvoir des projets interdisciplinaires sur la littératie des espèces marines à travers la création de ressources linguistiques multilingues favorisant un premier niveau de vulgarisation du vaste patrimoine naturel de la faune marine dans les secteurs de la recherche, de l'éducation et de la science citoyenne. C'est dans ce contexte que nous avons créé *ZooCor*, un corpus bilingue comparable (français-italien) visant plusieurs objectifs, tels que la création de matériel linguistique exploitable pour les chercheurs en sciences du langage (extraction automatique d'UT, études quantitatives et qualitatives sur les caractéristiques lexicales, morphosyntaxiques et discursives, etc.) et la consultation en libre accès de textes écrits authentiques à utiliser dans les contextes pédagogiques – que ce soit pour des applications didactiques ou des projets de science citoyenne – dans le but de sensibiliser le grand public à la langue de spécialité en question.

### ***1.2. Le corpus ZooCor : constitution et architecture générale***

Le système mis en place pour la constitution du corpus *ZooCor* s'est déroulé en trois phases :

---

<sup>4</sup> Pour des approfondissements sur l'histoire des nomenclatures des sciences nouvelles, consulter Selosse (2012) et Grimaldi (2017).

- (1) l'élaboration d'un modèle pour la construction et la gestion du corpus ;
- (2) la sélection et le traitement textuel ;
- (3) le stockage des textes et l'étiquetage des métadonnées.

Dans la phase d'élaboration, nous avons tout d'abord mené une analyse des besoins linguistiques et extralinguistiques pour affiner le périmètre de notre corpus, en fonction du public cible et des finalités du projet. Pour ce faire, nous avons consulté des experts et des chercheurs en biologie marine du Département de Sciences et Technologies de notre université et de la Stazione Zoologica Anton Dohrn de Naples. Les résultats de l'analyse des besoins ont permis d'élaborer l'architecture générale du corpus, basée sur une catégorisation à plusieurs niveaux qui prend en compte la langue, le canal de réception, le domaine, le discours, le genre textuel et la fenêtre temporelle.

Comme nous l'avons signalé ci-dessus (§ 1.1.), *ZooCor* est un corpus comparable qui implique actuellement deux langues romanes : le français et l'italien. Pour des raisons scientifiques et techniques, le français se trouve en tant que langue pivot du corpus : tout le projet est centré, en particulier, sur la description de la terminologie en langue française du domaine en question et sur des recherches contrastives du français par rapport à d'autres langues<sup>5</sup>, parmi lesquelles l'italien. Malgré l'apparente similitude entre les deux systèmes linguistiques, les deux communautés discursives concernées conceptualisent de manière distincte le lexique et les terminologies scientifiques de la faune marine, en raison de leur diversité culturelle en matière de communication et de gestion socioenvironnementale, juridique et managériale de la biodiversité marine<sup>6</sup>.

En ce qui concerne le canal, nous n'avons inclus que les textes écrits qui l'emportent sur la communication orale dans les contextes académiques, scientifiques et de vulgarisation. Quant au critère thématique, nous avons choisi d'articuler le corpus *ZooCor* autour de cinq sous-domaines de la faune marine, correspondant aux familles des Chéloniens, des Delphinidés, des Phocidés, des Coralliaires et des Astéries. Trois éléments fondamentaux ont déterminé le choix de circonscrire la sélection textuelle à ces sous-domaines : la présence massive de ces espèces dans nos mers, leur importance dans le monde de la recherche, de l'éducation et de la science citoyenne, et les récents développements théoriques et pratiques que la biologie de la conservation a connus en Europe

---

<sup>5</sup> Le corpus a vocation de s'élargir à d'autres langues représentées au sein de notre équipe de recherche ainsi que d'envisager des sous-corpus plurilingues (traduits ou comparables).

<sup>6</sup> Pour ne citer qu'un exemple, le terme complexe *grand dauphin*, ainsi dénommé pour sa taille, se traduit en italien par le terme simple *tursiope* (ou *tursio*). Une étude contrastive (français-italien) permettrait de mieux repérer les spécificités lexicales, morphosyntaxiques et discursives dans les deux systèmes linguistiques en question.

les deux dernières décennies grâce aux plans d’actions et aux stratégies politiques de préservation des animaux présents dans les eaux marines. Ces raisons nous ont poussés à sélectionner les textes appartenant à la fenêtre temporelle allant de l’année 2000 à l’année 2022, une période particulièrement prolifique pour les programmes de recherche et de recensement de la vie marine.

Le critère discursif a été établi sur la base des principales composantes communicatives caractérisant les textes collectés, en particulier le rôle des auteurs et la finalité des textes, tandis que les genres ont été envisagés comme des outils communicatifs marqués par des conventions structurelles et thématiques récurrentes partagées au sein d’une communauté discursive. Nous avons ainsi sélectionné trois typologies de discours :

1) le discours académique (articles scientifiques, livres, monographies, thèses de doctorat, etc.) ;

2) le discours professionnel (rapports et descriptions de projets, manuels techniques, matériel promotionnel, etc.) ;

3) le discours de vulgarisation (livres de vulgarisation pour le grand public, fiches documentaires et pédagogiques, glossaires, etc.).

Après avoir établi les critères linguistiques et extralinguistiques, nous avons procédé à la sélection et au classement des textes. Tous les documents de départ au format .pdf ont été convertis au format .txt avec un encodage UTF-8 à l’aide d’*Abbyy FineReader*, un logiciel de reconnaissance optique. À ce stade, il a également fallu « nettoyer » les textes pour garantir leur intégrité formelle et leur qualité (suppression des tableaux, des en-têtes de page, des notes, de la bibliographie finale, d’éventuelles coquilles, etc.). Les fichiers ainsi créés ont été sauvegardés sur le nuage informatique OneDrive mis à disposition par notre université. Ensuite, chaque texte a été étiqueté sur un fichier Excel, selon un ensemble de 11 classes de métadonnées textuelles (Tab. 1) permettant la récupération automatique des documents et la création d’éventuels sous-corpus pour des recherches ciblées sur des sous-domaines ou des genres textuels spécifiques.

Une quatrième et dernière phase, basée sur le traitement informatique du corpus (stockage textuel, lemmatisation et étiquetage morphosyntaxique) et sa mise en ligne, est prévue l’année prochaine en collaboration avec des spécialistes en TALN. Cette étape permettra d’interroger le corpus, de générer des listes de mots, d’étudier le comportement des TC en fonction de plusieurs critères, par exemple, selon le contexte de droite ou de gauche ou les propriétés de KWIC (forme, lemme, marque morphologique, etc.) et de faire un traitement statistique des données recueillies (calcul des collocations, score de spécificité, t-score, distribution de la fréquence absolue et relative, etc.).



**Tab. 1.** Métadonnées textuelles du corpus *ZooCor*.

<b>Métadonnées textuelles</b>	Domaine et sous-domaines
	Langue
	Titre
	Auteur
	Année de publication
	Année de consultation
	Éditeur
	N. de pages
	Format du texte (.pdf et .txt)
	Mots-clés
	Genres textuels
	Remarques

À présent, le corpus *ZooCor* se compose de 350 textes en français (dont 37 % de textes appartenant au discours académique, 34 % appartenant au discours professionnel et 29 % au discours de vulgarisation) pour un total de 425 000 *tokens*, et de 110 textes en italien (dont 44 % appartenant au discours académique, 27 % au discours professionnel et 29 % au discours de vulgarisation) pour un total de 210 000 *tokens*<sup>7</sup>.

## **2. L'extraction terminologique : test expérimental**

En vue de tester le corpus *ZooCor* pour la langue française, nous avons effectué une extraction terminologique automatique via l'outil *TermoStat* (Drouin 2003), qui se base sur une méthode probabiliste s'inspirant du calcul des spécificités de Lafon (1980) et qui associe la fréquence d'un terme observée dans un corpus spécialisé avec celle prévisible dans un corpus de langue générale, dénommé corpus de référence. Plus la fréquence mesurée dans le corpus spécialisé est supérieure à celle du corpus de référence, plus le candidat terme est potentiellement lié au domaine de spécialité.

L'extraction a été réalisée à partir d'un échantillon de 80 textes (38 % relevant du discours académique, 28 % du discours professionnel et 34 % du discours de vulgarisation) du corpus français, allant de 2015 à 2022 et portant sur le sous-domaine des Chéloniens et les stratégies de conservation des tortues marines. Après avoir soumis l'échantillon à l'étiquetage automatique, le logiciel a fourni une liste des candidats termes simples et complexes ayant reçu les scores les plus hauts en termes de fréquence et de spécificité (Fig. 1).

---

<sup>7</sup> Le corpus est constamment actualisé.

Corpus >> tortues1				
Résultats				
<span>Liste des termes</span> <span>Usage</span> <span>Statistiques</span> <span>Structuration</span> <span>Bigrammes</span>				
Candidat de regroupement	Fréquence	Score (Spécificité)	Variants orthographiques	Matrice
tortue	794	515.18	tortue tortues	Nom
tortue marine	461	406.57	tortue marine tortues marines	Nom Adjectif
marin	614	373.41	marin marine marines	Adjectif
tortue	358	348.43	tortue tortues	Adjectif
ponte	319	322.99	ponte pontes	Nom
plan national d' action	142	225.34	plan national d' action plan national d' actions plans nationaux d' actions	Nom Adjectif Préposition Nom
tortue verte	124	210.46	tortue verte tortues vertes	Nom Adjectif
plage de ponte	45	125.88	plage de ponte plages de ponte plages de pontes	Nom Préposition Nom
plan	582	121.19	plan plans	Nom
tortue imbriquée	41	120.03	tortue imbriquée tortues imbriquées	Nom Adjectif
espèce	265	119.95	espèce espèces	Nom
sensibilisation	80	115	sensibilisation sensibilisations	Nom
habitat	128	114.3	habitat habitats	Nom
priorité	38	113.94	priorité	Nom
zone d' alimentation	37	113.87	zone d' alimentation zones d' alimentation	Nom Préposition Nom
herbier	46	113.29	herbier herbiers	Nom
capture accidentelle	38	112.5	capture accidentelle captures accidentelles	Nom Adjectif
site d' alimentation	36	112.28	site d' alimentation sites d' alimentation	Nom Préposition Nom
caouanne	36	112.28	caouanne caouannes	Nom
sulvi	133	110.35	sulvi suvivis	Nom
imbriquée	46	105.35	imbriquée	PPAdj

**Fig. 1.** Liste des candidats termes triés selon la fréquence et la spécificité.

Bien que ce type de test ait fourni des résultats intéressants, l'exploitation d'une fréquence « anormale » n'a pas été suffisante pour isoler les termes et la liste comportait du bruit. Ainsi, nous avons fait un filtrage manuel des termes extraits, en excluant un nombre important de termes appartenant à d'autres domaines (économie bleue, droit environnemental, pollution lumineuse, pollution acoustique, etc.). Pour la sélection des UT, nous avons également eu recours aux critères proposés par L'Homme (2020a) :

- (1) l'UT a un sens spécifique lié au domaine examiné ;
- (2) les unités prédicatives ne sont choisies que si les actants renvoient à des UT sélectionnées en fonction du critère (1). Cependant, si l'UT se combine avec des actants non spécialisés, elle est exclue ;
- (3) les dérivés morphologiques sémantiquement proches aux UT choisies à partir des critères (1) ou (2) sont enregistrés ;
- (4) les UT qui ont des liens paradigmatiques avec les termes triés selon les principes (1), (2) ou (3) sont gardées.

Ces paramètres ont permis de systématiser nos décisions, de justifier la sélection des UT sur des bases tangibles et d'envisager le terme comme une entité qui évolue dans le lexique et dans le discours. Ensuite, nous avons mené une étude qualitative du TC *tortue marine*, en proposant un premier classement typologique des données linguistiques et des liens sémantiques récoltés (Fig. 2) via un modèle de fiche à développer dans la ressource lexicographique *ZooTerm*.



La description des concepts clés prévoit, d'un côté, l'identification des relations sémantico-lexicales entre les termes et, de l'autre, l'explication des nuances de sens. Outre les 50 concepts clés, *ZooTerm* contient un lexique de 30 termes en langue française (et ses équivalents en italien et anglais)<sup>8</sup> traités dans le corps des 50 articles. Chaque entrée se compose de 11 sections qui font l'objet des sous-sections suivantes : les deux premières sous-sections contiennent le terme vedette, le domaine, les informations grammaticales, les équivalents en italien et en anglais et la définition alors que la troisième renferme les contextes d'usage<sup>9</sup> extraits du corpus *ZooCor*.

Dans la quatrième sous-section figurent les synonymes (comprenant également les quasi-synonymes, les variantes diatopiques et diastratiques), le nom scientifique, le nom vulgarisé – n'apparaissant que si la fiche comporte le nom de l'espèce – et les relations sémantico-lexicales entre le terme vedette et d'autres termes appartenant au même champ (terme générique de l'espèce, termes désignant les espèces spécifiques, antonymes, parties de l'espèce, lieux et habitats typiquement fréquentés par l'espèce, actions réalisées par l'espèce, agents et instruments impliqués dans la sauvegarde de l'espèce, agents et instruments impliqués dans la menace de l'espèce, cycle de vie de l'espèce, etc.). Enfin, la dernière sous-section dénommée 'Informations supplémentaires' englobe les informations extralinguistiques (notes historiques ou encyclopédiques, curiosités, photos, etc.) sur le terme décrit. Le Tableau 2 montre un exemple concret de la mise en œuvre de cette architecture à partir du TC *tortue marine*.

Différentes relations sémantico-lexicales entre le terme vedette *tortue marine* et d'autres termes sont illustrées tout au long de la fiche terminographique (Tab. 2). Ces liens peuvent être classifiés comme suit : l'hyponymie et l'hyperonymie (*tortue marine* → *tortue rayonnée*, *tortue caouanne*, *tortue verte*, *tortue luth*, *tortue imbriquée*, *tortue franche*, *tortue olivâtre*, *tortue de Kemp*) ; la synonymie (*tortue marine* → *tortue de mer*, *tortue d'eau salée*) ; les agents impliqués dans le processus de conservation de l'espèce (*biologiste marin*, *biologiste de la conservation*, *chercheur*, *naturaliste*, *bénévole*, etc.) ; d'autres actants ou circonstants, comme la localisation typique (*océan*, *mer*, *plage*, *eau de mer*, *aire marine protégée*, *corridor biologique sous-marin*, *centre de soins*, etc.) ; le sens exprimé par des termes appartenant à d'autres parties du discours (*ponte* → *pondre* ; *éclosion* → *éclore* ; *nidification* → *nicher* ; *émergence* → *émerger*) et les relations syntagmatiques

<sup>8</sup> Les équivalents italiens ont été puisés au corpus *ZooCor* (version italienne) et les équivalents anglais ont été empruntés aux bases de données *TermiumPlus* et *GDT*.

<sup>9</sup> La méthode mise en œuvre pour identifier les contextes d'usage s'appuie sur l'exploitation des patrons de connaissances au sens de Meyer (2001) et Kilgarriff *et al.* (2008) et de plusieurs mesures d'association pour l'extraction des collocations. Pour des questions d'espace, nous ne l'abordons pas dans cet article. Elle fera l'objet d'une réflexion ultérieure.

telles que les collocations nominales et verbales (*menace de la tortue marine ; pondre des œufs*, etc.). D'autres liens lexicaux tels que la quasi-synonymie (*tortue, tortue océanique, tortue aquatique*), la méronymie (*carapace, dossière, bec, nageoires, écailles, plaques, plastron, crâne, vertèbres, os*) et l'antonymie (*tortue terrestre, tortue lacustre, tortue d'eau douce*) apparaissent dans le modèle de fiche.

**Tab. 2.** Modèle de fiche terminographique (terme vedette *tortue marine*).

<b>Terme vedette</b>	<b>Tortue marine</b>
<b>Domaine</b>	Zoologie – Herpétologie
<b>Informations grammaticales</b>	n. f.
<b>Équivalent(s) en italien</b>	tartaruga marina, tartaruga di mare
<b>Équivalent(s) en anglais</b>	marine turtle, sea turtle, turtle
<b>Définition</b>	Tortue vivant en eau salée, possédant de grands ailerons antérieurs en forme de palmes qu'elle fait bouger simultanément pour se déplacer, et une carapace aplatie lui conférant de l'hydrodynamisme. ( <i>GDT</i> )
<b>Contexte(s) d'usage</b>	« Plusieurs raisons peuvent expliquer ce défaut d'observation : d'une part, les <i>tortues marines</i> viennent pondre en masse de nuit et il est logiquement très difficile de patrouiller toutes les nuits pendant la nuit entière ». ( <i>ZooCor</i> )  « La totalité des traces laissées par les <i>tortues marines</i> ont été identifiées comme des traces de tortues vertes, les empreintes des nageoires étant apparues profondes et symétriques, typiques du mode de déplacement 'par rebond' de cette espèce ». ( <i>ZooCor</i> )
<b>Synonyme(s)</b>	tortue de mer, tortue d'eau salée tortue, tortue océanique, tortue aquatique
<b>Nom scientifique</b>	Chelonioidea
<b>Nom vulgarisé</b>	Tortue
<b>Relations sémantico-lexicales</b>	
<i>Terme générique de la t. m.</i>	animal reptile chélonien
<i>Espèces spécifiques</i>	tortue rayonnée, tortue caouanne, tortue luth, tortue verte, tortue imbriquée, tortue franche, tortue olivâtre, tortue de Kemp
<i>Antonymes de t. m.</i>	tortue terrestre, tortue lacustre, tortue d'eau douce
<i>Parties d'une t. m.</i>	carapace, dossière, bec, nageoires, écailles, plaques, plastron, crâne, vertèbres, os

<i>Lieux et habitats typiquement fréquentés par la t.m.</i>	océan, mer, plage, aire marine protégée, corridor biologique sous-marin, aquarium, centre de soins
<i>Actions réalisées par la t.m.</i>	pondre (ponte), éclore (éclosion), nicher (nidification), émerger (émergence), plonger (plongée), respirer (respiration)
<i>Agents et instruments impliqués dans le processus de conservation</i>	biologiste marin, biologiste de la conservation, vétérinaire, chercheur, naturaliste, bénévole
<i>Agents et instruments impliqués dans la menace</i>	homme, pêcheur, navire, bateau, pollution marine, sac en plastique, engins de pêche, réseaux, nasse, bobèche
<i>Cycle de reproduction de la t.m.</i>	parade nuptiale, accouplement, inter-ponte, ponte, nidification, incubation des œufs, éclosion, émergence
<b>Informations supplémentaires</b>	<a href="https://fr.wikipedia.org/wiki/Tortue_marine">https://fr.wikipedia.org/wiki/Tortue_marine</a>

Après avoir intégré ces relations dans la fiche terminographique, nous avons voulu les modéliser via un travail exploratoire basé sur l'emploi des fonctions lexicales. Dans le paragraphe suivant, nous présenterons les fonctions lexicales et expliquerons plus en détail leur intérêt pour l'analyse des relations sémantico-lexicales entre les termes en question.

#### 4. Les fonctions lexicales : pourquoi ?

La fonction lexicale (FL) est un outil formel conçu dans le cadre de la Théorie Sens-Texte (TST) (Mel'čuk 1997) pour décrire les relations lexicales paradigmatiques et syntagmatiques sur le modèle des fonctions mathématiques. Plus précisément « une fonction lexicale  $f$  est une fonction qui associe à une lexie  $L$  un ensemble d'expressions linguistiques (lexies, syntagmes, etc.)  $\{L_1, \dots, L_n\}$  qui ont le sens ' $f$ ' portant sur le sens de  $L$  (= sur ' $L$ ') et qui sont sélectionnées en fonction de  $L$  » (Mel'čuk *et al.* 2014, 210) :

$$f(L) = \{L_1, \dots, L_n\} \mid Li('f') \text{ et } 'f' ('L')$$

Elle peut également s'écrire de la façon suivante :  $f(x)=y$ , où  $f$  représente la fonction,  $x$  l'argument (ou mot-clé) et  $y$  la valeur exprimée par la fonction  $f$  lorsqu'elle est appliquée à un argument  $x$  donné. « La fonction lexicale  $f$  est telle que : l'expression  $f(L)$  représente l'application de  $f$  à  $L$  ; chaque élément de la valeur de  $f(L)$  est lié à  $L$  (à peu près) de la même façon » (Polguère 2016, 196). Il y a autant de fonctions lexicales que de liens sémantico-lexicaux et chaque FL est identifiée par une étiquette : Syn pour indiquer la synonymie, Anti pour signaler l'antonymie, Magn pour l'intensification, etc.

En principe, les FL ont été élaborées pour analyser les relations sémantico-lexicales entre les lexies dans les dictionnaires généraux (Picoche *et al.* 2001 ; Mel'čuk *et al.* 1984-1999, 2007). Ce n'est que dans la dernière décennie que leur applicabilité en terminologie et en lexicographie spécialisée a été examinée par plusieurs chercheurs pour la modélisation des terminologies techniques, telles que celles du commerce, du marketing, de l'informatique, de l'environnement et de la chimie (Dancette *et al.* 2004 ; L'Homme 2009 ; Ingrosso *et al.* 2015 ; León-Araúz *et al.* 2019 ; Bonadonna 2020 ; Frassi 2021). En s'inspirant de ces travaux, nous avons expérimenté l'usage des FL dans notre projet de recherche pour plusieurs raisons : (a) tout d'abord, le modèle est formellement compatible avec l'architecture de notre ressource lexicographique ; (b) les FL sont en mesure de regrouper un large éventail de sens entre les termes du champ lexical de la faune marine ; (c) plusieurs relations – paradigmatiques et syntagmatiques – présentes dans la nomenclature de départ de notre lexique peuvent être enregistrées par ce dispositif ; (d) les FL peuvent être encodées et vulgarisées pour les utilisateurs non experts (Polguère 2003), ce qui a l'avantage de rendre les connaissances accessibles à tous.

#### 4.1. L'attribution des fonctions lexicales aux termes de ZooCor

Au moment de l'attribution des FL à l'échantillon de termes extraits du corpus *ZooCor*, un certain nombre de choix méthodologiques ont été faits. Nous en illustrons un premier avec la modélisation de l'UT *espèce*, dont un examen attentif révèle qu'il s'agit d'un terme polysémique. La première acception *espèce<sub>I</sub>* fait référence au niveau de classification taxonomique des êtres vivants, alors que la seconde *espèce<sub>II</sub>* est strictement liée au concept biologique, à savoir l'« ensemble d'êtres vivants possédant des caractères anatomiques, morphologiques et physiologiques communs qui reproduisent entre eux des êtres semblables et également féconds » (*TLFi*), ce qui nous a obligées d'assigner les FL séparément à chaque sens, comme le montre le Tableau 3.

**Tab. 3.** FL attribuées à deux sens différents du terme *espèce*.

	<b>Espèce<sub>I</sub></b>	<b>Espèce<sub>II</sub></b>
<b>Domaine</b>	Taxonomie	Biologie
<b>FL</b>	S <sub>1</sub> ( <i>espèce<sub>I</sub></i> )= <i>taxonomiste</i> S <sub>loc</sub> ( <i>espèce<sub>I</sub></i> )= <i>classification</i>	S <sub>1</sub> ( <i>espèce<sub>II</sub></i> )= <i>biologiste, naturaliste</i> S <sub>loc</sub> ( <i>espèce<sub>II</sub></i> )= <i>habitat</i> LiquFunc <sub>0</sub> ( <i>espèce<sub>II</sub></i> )= <i>détruire</i> FinFunc <sub>0</sub> ( <i>espèce<sub>II</sub></i> )= <i>éteindre, menacer</i> CausContPred( <i>espèce<sub>II</sub></i> )= <i>sauvegarder, monitorer</i>

Dans le réseau des liens paradigmatiques d'*espèce<sub>I</sub>* et d'*espèce<sub>II</sub>*, nous avons détecté des dérivés sémantiques nominaux actanciels et circonstanciels respectivement représentés par les étiquettes  $S_1$  (actant syntaxique profond – AsyntP de L ou nom d'agent) et  $S_{loc}$  (circonstant de lieu typique de L ou nom de lieu). Quant aux liens syntagmatiques, plusieurs fonctions lexicales verbales apparaissent dans le réseau terminologique d'*espèce<sub>II</sub>*. Plus précisément, il s'agit de FL complexes, à savoir des fonctions qui se comportent comme un « enchaînement de FL simples syntaxiquement liées, cet enchaînement ayant une valeur globale cumulative, qui exprime de façon indécomposable, le sens de l'enchaînement entier » (Mel'čuk *et al.* 1995, 148). Elles sont écrites sous forme d'une chaîne de symboles de FL simples constituantes. Nous en illustrons trois exemples ci-dessous :

LiquFunc<sub>0</sub>(*espèce<sub>II</sub>*)=*détruire*. Cette FL désigne 'quelqu'un ou quelque chose qui détruit une espèce'. Elle est composée du verbe collocatif Liqu qui signifie 'faire cesser' et de Func, un verbe support assimilable à 'avoir lieu', 'exister'.

FinFunc<sub>0</sub>(*espèce<sub>II</sub>*)=*éteindre, menacer*. Cette FL prend le sens de 'cesser d'exister' du verbe collocatif phasique Fin qui signifie 'cesser' et du verbe support Func.

CausContPred(*espèce<sub>II</sub>*)=*sauvegarder, monitorer*. Cette FL signifie 'quelqu'un ou quelque chose qui fait en sorte qu'une espèce reste dans son état actuel'. Elle est constituée du verbe collocatif de causation Caus qui signifie 'causer', 'faire en sorte que', du verbe collocatif phasique Cont qui signifie 'continuer' et de Pred qui a le sens 'être'<sub>(copul)</sub>.

D'autres problèmes descriptifs concernent la modélisation des locutions faibles (très nombreuses dans le corpus) dont la nature sémantiquement non compositionnelle est parfois très difficile à cerner à cause du caractère « spécialisé » des définitions. Plus précisément, il s'agit des locutions nominales qui dénotent des types d'entités plutôt que des entités lexicales à proprement parler, telles qu'*espèce paravent*, *espèce parapluie*, *espèce-ingénieur*, *espèce clé-de-voûte*, *chien de mer*, *grand dauphin* et *dauphin bleu et blanc*<sup>10</sup>. Pour encadrer la

<sup>10</sup> À noter que Mel'čuk ne considère pas ces syntagmes comme des locutions : à son avis, ces entités lexicales n'ont pas le statut de lexie, contrairement au point de vue de Polguère qui – nous semble-t-il – met en lumière la nécessité de donner une définition lexicographique à ces entités contrôlant leur propre réseau de liens paradigmatiques et syntagmatiques. Pour ce faire, il a introduit la notion de « termème » (Mel'čuk 2023, 145) pour désigner une classe d'individus, d'entités ou faits spécifiques qui correspondent aux termes authentiques des



problématique, examinons le cas particulier du syntagme *dauphin bleu et blanc*. Ce syntagme est ainsi défini dans le corpus *ZooCor* :

« Le dauphin bleu et blanc est l'une des cinq espèces du genre *Stenella*. [...] Type de dauphin, dont la coloration supérieure est une sorte de bleu qui s'étire de l'œil vers la région anale et la coloration inférieure une sorte de blanc qui couvre ses flancs. Il a une nageoire dorsale falciforme et des comportements démonstratifs avec sauts fréquents [...]. La bouche, de forme droite, s'élanche vers l'œil et est munie de deux mandibules qui comportent chacune 80 à 110 dents fines, semblables, très pointues et légèrement recourbées [...] » (*ZooCor*).

Cette description est essentiellement composée d'une prédication sur le sémantème 'dauphin'. Nous nous posons donc la question de savoir s'il s'agit d'une collocation du terme *dauphin* plutôt que d'une locution ? Toutefois, si nous réfléchissons à la définition, qui se base sur des connaissances encyclopédiques, nous remarquons que celle-ci comprend plus que juste 'dauphin' 'bleu' et 'blanc' et que d'autres caractéristiques sont prises en considération par les spécialistes. Son sens n'est donc pas compositionnel car il ne dénote pas une entité, mais un type d'entité ayant des caractéristiques bien précises. En d'autres termes, le pivot sémantique de *dauphin bleu et blanc* n'est pas 'dauphin', mais bien 'type de dauphin', ce qui lui donnerait le statut de locution faible plutôt que de collocation de *dauphin*. Il en va de même pour *grand dauphin* qui désigne un 'type de dauphin' et pour *espèce paravent*, *espèce parapluie*, *espèce-ingénieur*, *espèce clé-de-voûte*, etc. qui dénotent des 'types d'espèce'.

En dépit de ce que nous venons de dire, certains syntagmes pourraient être décrits d'un point de vue lexicographique à la fois comme collocations et comme locutions faibles. Un exemple est illustré par l'UT *bébé tortue* dont le sens est exprimé de façon compositionnelle : 'tortue' est exprimé par la base *tortue* et le reste du sens 'très jeune animal'/'nouveau-né' par le collocatif *bébé*. Situé à l'interface de la langue générale et de la terminologie, ce syntagme dénote l'état initial des stades de vie (bébé, juvénile, jeune, subadulte et adulte) de la lexie L. On pourrait tout d'abord le décrire en tant que collocation du terme *tortue* via le collocatif nominal Germ, une fonction lexicale standard qui désigne l'état initial/l'origine de L : Germ(*tortue*)=*bébé tortue*. Cependant, on pourrait aussi considérer deux choses : la première est que le sens de *bébé tortue* dénote avant tout un 'type de tortue' ayant des spécificités bien précises dont la richesse et la complexité du sens ne sont pas incluses dans le sens de Germ. La seconde

---

langues de spécialité. Il les classe parmi les types de phrasèmes, comme une sous-classe des clichés linguistiques.

est que dans le domaine de spécialité en question, il existe une locution faible *bébé tortue* dont le pivot sémantique est ‘type de tortue de taille petite nécessitant des soins particuliers’ (et pas seulement ‘très jeune animal’). Dans ce dernier cas, on pourrait éventuellement modéliser le lien via une nouvelle fonction lexicale non standard nommée *Petit\_de*. Il s’agit, bien sûr, d’un cas singulier, mais très récurrent dans notre corpus (*bébé phoque, bébé dauphin*, etc.), ce qui nous oblige à considérer – dans la phase d’identification de la nature collocationnelle ou locutionnelle de ces termes – aussi bien le contexte communicatif que les intentions des locuteurs qui utilisent ces syntagmes.

Ainsi de suite, les termes ont été examinés en fonction des relations sémantico-lexicales à chaque fois qu’elles apparaissaient dans les fiches de *ZooTerm*. Leur relation avec chaque terme vedette ayant été analysée, nous avons pu détecter et corriger les éventuelles anomalies dans l’attribution des fonctions lexicales. Au total, 23 FL différentes ont été utilisées, un nombre jugé suffisant et nécessaire pour examiner les relations entre les termes véhiculant les informations les plus importantes sur les concepts fondamentaux de *ZooTerm*. Toutefois, comme nous le verrons dans les sections suivantes, l’attribution de certaines FL originales a été modifiée. En premier lieu, nous avons simplifié quelques FL, en généralisant certains détails parce que nous ne disposions pas d’un nombre suffisant d’occurrences pour justifier leur emploi. En second lieu, d’autres fonctions ont été créées pour modéliser des relations non prises en considération dans le modèle original, à savoir des relations qui s’avèrent fondamentales pour la description des termes du champ lexical de la faune marine (nom vulgarisé, nom scientifique, etc.), mais que la LEC n’évaluerait pas congruentes pour les UL.

Nous avons examiné trois catégories de relations sémantiques à partir de l’échantillon de termes extraits du corpus *ZooCor* : les relations classiques en terminologie, à savoir la synonymie, l’antonymie et les relations taxonomiques et méronymiques (§ 4.1.1.) ; les relations actanciennes et circonstancielles (§ 4.1.2.) et, enfin, quelques relations syntagmatiques et dérivationnelles (§ 4.1.3.).

#### **4.1.1. Les relations classiques en terminologie**

Dans la phase de modélisation des relations sémantico-lexicales, nous avons appliqué la plupart des FL élaborées dans le cadre de la LEC avec quelques adaptations selon les données contenues dans notre ressource lexicographique et ses finalités.

Les relations sémantiques typiquement employées en terminologie correspondent – dans le modèle original des FL – à *Hyper*, *Hypo*, *Syn* (et autres types de synonymies, tels que  $Syn_{\supset}$ ,  $Syn_{\subset}$ ,  $Syn_{\cap}$ ), *Anti* (et ses nuances  $Anti_{\supset}$ ,  $Anti_{\subset}$ ,  $Anti_{\cap}$ ) et *Contr*. Les étiquettes *Hyper* et *Hypo* (Polguère 2016, 183) se sont

avérées très utiles pour structurer les hiérarchies sémantiques des termes, mais à la suite de L'Homme (2020b) et de Mel'čuk *et al.* (2021), nous avons ressenti le besoin de les remplacer par les FL Gener et Spéc, extrêmement productives dans les séries taxonomiques pour lister les différents termes spécifiques liés à une espèce marine. Par exemple, *tortue rayonnée*, *tortue verte*, *tortue caouanne*, *tortue luth*, *tortue franche*, *tortue olivâtre* et *tortue imbriquée* sont décrits comme des Spéc de *tortue marine* :

Spéc(*tortue marine*)= *tortue rayonnée*, *tortue verte*, *tortue caouanne*,  
*tortue luth*, *tortue franche*, *tortue olivâtre*, *tortue imbriquée*

Quant au phénomène de la synonymie, nous avons remarqué que dans notre corpus la synonymie exacte – Syn(*tortue marine*)=*tortue de mer*, *tortue d'eau salée* – est aussi fréquente que la synonymie approximative. C'est le cas de *tortue marine* (qui vit en mer) qui est un « synonyme à intersection » (Mel'čuk *et al.* 2021, 24) de *tortue océanique* (qui vit dans les océans) et de *tortue aquatique* (qui vit dans l'eau ou à proximité de l'eau). Nous l'avons donc représenté à l'aide de la fonction lexicale suivante :

Syn<sub>∩</sub>(*tortue marine*)=*tortue océanique*, *tortue aquatique*

Pareillement, les termes *tortue terrestre* (qui est adaptée à la vie sur les terres émergées), *tortue lacustre* et *tortue d'eau douce* (qui vit sur les bords ou dans les eaux des lacs) partagent un certain nombre de caractéristiques sémantiques avec le terme *tortue marine*, mais pas toutes, d'où l'étiquette Contr :

Contr(*tortue marine*)=*tortue terrestre*, *tortue lacustre*, *tortue d'eau douce*

Nous avons également ajouté l'étiquette Syn<sub>dia</sub> pour les termes qui se réfèrent à la même réalité mais la considèrent d'un point de vue différent, en fonction du contexte d'usage ou du registre linguistique. À titre d'exemple, *grosse tête* est une variante diatopique de *tortue caouanne* employée en Nouvelle Calédonie pour désigner la même espèce de tortue. Il en va de même pour *tortue luth* dénommée en Guyane *tortue franche*, *tortue-cuir*, etc. et pour *tortue verte* dénommée *cochon de mer*, *tôti blan*, *tôti vé* dans les Antilles françaises :

Syn<sub>dia</sub>(*tortue caouanne*)=*grosse tête*, *caret*, *coffre*, *tortue à bahut*  
Syn<sub>dia</sub>(*tortue luth*)=*tortue franche*, *tortue-cuir*, *tortue cercueil*  
Syn<sub>dia</sub>(*tortue verte*)=*tortue comestible*, *chélonée marine*, *cochon de mer*,  
*tôti blan*, *tôti vé*

En plus, nous avons intégré l'étiquette  $Syn_{Scient}$  pour les variantes synonymiques représentées par les noms scientifiques des espèces. À titre d'exemples :

$Syn_{Scient}(orque\ naine)=Feresa\ attenuata$   
 $Syn_{Scient}(globicéphale\ tropical)=Globicephala\ macrorhynchus$

Pour les relations méronymiques, nous nous sommes appuyés sur les FL Mult ('ensemble régulier de') et Sing ('unité minimale régulière de') pour saisir des relations telles que celles partagées par les termes *tortue marine* et *population de tortues marines*, *dauphin* et *troupe de dauphins*, *bande de dauphins* et *groupe dense*. Nous avons proposé une simplification des deux étiquettes en les remplaçant par Coll et Part – cette dernière suggérée par Fontenelle (2014) – pour représenter des relations partagées par des parties et des ensembles :

$Coll(tortue\ marine)=population\ de\ tortues\ marines$   
 $Part(tortue\ marine)=carapace, bec, plastron, nageoire, écaille, plaque$

Pour rendre compte des phases chronologiques liées au processus de reproduction des tortues marines, nous avons créé l'étiquette Phase suivie par une numérotation, en fonction des stades concernés. Par exemple, la *parade nuptiale* est une Phase<sub>1</sub> de la reproduction des tortues marines :

$Phase_1(reproduction)=parade\ nuptiale$   
 $Phase_2(reproduction)=accouplement$   
 $Phase_3(reproduction)=inter-ponte$   
 $Phase_4(reproduction)=ponte$   
 $Phase_5(reproduction)=nidification$   
 $Phase_6(reproduction)=incubation\ des\ œufs$

#### **4.1.2. Les relations actancielles et circonstancielles**

Un grand nombre de relations repérées sont liées aux noms d'actants et de circonstants. Dans le modèle original des FL, les actants sont notés en fonction de leur position (S<sub>1</sub> « nom d'agent », S<sub>2</sub> « nom de patient », S<sub>3</sub> « nom typique de l'ASyntP III, etc.) et des noms typiques des circonstants comme S<sub>instr</sub>, S<sub>med</sub>, S<sub>res</sub>, S<sub>loc</sub>, mais nous avons opté pour une notation plus systématique et plus transparente pour les usagers potentiels de notre ressource lexicographique. Par exemple, le terme prédicatif *protection* (protection faite par X de Y contre Z au moyen de W) appelle les noms d'actants (S<sub>1</sub>) *biologiste marin*, (S<sub>2</sub>) *tortue marine*, (S<sub>3</sub>) *braconnier* et les noms de moyen (S<sub>med</sub>) *plan d'action* et (S<sub>med</sub>) *programme de repeuplement*. Dans notre ressource, nous avons étiqueté le nom

de patient *tortue marine* comme l'« objet visé » (Objet), les noms d'actants *biologiste marin* et *braconnier* comme les agents impliqués respectivement dans les processus de sauvegarde (Agent+) et de menace (Agent-) de l'espèce et les dérivés sémantiques nominaux circonstanciels *plan d'action* et *programme de repeuplement* comme les moyens (Moyen). Bien que notre interprétation des sens soit conforme au modèle original, nous avons préféré utiliser des notations différentes.

D'autres exemples de dérivés sémantiques nominaux actanciels et circonstanciels ont été repérés dans le corpus :

$S_1(\text{braconnage})=\text{braconnier}$   
 $S_1(\text{menace})=\text{homme, pêcheur, prédateur}$   
 $S_1(\text{conservation})=\text{zoologiste, naturaliste, chercheur, vétérinaire, bénévole}$   
 $S_{\text{instr}}(\text{capturer})=\text{engins de pêche, bobèche, filet}$   
 $S_{\text{instr}}(\text{menacer})=\text{hameçon, déchets marins flottants, sac plastique, hélice, engins de pêche, source lumineuse}$   
 $S_{\text{med}}(\text{se nourrir})=\text{méduse, crevette, posidonie, algue}$   
 $S_{\text{res}}(\text{réhabiliter})=\text{guérison, bonne santé}$   
 $S_{\text{res}}(\text{pondre des œufs})^{11}=\text{nouveau-né, hatchling}$   
 $S_{\text{res}}(\text{chasser})=\text{mort, menace, extinction}$   
 $S_{\text{loc}}(\text{tortue marine})=\text{océan, mer, plage, aire marine protégée, corridor biologique sous-marin, aquarium, centre de soins}$

#### 4.1.3. Les relations entre le nom et les autres parties du discours

Dans l'échantillon de termes candidats issus de l'extraction automatique, les relations entre le nom et les autres parties du discours se sont avérées aussi productives que celles décrites dans les sections précédentes.

En ce qui concerne les liens entre le nom et le verbe, la modélisation des FL a été délibérément abandonnée à cause de sa complexité<sup>12</sup>. Nous ne donnons ci-dessous qu'un inventaire de tendances et quelques pistes de classification des verbes regroupés en trois catégories :

(1) les dérivés syntaxiques  $V_0$  (*ponde=pondre, éclosion=éclore, nidification=nicher, émergence=émerger, plongée=plonger, respiration=respirer, etc.*) ;

(2) les verbes correspondant aux actions faites par l'espèce marine (*~ se déplacer, ~ migrer, ~ nager, ~ se reproduire, ~ survivre, ~ se nourrir, ~ s'alimenter, ~ s'accoupler, ~ nidifier, ~ disparaître, ~ s'éteindre, etc.*) ;

<sup>11</sup> Très intéressant ce cas de pléonasme modélisable via la fonction Redun proposée par Polguère : Redun(*pondre*)=*pondre les œufs*.

<sup>12</sup> Une représentation de ces données, basée sur les principes théoriques de la sémantique des cadres (Fillmore 1982 ; Faber *et al.* 2014), pourrait avoir une force explicative et modélisatrice très puissante.

(3) les verbes associés aux actions typiquement liées aux activités de conservation et de protection de l'espèce marine (*rétablir* ~, *restaurer* ~, *protéger* ~, *conserver* ~, *préserver* ~, *soigner* ~, *mesurer* ~, *suivre* ~, *sensibiliser* ~, *géolocaliser* ~, etc.) ;

(4) les verbes associés aux actions typiquement liées aux activités de menace et d'extinction de l'espèce marine (*détruire* ~, *tuer* ~, *chasser* ~, *menacer* ~, *capturer* ~, *pêcher* ~, *braconner* ~, *prendre* ~, etc.).

Cette schématisation semblerait être beaucoup plus simple et lisible pour des usagers non-experts de la ressource, censés se fier à leurs compétences linguistiques, contrairement au modèle de la LEC, qui peut être utilisé à des fins d'encodage chez les chercheurs.

Nombreux sont les adjectifs, tels que *commun*, *gris*, *blanc*, *petit*, *vert*, *franche*, etc. qui prennent des significations spécialisées lorsqu'ils sont associés à des bases nominales extrêmement spécifiques (*phoque commun*, *phoque grise*, *ours blanc*, *petit dauphin*, *tortue verte*, *tortue franche*, *nouveau-né*, *grosse tête*, *grand dauphin*, etc.) et qui mériteraient une réflexion plus approfondie. Pour des raisons d'espace, l'analyse des adjectifs n'a pas été retenue dans cette étude. Toutefois, nous avons commencé à réfléchir sur les critères de sélection, parmi lesquels celui de l'indissociabilité du terme : ainsi, si *vitellin(e)* ne s'associe qu'à *sac*, *canal*, *membrane* et *vésicule*, alors *sac vitellin* est considérée comme une UT, contrairement aux syntagmes libres *tortue charismatique*, *tortue mythique*, *tortue sauvage*, etc. qui s'associent à un plus grand nombre de substantifs.

## 5. Conclusions

Dans cette étude nous avons mis en lumière l'importance des corpus et des lexiques de spécialité pour la diffusion des connaissances et des terminologies dans le domaine émergent de la faune marine. Bien qu'expérimentale, la méthodologie adoptée pour la constitution du corpus *ZooCor*, nous a semblé particulièrement utile pour les finalités du projet *Littératie océanique*. La consultation d'experts du domaine – aussi bien académiques que professionnels – a permis, d'un côté, de cartographier les besoins linguistiques et extralinguistiques des potentiels usagers et, de l'autre, d'identifier les domaines, les genres textuels et les pratiques communicatives qui constituent souvent un obstacle pour le citoyen, désireux d'approcher ce monde dans une perspective de science ouverte, cette dernière étant aujourd'hui de plus en plus fréquente et requise dans les projets européens (Jacobi *et al.* 1988 ; Ledouble 2023). Ainsi, le corpus et la ressource lexicographique en cours d'élaboration pourraient constituer un outil précieux pour la promotion du patrimoine naturel marin et des pratiques professionnelles et bénévoles liées à la sauvegarde de la faune marine.

En outre, conformément aux objectifs purement linguistiques, nous avons expérimenté le recours au cadre théorique de la LEC. Cette expérimentation a permis d'explicitier les différences d'encodage entre l'approche terminologique et l'approche lexicographique. Certes, nous sommes basés sur un cadre lexical extrêmement spécifique, mais une grande partie de ce qui a été testé dans cette étude pourrait s'appliquer à d'autres modèles formels, à tout le moins pour ce qui est de la terminologie d'un *domain-sensitive*. Dans certains cas, la description d'UL liées à un sous-domaine particulier nous a obligés de choisir entre l'approche conceptuelle et l'approche lexico-sémantique. Même en essayant de combiner les deux, nous avons dû écarter de nombreux traits lexicaux et plusieurs choix méthodologiques ont été faits pour l'articulation entre les classes de termes et pour la modélisation lexicographique de certains liens sémantiques (polysémie, dérivés nominaux actanciels et circonstanciels, FL complexes, création de nouvelles étiquettes pour la structuration des connaissances spécialisées relevant du domaine, etc.).

Toutefois, les FL ont permis de modéliser un nombre majeur de relations que celles de l'approche terminologique classique, étant étroitement liées à la définition actancielle du terme, à la différence des définitions terminologiques qui répondent à des questions sur la nature des choses (quoi, où, quand, comment, etc.). À ce propos, nous avons rencontré un problème émergent : les noms des espèces marines et les activités liées pour leur conservation dépendent de la culture et du lieu, et leurs définitions varient conséquemment. Si l'approche lexicale a contribué à réduire l'imprécision définitoire, l'attribution des FL et des rôles sémantiques a été souvent risquée, en raison du flou terminologique : parfois, elle aurait été pratiquement impossible sans la confrontation avec les experts du domaine. Dans ces cas, les structures syntaxiques se sont révélées moins fiables que les connaissances encyclopédiques et technoscientifiques.

De même, l'analyse exploratoire a fait apparaître plusieurs questions théoriques fondamentales liées aux notions de terme, locution faible taxinomique, collocation, compositionnalité et non compositionnalité des entités terminologiques de la faune marine, qui mériteraient d'être approfondies dans le futur. Pour ce faire, outre la perspective de la compositionnalité sémantique, il serait nécessaire d'adopter la perspective du locuteur et d'insister sur l'importance de la communauté discursive (et ses intentions communicatives) utilisant ce lexique situé à la charnière entre langue générale et langue de spécialité.

En dépit de quelques limites, les FL s'avèrent – nous semble-t-il – un outil efficace pour saisir les relations sémantico-lexicales entre les termes de ce domaine : elles contribuent à enrichir la description des propriétés lexicales des UT et offrent un moyen pour mieux interpréter les liens paradigmatiques et syntagmatiques. Il nous reste à comprendre si la fusion des approches

terminologique et lexicologique pourrait aider à structurer formellement et sémantiquement les différentes facettes des termes et des concepts du domaine de la faune marine (à partir d'un échantillon de données plus vaste) et, plus généralement, vérifier si une telle combinaison d'approches serait bénéfique et offrirait de nouvelles perspectives pour les projets terminographiques émergents.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- Altmanova, Jana, Maria Centrella et Katherine Russo. 2018. *Terminology & Discourse/ Terminologie et discours*. Bern : Peter Lang.
- Binon, Jean, Serge Verlinde et Jan Van Dyck. 2000. *Dictionnaire d'apprentissage du français des affaires*. Paris : Didier.
- Bonadonna, Maria Francesca. 2020. « Using lexical functions to describe adjectives in terminography ». *Terminology* 26, no. 1: 7-32.
- Bowker, Lynne et Jennifer Pearson. 2002. *Working with specialized language : a practical guide to using corpora*. London : Routledge.
- Branca-Rosoff, Sonia. 1999. « Types, modes et genres : entre langue et discours ». *Langage & Société* 87, 5-24.
- Dancette, Jeanne. 2004. *Dictionnaire analytique de la mondialisation et du travail*. Montréal : Les Presses de l'Université de Montréal.
- Drouin, Patrick. 2003. « Term extraction using non-technical corpora as a point of leverage ». *Terminology* 9, no. 1 : 99-117.
- Faber, Pamela et Marie-Claude L'Homme. 2014. « Lexical semantic approaches to terminology : an introduction. *Terminology* 20, no. 2 : 143-150.
- Fillmore, Charles. 1982. « Frame Semantics ». *Linguistics in the Morning Calm*. Seoul : Hanshin Publishing Co., 111-137.
- Fontenelle, Thierry. 2014. *Turning a Bilingual Dictionary into a Lexical-Semantic Database*. Berlin/Boston : De Gruyter.
- Frassi, Paolo. 2021. « DIACOM-FR, une base de données terminologiques de type diachronique ». *Cahiers de lexicologie* 118, no. 2 : 23-49.
- Grimaldi, Claudio. 2017. *Discours et terminologie dans la presse scientifique française (1699-1740). La construction des lexiques de la botanique et de la chimie*. Bern : Peter Lang.
- Ingresso, Francesca et Alain Polguère. 2015. « How terms meet in small-world lexical networks : the case of chemistry terminology ». In *Proceedings of the 11<sup>th</sup> Conference Terminology and Artificial Intelligence*, 167-171.
- Jacobi, Daniel et Bernard Schiele. 1988. *Vulgariser la science. Le procès de l'ignorance*. Seyssel : Champ Vallon.
- Kilgariff, Adam *et al.* 2008. « GDEX: Automatically Finding Good Dictionary Examples in a Corpus ». In *Proceedings of the 13<sup>th</sup> EURALEX International Congress*, 425-432.



- L'Homme, Marie-Claude. 2008. « Le DiCoInfo. Méthodologie pour une nouvelle génération de dictionnaires spécialisés ». *Traduire* 217 : 78-103.
- L'Homme, Marie-Claude. 2009. *DiCoEnviro. Le dictionnaire fondamental de l'environnement*. <http://olst.ling.umontreal.ca/dicoenviro/manuel-DiCoEnviro.pdf>.
- L'Homme, Marie-Claude. 2020a. *La terminologie : principes et techniques*. Montréal : Les Presses de l'Université de Montréal.
- L'Homme, Marie-Claude. 2020b. *Lexical semantics for terminology*. Amsterdam/Philadelphia: John Benjamins.
- Ledouble, Hélène. 2023. *Médiatisation de la science et diffusion des connaissances*. Londres : ISTE.
- León-Araúz, Pilar, Arianne Reimerink et Pamela Faber. 2019. « EcoLexicon and by-products : integrating and reusing terminological resources ». *Terminology* 25, no. 2 : 222-258.
- Lerat, Pierre. 2016. *Langue et technique*. Paris : Hermann.
- Mel'čuk, Igor et al. 1984-1999. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexicos-sémantiques I-IV*, Montréal : Les Presses de l'Université de Montréal.
- Mel'čuk, Igor et Alain Polguère. 2007. *Lexique actif du français*, Bruxelles : De Boeck.
- Mel'čuk, Igor et Alain Polguère. 2021. « Les fonctions lexicales dernier cri », In *La Théorie Sens-Texte. Concepts-clés et applications*, édité par Sébastien Marengo, 75-156. Paris : L'Harmattan.
- Mel'čuk, Igor et Jasmina Miličević. 2014. *Introduction à la linguistique 1*, Paris : Hermann.
- Mel'čuk, Igor, André Clas et Alain Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.
- Mel'čuk, Igor. 1997. *Vers une linguistique Sens-Texte. Leçon inaugurale (10 janvier 1997)*. Collège de France. Chaire internationale. <http://olst.ling.umontreal.ca/pdf/melcukColldeFr.pdf>.
- Mel'čuk, Igor. 2023. *General Phraseology. Theory and Practice*, Amsterdam/Philadelphia : John Benjamins.
- Meyer, Ingrid. 2001. « Extracting knowledge-rich contexts for terminography. A conceptual and methodological framework », In *Recent Advances in Computational Terminology*, édité par Didier Bourigault, Christian Jacquemin et Marie-Claude L'Homme. 279-302. Amsterdam/Philadelphia : John Benjamins.
- Picoche, Jacqueline et Jean-Claude Rolland. 2001. *Dictionnaire du français usuel*. Bruxelles : De Boeck/Diculot.
- Polguère, Alain. 2003. « Collocations et fonctions lexicales : pour un modèle d'apprentissage ». *Revue Française de Linguistique Appliquée* 1 : 117-133.
- Polguère, Alain. 2016. *Lexicologie et sémantique lexicale. Notions fondamentales*. Montréal : Les Presses de l'Université de Montréal.
- Poudat, Céline, Guillaume Cleuziou et Viviane Clavier. 2006. « Catégorisation de textes en domaines et genres. Complémentarité des indexations lexicale et morpho-syntaxique », *Document numérique* 9, no. 1 : 61-76.

- Rakotonoelina, Florimond. 2014. « Perméabilité des frontières entre l'ordinaire et le spécialisé dans les genres et les discours ». *Les Carnets du Cediscor* 12.
- Sandulli, Roberto. 2011. *Biologia marina*. Milano : McGrawHill.
- Selosse, Philippe. 2012. « Peut-on parler de classification à la Renaissance : les concepts d' « ordre » et de « classe » dans les ouvrages sur les plantes », *Seizième Siècle* 8, 39-56.
- Swales, John. 1990. *Genre Analysis English in Academic and Research Setting*. Cambridge : Cambridge University Press.
- Tillier, Simon. 2005. « Terminologie et nomenclatures scientifiques : l'exemple de la taxonomie zoologique ». *Langages* 157 : 103-116.
- TLFi : Trésor de la langue Française informatisé*. ATILF - CNRS & Université de Lorraine. <http://www.atilf.fr/tlfi>.
- Tutin, Agnès et Francis Grossmann. 2014. *L'écrit scientifique : du lexique au discours. Autour de Scientext*. Rennes : Presses Universitaires de Rennes.
- Wanner, Leo. 1996. *Lexical functions in lexicography and natural language processing*. Amsterdam/Philadelphia : John Benjamins.
- Zanola, Maria Teresa. 2018. *Che cos'è la terminologia*. Roma : Carocci.
- Zanola, Maria Teresa. 2023. « La formation des phrasèmes terminologiques entre diachronie et variation ». In *Phraséologie et terminologie*, édité par Paolo Frassi, 47-58, Berlin/Boston : De Gruyter.