

### CAPITOLO 3. QUADRO TEORICO-METODOLOGICO PER L'ANALISI DEL LESSICO SPECIALISTICO

#### *Abstract*

Nel terzo capitolo – attraverso una rassegna delle principali teorie di nostro riferimento – vengono illustrate le opzioni metodologiche offerte dalla tradizione terminologica, che – come sottolinea Zanola – si articolano, sia in chiave “[...] puramente onomasiologica” per la standardizzazione dei linguaggi specialistici, sia “[...] in un'altra direzione, altrettanto fruttuosa, quella semasiologica”<sup>137</sup>, per la disambiguazione dei termini in relazione al dominio di appartenenza, attraverso l'osservazione e l'analisi di testi specialistici.

Tuttavia, la presente ricerca non adotta in modo esclusivo nessuna delle due impostazioni, privilegiando invece un approccio integrato che ha permesso di raggiungere gli obiettivi di ricerca attraverso l'indagine, l'analisi e la valorizzazione del lessico relativo al patrimonio naturalistico marino, nonché mediante la sua standardizzazione in chiave multilingue, in un'ottica di promozione turistica del territorio.

A tale scopo, il lavoro ha approfondito principalmente l'applicazione della Lessicologia Esplicativa e Combinatoria (LEC) e l'utilizzo delle Funzioni Lessicali (FL), come strumenti per descrivere in modo esaustivo le relazioni semantiche e combinatorie dei termini (Mel'čuk, Clas & Polguère, 1995; Mel'čuk & Polguère, 2021). Questa metodologia si è rivelata innovativa per la creazione di un sistema di rappresentazione del lessico specialistico che non solo ne cattura la complessità, ma ne facilita anche la comprensione e l'accessibilità per diverse categorie di utenti, inclusi i non esperti, come dimostrato nell'applicazione al comparto turistico (§ capitolo 5).

---

<sup>137</sup> Zanola M.-T., *Cos'è la terminologia*, Carocci, Roma, 2018, pp. 32-33.

### 3.1. La terminologia, dalla *langue de spécialité* alla divulgazione scientifica

Al giorno d'oggi, il mondo è abitato da una « société du savoir »<sup>138</sup>, ovvero da una popolazione assetata di apprendimento, in cui la maggior parte degli individui nutre costantemente il bisogno di possedere delle conoscenze specialistiche, che, inoltre, sempre più frequentemente caratterizzano aspetti della vita quotidiana.

La terminologia rappresenta la porta d'accesso principale ai molteplici ambiti del sapere, assicurando precisione concettuale ed appropriatezza espressiva nella comunicazione tecnica e scientifica. In particolare, come indicato e stabilito dal *Groupe de travail Terminologie de la Conférence des Services de traduction des États européens (CST)*:

Le terme « terminologie » désigne en premier lieu le vocabulaire des langues de spécialité (vocabulaire spécialisé) ; il désigne également la science qui étudie, d'une part, les notions et leurs dénominations dans le cadre des vocabulaires spécialisés (étude théorique) et, d'autre part, les méthodes propres au travail terminologique.<sup>139</sup>

Essa, quindi, non solo facilita l'identificazione e la denominazione di oggetti, operazioni e processi, ma al contempo documenta l'evoluzione tecnico-scientifica e definisce i campi di conoscenza, come riportato da Zanola da un punto di vista semasiologico:

La terminologia [...] documenta lo sviluppo tecnico e scientifico, lo comunica e favorisce l'accesso ai domini, permette l'esattezza concettuale e la pertinenza dell'espressione nella comunicazione tecnica e scientifica, contribuendo alla chiarezza nella denominazione degli oggetti, delle operazioni, dei processi astratti e concreti, teorici e pratici che siano.<sup>140</sup>

---

<sup>138</sup> Meyer I., Mackintosh K., « L'étirement du sens en terminologie : aperçu du phénomène de la déterminologisation », in Béjoint H., Thoiron Ph. (a cura di), *Le sens en terminologie*, Presses universitaires de Lyon, Lyon 2000, p. 201.

<sup>139</sup> Chancellerie fédérale: Section de terminologie, *Terminologie*, Berne 2014.

<sup>140</sup> Zanola 2018, *op. cit.*, p. 33.

L'intero patrimonio terminologico, infatti, è fruibile da diverse categorie di utenti ed ogni professione arricchisce il proprio bagaglio linguistico attraverso neologismi, prestiti, calchi o processi di derivazione e ridefinizione semantica: si può quindi affermare che un uso appropriato della terminologia sia strettamente legato all'efficacia della comunicazione<sup>141</sup>, questo perché gli esperti sono chiamati a presentare i risultati delle loro ricerche ed a spiegare le metodologie e le conseguenze dei loro lavori ad un pubblico ampio e variegato, che va dal professionista del settore al « grand public »<sup>142</sup>.

Tale processo, noto anche come “divulgazione scientifica”, rappresenta “l'adattamento delle conoscenze tecniche e scientifiche al fine di renderle comprensibili ad un pubblico non specializzato”<sup>143</sup>, attraverso il bilanciamento tra la libera accessibilità delle informazioni e la loro corretta diffusione: esistono infatti vari canali di comunicazione divulgativa, tra cui periodici generali, riviste scientifiche, programmi televisivi, mostre, podcast, interviste, ed altri mezzi, che denotano la crescente interdisciplinarietà della sua attività discorsiva, collocando le differenti produzioni che ne scaturiscono in una dimensione-multidimensionale<sup>144</sup>. La concezione di una divulgazione « [...] insaisissable comme écriture et comme genre discursif »<sup>145</sup>, suggerisce l'eterogeneità delle conoscenze che questa si prefissa di diffondere<sup>146</sup>, di conseguenza ridurla al mero aspetto scientifico escluderebbe tipi di conoscenza diversi, bensì è fondamentale considerare anche altri aspetti come quello sociale e formale, in cui la divulgazione si basa su una scommessa, quella della « dissolution discursive du jargon technique »<sup>147</sup>.

In quest'ottica, la divulgazione non si realizza soltanto attraverso i canali mediatici o le forme testuali che la veicolano, ma prende corpo anche negli strumenti linguistici che ne rendono possibile l'attuazione, rendendoli a pieno titolo delle

---

<sup>141</sup> Zanola 2017, *op. cit.*

<sup>142</sup> Delavigne V., «Quand le terme entre en vulgarisation», in *Terminologie et Intelligence artificielle*, Strasbourg 2003, p. 80.

<sup>143</sup> “Divulgazione scientifica”, *Le Robert. Dico en ligne*, consultato nel 2025.

<sup>144</sup> Delavigne 2003, *op. cit.*

<sup>145</sup> Beacco 2000, in Delavigne 2003, *op. cit.*, p. 83.

<sup>146</sup> Tuttavia, nonostante siano strettamente legate, la trasmissione di conoscenze si differenzia dalla più precisa divulgazione scientifica per l'utilizzo di strategie diverse, come documentato dallo studio sulla *popular narrative* di Pilkington 2019, p. 582.

<sup>147</sup> Delavigne 2003, *op. cit.*, p. 84.

forme di divulgazione: essi intervengono direttamente sul linguaggio specialistico, attraverso un'azione che si configura non soltanto, dunque, come descrizione scientifica, ma come pratica di mediazione tra sapere esperto e pubblico non specializzato. In tale prospettiva si inseriscono le risorse linguistiche sviluppate ed in corso d'opera nel nostro studio – le reti lessicali, le schede terminografiche ed i glossari multilingue – per la cui realizzazione è parso necessario e fondamentale definire anche le relazioni semantico-lessicali che intercorrono tra i termini estratti dal corpus specialistico: di conseguenza, in linea con gli obiettivi del presente elaborato, tra i quali la promozione turistica del patrimonio naturale, abbiamo sperimentato la metodologia della Lessicologia Esplicativa e Combinatoria (LEC)<sup>148</sup>, per consentire una descrizione strutturata e relazionale del lessico della fauna marina. Ciò si è concretizzato attraverso l'utilizzo delle Funzioni Lessicali per l'esplicitazione delle complesse relazioni semantico-lessicali che i singoli termini raccolti (come *blanchissement*, *phoque*, *capture*) intrattengono con altri: si pensi ad esempio al termine *phoque*, per il quale si possono indicare rispettivamente gli habitat (*grotte rocheuse*, *océan atlantique*), i comportamenti (*allaiter le bébé-phoque*, *chasse benthique*), un potenziale pericolo (*chasse commerciale*) e la categoria iperonimica (*mammifère*). Inoltre, la LEC permette di esplicitare combinazioni tipiche, come *blanchissement de corail*, *aire marine protégé*, *espèce marine vulnérable*, ovvero espressioni che possono essere poi riutilizzate in materiali divulgativi in lingua semplice o in traduzioni adattate.

Il nostro approccio, che applica la LEC al lessico estratto da un corpus specialistico, ci ha permesso di coniugare rigore terminologico e funzionalità applicativa, constatando la fattibilità dell'edificazione di risorse lessicografiche (anche multilingue) controllate, coerenti ed efficaci per i progetti di *citizen science* (che possono così beneficiare di lessici semplificati e contestualizzati), per la redazione di pannelli esplicativi nei siti naturali (*i.e.* “scheda divulgativa di *phoque*” § capitolo 5), per la formazione mirata di guide turistiche naturalistiche e per la creazione di itinerari didattici e/o digitali (come app multilingue per le visite nelle aree marine protette).

---

<sup>148</sup> Mel'čuk I., *et al.*, *Introduction à la lexicologie explicative et combinatoire*, Duculot, Louvain-la-Neuve 1995.

### 3.2. Lessicologia Esplicativa e Combinatoria (LEC): un modello per l'analisi terminologica

Nel nostro studio, inserendoci nel dialogo tra terminologia e lessicologia<sup>149</sup>, abbiamo constatato la fruttuosità della Lessicologia Esplicativa e Combinatoria (LEC) che, come sottolineato da Mel'čuk *et al.*, nasce nell'ambito della lessicologia generale e, più nel dettaglio, della lessicografia:

Disons ici que *explicatif* dans son titre signifie que tout élément lexical qui y est consigné est infailliblement accompagné d'une EXPLICATION SEMANTIQUE formelle, de sorte que les représentations sémantiques sont sous-jacentes au reste de la description lexicographique dont est munie chaque unité lexicale. L'adjectif *combinatoire*, quant à lui, indique que [...] accorde une valeur particulière à la présentation rigoureuse et exhaustive de la COMBINATOIRE LEXICALE.<sup>150</sup>

Tuttavia, la LEC offre un modello formale e coerente per analizzare il comportamento sintattico e semantico delle unità lessicali che si rivela estremamente utile anche per la terminologia, soprattutto nei casi in cui mancano criteri stabili per l'identificazione e la descrizione dei termini nei contesti specialistici: è proprio a partire da questo incrocio disciplinare che si colloca il nostro studio pilota illustrato nel capitolo 5, che mette alla prova l'adattamento del modello LEC nell'analisi dei termini predicativi e semi-predicativi e delle loro strutture attanziali nel linguaggio specialistico.

In questa prospettiva, la realizzazione dei risultati si fonda su una metodologia che, coniugando l'osservazione empirica e la formalizzazione descrittiva dei dati, si sviluppa attraverso una serie di passaggi che mirano a definire con precisione i singoli elementi linguistici nel contesto specialistico, in particolare:

---

<sup>149</sup> Una delle differenze più evidenti tra le due branche – come sottolinea Ledrew (1997, p. 29) – risiede nel modo in cui il significato viene studiato: se da un lato i terminologi indagano il significato dei termini *in vivo* – cioè nel discorso naturale ed autentico – d'altro canto i lessicografi fissano tutti i significati di una parola per studiarla *in vitro*, ovvero spesso in modo più astratto o decontestualizzato.

<sup>150</sup> Mel'čuk *et al.*, 1995, *op. cit.*, p. 10.

- i ruoli semantici e la classificazione delle unità predicative, non predicative e quasi-predicative (§ 3.4.2.);
- la modellizzazione della struttura attanziale (§ 3.4.3.);
- il trattamento di collocazioni nei linguaggi settoriali (§ 3.4.4.);
- l'impiego delle funzioni lessicali per rappresentare le relazioni lessico-semantiche (§ 3.4.5.);
- la costruzione di reti terminologiche basate sulle funzioni lessicali, utile alla rappresentazione sistemica dei dati lessicali (§ 3.4.5.).

Questa articolazione metodologica consente di avvicinare l'analisi terminologica alla profondità descrittiva offerta dalla LEC, mantenendo al tempo stesso un ancoraggio empirico nei testi e nei corpora specialistici.

### 3.2.1. Il predicato semantico e le unità predicative, non predicative, quasi-predicative

La disamina delle unità linguistiche nei linguaggi settoriali richiede un'attenta considerazione della loro natura e del loro ruolo nel contesto specialistico. Tradizionalmente, la terminologia si è concentrata prevalentemente sui sostantivi che denotano entità, trascurando aggettivi e verbi. Tuttavia, la questione di quando un'unità predicativa (verbo o aggettivo) possa essere considerata un termine è di fondamentale importanza, come evidenziato da L'Homme, in quanto:

[...] the strategy normally employed by terminologists to decide whether a unit can qualify as a term (by defining its position in a knowledge structure specific to a domain) cannot be applied to parts of speech such as verbs or adjectives.<sup>151</sup>

Infatti, sebbene i terminologi definiscano un termine in base alla sua posizione in una struttura di conoscenza, questo criterio è meno applicabile a verbi o aggettivi che non si riferiscono a entità concrete, come suggerito anche da Sager:

---

<sup>151</sup> L'Homme M.-C., «Predicative lexical units in terminology», in A. Rossi (a cura di), *Language Production, Cognition, and the Lexicon*, Springer International Publishing, Cham, 2014, p. 80.

Concepts represented in terminological dictionaries are predominantly expressed in the form of nouns; concepts which are linguistically expressed as adjectives and verbs in technical languages are frequently found only in the corresponding noun form and some theorists deny the existence of adjective and verb concepts.<sup>152</sup>

L'Homme propone due criteri principali per identificare le unità predicative come termini:

1. se gli argomenti (o “attanti”, § paragrafo 3.4.2.) di un verbo sono già riconosciuti come termini specifici in un dato dominio, è probabile che anche il verbo stesso sia un termine: l'autrice adotta l'esempio, nel settore del ciclismo, del verbo “pedalare” (o in inglese “to ride”<sup>153</sup>), che può essere considerato un termine se i suoi argomenti, come “ciclista” e “bicicletta”, sono termini tecnici del campo;
2. se un verbo è collegato morfologicamente e semanticamente ad alcune unità lessicali già definite come termini, allora è anch'esso un termine. L'autrice usa come esempio il verbo “frenare” (in inglese “to brake”<sup>154</sup>): se il sostantivo “freno” (“brake”) è un termine, allora anche il verbo correlato lo sarà.

Quando si analizzano le unità predicative in un contesto terminologico, è essenziale considerare la loro struttura attanziale (vedasi paragrafo successivo), o *argument structure* secondo L'Homme, ovvero i partecipanti obbligatori e facoltativi coinvolti nell'azione o nello stato descritto:

The ordered list of semantic arguments of a lexical unit is called the argument structure. Defining the argument structure of an LU or term consists in stating its obligatory participants and the canonical order in which they appear with respect to the predicative LU.<sup>155</sup>

Questi partecipanti, chiamati attanti e circostanti, vengono etichettati in base a ruoli semantici generali come Agente, Paziente e Strumento (§ paragrafo 2.5.2.).

---

<sup>152</sup> Sager J. C., *A Practical Course in Terminology Processing*, John Benjamins Publishing Company, Amsterdam 1990, p. 58.

<sup>153</sup> L'Homme 2014, *op. cit.*, p. 81.

<sup>154</sup> *Ibidem*.

<sup>155</sup> L'Homme 2020, *op. cit.*, p. 141.

Il ruolo dello Strumento, sebbene talvolta oggetto di dibattito in linguistica generale, assume rilevanza cruciale in ambiti specialistici dove può essere un partecipante obbligatorio, evidenziando alternanze regolari (es. “L’utente stampa un documento con una stampante laser” vs. “La stampante stampa il file”).

Questa metodologia, che include l’identificazione dell’unità predicativa, la distinzione tra partecipanti obbligatori e non, l’attribuzione dei ruoli semantici e l’annotazione delle funzioni sintattiche, permette di catturare le relazioni tra unità predicative semanticamente correlate. Tale sistema si estende anche alle unità quasi-predicative, suggerendo una flessibilità nella modellizzazione delle relazioni terminologiche.

Per approfondire la natura delle unità linguistiche in contesti specialistici, è utile fare riferimento alla classificazione di L’Homme<sup>156</sup>, che – inserendosi nel quadro teorico esposto da Mel’čuk e Polguère<sup>157</sup> – distingue tra:

- unità predicative (i.e. *X migre de Y à Z*, *X est menacé*, *X s’êteint*): queste unità denotano un fatto, un evento o una situazione che coinvolge dei partecipanti (chiamati “attanti semantici”). Tali partecipanti corrispondono alle “posizioni attanziali” all’interno della struttura semantica del predicato. Per esempio, il verbo “combattere” implica un’azione che coinvolge qualcuno che combatte (attante X) contro qualcun altro (attante Y), con eventuali ulteriori dettagli (ad esempio, il luogo, lo strumento, ecc.).
- unità non predicative (i.e. *animal*, *tortue*): queste unità denotano entità autonome che non presuppongono la partecipazione ad un fatto specifico. Sono considerate “non leganti”, non controllano posizioni attanziali, come nel caso di sostantivi che rappresentano oggetti concreti (i.e. “fiore”, “leone”), non richiedono la presenza di attanti per esprimere il loro significato e non controllano alcuna struttura semantica complessa come avviene per i predicati.
- unità quasi predicative (i.e. *site de X*, *espèce de X*): occupano una posizione intermedia tra le unità predicative e non predicative. Sebbene queste unità denotino entità, non sono puramente descrittive come le unità non predicative e

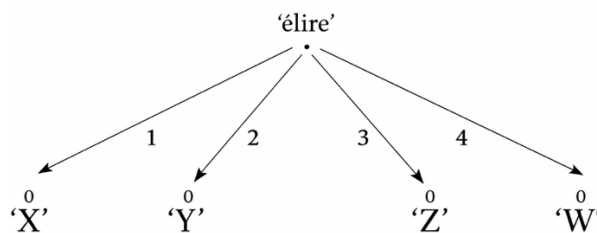
---

<sup>156</sup> L’Homme 2014, *op. cit.*

<sup>157</sup> Mel’čuk I., Polguère A., « Prédicats et quasi-prédicats sémantiques dans une perspective lexicographique », *Lidil. Revue de linguistique et de didactique des langues*, 37, 2008, pp. 99-114.

per comprendere pienamente il loro significato, è necessario tener conto di alcune posizioni semantiche che esse controllano, simili a quelle dei predicati. Ad esempio, il termine “ministro” denota un’entità (una persona), ma questa entità è definita dalla sua funzione all’interno di un contesto specifico: un ministro (X) è responsabile di un certo ambito (Y) all’interno di un governo (Z). Allo stesso modo, “ospedale” non rappresenta solo un luogo, ma un’entità (X) in cui si svolge un’attività (curare Y). Ad esempio, “ministro” denota una persona, ma implica anche relazioni con un ambito di responsabilità e un governo. Le unità quasi predicative sono quindi definite dalle relazioni implicite che instaurano tra diverse entità, anche se non denotano direttamente un fatto o un’azione come i predicati propriamente detti.

IGOR MELČUK ET ALAIN POLGUÈRE



**Fig. 3.1.** Legami dell’unità predicativa “élire” : « X élit l’individu Y au poste Z pour la période W » (Mel’čuk e Polguère 2008, p. 2).

Come ricapitolato nella seguente tabella, la distinzione tra unità predicative, non predicative e quasi-predicative si basa sulla capacità di queste unità lessicali di denotare fatti e di controllare posizioni attanziali: se da un lato le prime denotano fatti e coinvolgono attanti, dall’altro le seconde rappresentano entità senza richiedere attanti, mentre le quasi-predicative, pur raffigurando delle entità, implicano comunque una struttura semantica complessa che include posizioni attanziali<sup>158</sup>.

<sup>158</sup> Effettuare una tale distinzione su un campione ingente di termini (come vedremo nei paragrafi successivi) risulta più agevole ed efficiente all’interno di un’analisi che segue un metodo che combina un estrattore automatico di termini e la convalida umana, in quanto consente di far emergere dal corpus termini di diversa natura, come nomi, ovviamente, ma anche verbi e aggettivi.

Tipo di unità	Caratteristiche principali	Esempi
Unità predicative <i>Prédicats</i> <i>sémantiques</i>	Denotano un fatto, evento o situazione; richiedono partecipanti (attanti semantici) che occupano posizioni attanziali. Possono essere verbi, sostantivi o aggettivi che esprimono relazioni.	<i>X migre de Y à Z ; X est menacé ; X s'éteint</i>
Unità non predicative <i>Noms sémantiques</i>	Denotano entità autonome; non presuppongono partecipanti né relazioni; non controllano posizioni attanziali.	<i>animal ; tortue</i>
Unità quasi predicative <i>Quasi-prédicats</i> <i>sémantiques</i>	Denotano entità ma implicano relazioni con altre (controllano alcune posizioni semantiche).	<i>site de X ; espèce de X</i>

**Fig. 3.2.** Tipologie di unità terminologiche.

L'analisi delle unità predicative, quindi, segue un processo che include l'identificazione dell'unità, la classificazione dei partecipanti come obbligatori o facoltativi, l'attribuzione dei ruoli semantici e l'annotazione delle funzioni sintattiche all'interno della frase: questo approccio (adottato nell'analisi al capitolo 5) permette di comprendere in modo più sistematico come i termini predicativi (e non) interagiscono con il resto della terminologia in un dominio specifico.

### 3.2.2. I ruoli semantici e la struttura attanziale

La comprensione di come i significati lessicali, ed in particolare quelli associati ai predicati, si articolino e richiedano una specifica configurazione di partecipanti, è un concetto centrale nella linguistica moderna. Come già anticipato, il quadro teorico di riferimento per l'analisi della struttura attanziale trova una delle sue formulazioni più influenti nella *Théorie de la lexicologie explicative et combinatoire* (LEC)<sup>159</sup>, la quale propone un modello dettagliato che distingue tra i ruoli semantici dei partecipanti e le loro manifestazioni sintattiche.

Mel'čuk *et al.* introducono due concetti chiave per descrivere la struttura attanziale dei predicati: l'*Actant Sémantique (ASém)* e l'*Actant Syntaxique Profond (ASyntP)*, una distinzione fondamentale per comprendere i diversi livelli di rappresentazione

<sup>159</sup> Mel'čuk *et al.*, 1995, *op. cit.*

degli argomenti. Gli attanti sono entità (persone, oggetti, concetti, et al.) che occupano specifiche posizioni nella struttura semantica di un predicato. Per esempio, nel caso del verbo *chasser*, abbiamo tipicamente un attante che è chi caccia (il soggetto) e un attante che è l'oggetto cacciato (il complemento oggetto); inoltre, gli attanti sono assegnati a ruoli specifici nella struttura semantica del predicato: questi ruoli attanziali (detti anche “ruoli tematici” o “ruoli semantici”) rappresentano la funzione che un attante svolge all'interno dell'evento descritto dal predicato. Analizzando la frase *Les phoques chassent les proies pour leurs petits chiots affamés*, estratta dal corpus, evidenziamo alcuni ruoli attanziali come:

- agente: l'entità che compie l'azione (*Les phoques* sono gli agenti);
- paziente: l'entità che subisce l'azione o viene influenzata da essa (*les proies* sono i pazienti);
- esperiente: l'entità che percepisce o esperisce un evento senza necessariamente compiere un'azione (*leurs petits chiots affamés* sono gli esperienti).
- beneficiario: l'entità che beneficia dell'azione (*leurs petits chiots* sono i beneficiari).

Di conseguenza, un *Actant Sémantique* (ASém) è definito come un ruolo semantico che corrisponde ad una variabile all'interno della definizione di un predicato, ovvero rappresenta l'argomento logico-semantico del predicato, vale a dire il partecipante concettuale necessario per il significato completo dell'unità lessicale, come affermato da Mel'čuk *et al.*:

Un actant sémantique de L est un sens qui correspond à une des variables dans la définition de L ; cette variable représente un argument (au sens logico-sémantique) du prédicat (L) qui est le sens de L.<sup>160</sup>

Solo le *lexies* (unità lessicali) – il cui senso è un predicato – possiedono un ASém (pertanto, nomi come “acqua” o “sole” non ne hanno), che non si limita ai soli verbi

---

<sup>160</sup> Mel'čuk *et al.*, 1995, *op. cit.*, p. 36.

ma si estendono anche a nomi di azione, aggettivi, avverbi, preposizioni e congiunzioni.

Un *Actant Syntaxique Profond* (ASyntP), invece, è un sintagma (un gruppo di parole che funziona come unità sintattica) che esprime, a livello della sintassi profonda, un ASém del predicato. Mel'čuk sottolinea che ogni ASém può essere realizzato attraverso uno o più ASyntP, i quali fungono da espressione sintattica, corrispondente ad uno degli argomenti semantici: « Nous appelons actant syntaxique profond [= ASyntP] de la lexie L un syntagme qui dépend de L syntaxiquement et en exprime un actant sémantique »<sup>161</sup>.

Gli ASyntP sono numerati con cifre romane (I, II, III...), seguendo un principio di assegnazione che vede il *S(ujet) G(rammatical)*<sup>162</sup> corrispondere al primo ASyntP, e gli oggetti diretti e indiretti ai successivi, in ordine di importanza decrescente<sup>163</sup>, come dimostra la tabella semplificativa che segue:

Ruolo sintattico	Definizione	Esempi
Agente <i>ASyntP I</i>	Chi compie l'azione	<i>Les phoques chassent les proies pour leurs petits chiots affamés</i> → Les phoques
Paziente <i>ASyntP II</i>	Chi subisce l'azione	<i>Les phoques chassent les proies pour leurs petits chiots affamés</i> → les proies
Destinatario/Beneficiario/Luogo <i>ASyntP III</i>	A chi/per chi/dove	<i>Les phoques chassent les proies pour leurs petits chiots affamés</i> → leurs petits chiots affamés

**Fig. 3.3.** Ruoli sintattici degli attanti semantici nella sintassi profonda.

La distinzione tra ASém e ASyntP è centrale: il primo si colloca a un livello logico-semantico, è quindi legato alle relazioni logiche intrinseche di un predicato, e al suo significato, ma astratto dalla sua realizzazione sintattica; mentre il secondo ne è la manifestazione strutturale concreta a livello sintattico profondo, dove si decide

<sup>161</sup> Mel'čuk *et al.*, 1995, *op. cit.*, p. 117.

<sup>162</sup> *Ibidem.*

<sup>163</sup> È importante notare che un ASyntP non corrisponde sempre in modo univoco a una funzione sintattica di superficie, potendo talvolta corrispondere a complementi circostanziali o modificatori, ed il concetto di ASyntP è generalizzato per coprire i dipendenti sintattici di lexie di tutte le parti del discorso, purché esprimano un ASém della lexie in questione. (Per gli esempi di applicazione, si rinvia a Mel'čuk *et al.* 1995, pp. 117-118).

come quella relazione semantica verrà strutturata grammaticalmente nella frase, mostrando come quegli stessi partecipanti vengono espressi concretamente.

Successivamente, se da un lato la teoria di Mel'čuk e Polguère<sup>164</sup> ha approfondito ulteriormente la struttura attanziale, focalizzandosi sulla valenza attanziale del predicato, ovvero il numero di attanti che esso richiede<sup>165</sup>, d'altra parte, L'Homme sottolinea l'importanza dei ruoli semantici (o ruoli tematici/*theta roles*<sup>166</sup>), che descrivono la funzione che un attante svolge nell'evento<sup>167</sup>. Sebbene non esista una lista universalmente consensuale, alcuni ruoli sono ampiamente riconosciuti e utilizzati in risorse lessicali e in dizionari terminologici quali:

- agente: l'entità, solitamente animata, responsabile dell'azione o della creazione di un'entità;
- paziente/tema: l'entità che subisce l'azione o è influenzata da essa. Il paziente è specificamente affetto, mentre il tema subisce l'azione senza esserne necessariamente alterato;
- strumento: l'oggetto inanimato utilizzato da un agente per compiere un'azione.
- destinatario/ricevente: il bersaglio dell'azione o l'entità che beneficia dell'azione;
- locazione, sorgente, destinazione: argomenti che esprimono diverse relazioni spaziali rispetto all'unità predicativa.

Già in precedenza è stato possibile notare che l'approccio di L'Homme alla struttura attanziale, pur condividendo la centralità della descrizione dei ruoli argomentali, presenta delle differenze rispetto alla LEC di Mel'čuk: mentre la LEC utilizza variabili (X, Y, Z) e assegna loro etichette semantiche generali, per rappresentare gli argomenti L'Homme tende a enfatizzare l'utilizzo diretto di ruoli semantici (come Agente, Paziente, Strumento) per etichettare gli argomenti, sebbene riconosca l'uso delle variabili come sistema di rappresentazione più neutro:

---

<sup>164</sup> Mel'čuk & Polguère 2008, *op.cit.*

<sup>165</sup> Si distinguono in predicati monovalenti: ( che richiedono un solo attante), predicati bivalenti (che richiedono due attanti), predicati trivalenti (richiedono tre attanti). Si rinvia agli esempi esposti in Mel'čuk & Polguère 2008, *op.cit.*

<sup>166</sup> L'Homme 2020, *op.cit.*, p. 149.

<sup>167</sup> L'Homme sottolinea che la loro importanza è stata evidenziata da Fillmore con il suo concetto di *cases* (casi), dimostrando come diverse funzioni sintattiche di superficie (soggetto, complemento) possano esprimere lo stesso ruolo semantico profondo. (Fillmore 1968, in L'Homme 2020, *op.cit.*, con l'esempio del verbo inglese "break").

For instance, Explanatory Combinatorial Lexicology (ECL) uses the term *actant*, whereas that Frame Semantics prefers the label frame element (and more specifically core frame element when participants are obligatory).<sup>168</sup>

Inoltre, come abbiamo notato nel paragrafo precedente, l’approccio di L’Homme distingue esplicitamente tra unità predicative, quasi-predicative e non-predicative per categorizzare il coinvolgimento delle unità lessicali nella struttura argomentale, invece, la LEC si concentra sulla distinzione tra ASém e ASyntP, analizzando come i ruoli semantici si realizzano a livello sintattico profondo, senza una classificazione per tipi di unità basata sulla loro natura predicativa.

Di seguito riportiamo uno schema delle differenze nella concezione della struttura attanziale delle teorie di riferimento del nostro studio:

CARATTERISTICA	<i>L’Homme (2015, 2020b)</i>	<i>Mel’čuk et al. (1995, 2008)</i>
CONCETTO DI STRUTTURA ATTANZIALE	Insieme di ruoli e funzioni semantiche che i partecipanti assumono per completare il significato di un predicato, organizzati in una griglia di significati. Focus sui ruoli semantici e predicativi.	Descritta attraverso <i>Actant Sémantique</i> (ASém) e <i>Actant Syntaxique Profond</i> (ASyntP). L’ASém è il ruolo semantico necessario per il predicato, l’ASyntP la sua espressione sintattica profonda.
RAPPRESENTAZIONE DEGLI ARGOMENTI	Utilizza prevalentemente etichette semantiche dirette (es. Agente, Paziente).	Utilizza variabili (es. X, Y, Z) per indicare gli “slot” da riempire, specificando poi che sono ruoli semantici (ASém).
CLASSIFICAZIONE DELLE UNITÀ LESSICALI	Distingue tra unità predicative, quasi-predicative e non-predicative in base al loro coinvolgimento nella struttura predicativa.	Non presenta una classificazione simile. Si concentra su come i ruoli semantici si realizzano sintatticamente (ASyntP).
RELAZIONE SEMANTICA-SINTASSI	Utilizza una “griglia di significato” per rappresentare ruoli semantici e le relazioni tra unità predicative e non-predicative. Focus sulla funzione dei partecipanti.	Enfatizza la concretizzazione dei ruoli semantici (ASém) nella sintassi profonda tramite gli ASyntP. Si concentra su come i ruoli semantici si realizzano a livello sintattico e come queste realizzazioni possono variare.
AMBITO DI APPLICAZIONE PREFERENZIALE	Linguistica cognitiva e semantica lessicale, analisi della struttura dei significati predicativi nei testi.	Linguistica computazionale e lessicografia, focalizzata sulla rappresentazione di ruoli semantici

<sup>168</sup> L’Homme 2020, *op.cit.*, p. 140.

**Fig. 3.4.** Differenze negli approcci di L’Homme (2014, 2020b) e Mel’čuk et al. (1995, 2008).

Dunque, mentre entrambe le teorie riconoscono l’importanza della struttura attanziale e dei ruoli semantici, da un lato la LEC di Mel’čuk si focalizza sulla relazione tra i ruoli semantici astratti (ASém) e la loro espressione sintattica profonda (ASyntP), fornendo un quadro robusto per la lessicografia computazionale; L’Homme, d’altro canto, enfatizza la classificazione delle unità in base alla loro natura predicativa e l’applicazione dei ruoli semantici per una descrizione approfondita dei termini in contesti specifici, mantenendo al contempo un ancoraggio empirico nei testi e nei corpus specialistici.

Per concludere, la comprensione della struttura attanziale, sia nella sua distinzione tra ruoli semantici e realizzazioni sintattiche come proposto da Mel’čuk, sia nell’analisi delle categorie predicative di L’Homme, è cruciale per una descrizione lessicografica e terminologica accurata. Questi modelli non solo illuminano le relazioni profonde tra le unità lessicali e i loro partecipanti, ma gettano anche le basi per affrontare la complessità di fenomeni linguistici più articolati.

### 3.2.3. Le Funzioni Lessicali (FL)

Le Funzioni Lessicali (FL) sono uno degli strumenti più innovativi sviluppati dalla Teoria Senso-Testo (TST)<sup>169</sup>, e, come suggeriscono Mel’čuk & Polguère « servent à modéliser l’expression d’un sens donné qui se fait de façon contingente à une expression préalablement sélectionnée d’un autre sens »<sup>170</sup>, ovvero sono state concepite per affrontare le sfide legate alla selezione precisa dei termini ed alla

<sup>169</sup> Marengo S., *La théorie Sens-Texte. Concepts-clés et applications*, L’Harmattan, Paris, 2021.

<sup>170</sup> Mel’čuk & Polguère 2021, *op. cit.*, p. 15.

produzione di combinazioni lessicali appropriate, un aspetto centrale soprattutto nei linguaggi specialistici dove l'accuratezza terminologica è imprescindibile<sup>171</sup>.

Nel quadro teorico della *Zone de combinatoire lexicale* di Mel'čuk, Clas e Polguère<sup>172</sup>, che si prefigge di modellare l'organizzazione e la struttura delle combinazioni lessicali nel linguaggio, le Funzioni Lessicali emergono come lo strumento chiave per descrivere queste ultime, soprattutto se non prevedibili: esse sono concepite come « fonction[s] au sens mathématique »<sup>173</sup> applicate al linguaggio ed il loro obiettivo è rappresentare in modo sistematico le relazioni che determinano le co-occorrenze fisse. Quindi, la notazione per l'applicazione di una FL ( $f$ ) ad un lessema di partenza ( $L$ ) è la seguente:

$$f(L)=L_1, L_2, \dots, L_n^{174}$$

dove  $L$  è l'argomento (il lessema di base) e  $L_1, L_2, \dots, L_n$  sono i valori o i risultati dell'applicazione della funzione, ovvero le parole che si combinano con  $L$  o lo sostituiscono secondo una relazione specifica<sup>175</sup>. Ad esempio, per esprimere l'intensificazione di *lent* (lento) o *ramifié* (ramificato), le FL permettono di formalizzare l'uso di espressioni fisse come *lent comme une tortue* o *ramifié comme un corail*.

Mel'čuk classifica le FL standard come operatori semantico-lessicali che descrivono relazioni ricorrenti e sistematiche tra un lemma e i suoi collocativi, applicabili a un vasto numero di basi, come ad esempio:

---

<sup>171</sup> Inoltre rappresentano lo strumento cardine del *Dictionnaire Explicatif et Combinatoire* (DEC). Mel'čuk, I., *Dictionnaire explicatif et combinatoire du français contemporain IV - Recherches lexico-sémantiques IV*, HORS COLLECTION, Les Presses de L'Université de Montréal, 1999.

<sup>172</sup> Mel'čuk et al., 1995, *op. cit.*, p. 191.

<sup>173</sup> Mel'čuk et al., 1995, *op. cit.*, p. 149.

<sup>174</sup> In Wanner,  $f(L)=\{L_i\}$  dove appunto  $i=1,2,\dots,n$ , (Wanner L., *Lexical functions in lexicography and natural language processing*. Amsterdam/Philadelphia, John Benjamins, 1996) e come sottolineato in Mel'čuk & Polguère 2021, *op. cit.*, p. 3: « [...]  $f$  appliquée à l'unité lexicale  $L$  – ce qui est noté  $f(L)$  – fournit [...] un ensemble d'expressions alternatives ».

<sup>175</sup> L'espressione si rifà proprio alla funzione matematica  $f(x)=y$ , per cui  $y$  è il prodotto del rapporto tra  $f$  e  $x$ . In particolare, nella funzione lessicale la  $x$  rappresenta un lemma (ovvero la “parola chiave”, la *mot-clé* di Mel'čuk 1995, *op. cit.*, p. 127) e la  $y$  è un valore della funzione che esprime un significato specifico se associato ad  $f$ . A differenza della funzione matematica, infatti, una funzione lessicale può avere più di un valore per un dato argomento.

- verbi supporto (OPER, FUNC, LABOR): descrivono azioni generiche legate a un nome<sup>176</sup>;
- verbi di realizzazione (REAL, FACT, LABREAL): indicano la concretizzazione di un'azione o uno stato<sup>177</sup>;
- intensificatore (MAGN): esprime un alto grado della proprietà o dell'azione<sup>178</sup>.

Nello specifico, la definizione di una Funzione Lessicale Standard, come proposta dalla teoria di nostro riferimento, si basa su quattro criteri principali che descrivono come le unità lessicali si combinano in modo prevedibile ma non banale nel linguaggio. Infatti, secondo la definizione di Mel'čuk *et al.*, una funzione lessicale standard è «[...] une fonction f qui associe à une lexie L un ensemble de lexies f(L) [...]»<sup>179</sup>, seguendo le seguenti condizioni:

1. condizione di “proporzionalità”: per due lessie diverse (L<sub>1</sub> e L<sub>2</sub>), le parole generate dalle rispettive funzioni devono mostrare relazioni semantiche e sintattiche simili. Ad esempio, la funzione di intensificazione MAGN, applicata a *nager*, produrrà *nager rapidement*, *nager vigoureusement*, *nager avec force*, mentre applicata a *courant* produrrà *courant puissant*, *courant impétueux*, *courant fort*.

Bien entendu, nous ne disons pas que *comme une Madeleine* et *grosse* sont sémantiquement ou syntaxiquement équivalents ; mais l'expression *comme une Madeleine* remplit par rapport à PLEURER (à peu près) le même rôle que l'adjectif préposé *grosse* par rapport à PLUIE : les deux sont des modificateurs intensificateurs qui signifient – dans le contexte donné – ‘beaucoup’, ‘très’, ‘intense/intensément’.<sup>180</sup>

---

<sup>176</sup> Esempi: OPER(*problème*) = *poser un problème* ; OPER(*conclusion*) = *tirer une conclusion* ; FUNC(*blague*) = *une blague circule* ; FUNC(*décès*) = *un décès survient*. (Mel'čuk, I., *Phraséologie dans la langue et dans le dictionnaire. Repères & Applications*, 6, 2008, p. 8).

<sup>177</sup> Esempi: REAL(*envie*) = *assouvir son envie* ; FACT (*film*) = *le film est à l'affiche* ; (Ibidem).

<sup>178</sup> Esempi: MAGN(*haine*) = *haine aveugle* ; MAGN(*coûter*) = *coûter la peau des fesses* ; (Ibidem).

<sup>179</sup> Mel'čuk *et al.* 1995, *op. cit.*, p. 128.

<sup>180</sup> *Ibidem*.

$$\frac{f(L_1)}{L_1} \approx \frac{f(L_2)}{L_2}$$

Fig. 3.5. Criterio di “proporzionalità” tra i *lexies*<sup>181</sup>.

2. condizione di “non trivialità”: le parole generate dalla funzione per due lessie diverse devono essere differenti. Per esempio, Mel’čuk spiega come il significato di « cher’ [= ‘qui est d’un prix élevé’] »<sup>182</sup> non costituisce una Funzione Lessicale perché si applica uniformemente a qualsiasi sostantivo (*cher voiture*, *cher voyage*), rendendo la relazione trivialmente prevedibile:

L’importance de la Condition 2 peut être illustrée comme suit. Si la fonction *f* considérée donne lieu à des proportions comme celles ci-dessus (donc, si elle satisfait la Condition 1) mais que l’on a toujours le même numérateur pour des dénominateurs différents, une telle dépendance *f* est triviale en *L* : elle n’a pour nous aucun intérêt, puisque le résultat de son application n’est pas une collocation ; nous ne voulons pas que *f* soit retenue comme FL. Ce n’est qu’un simple cas de signification lexicale.<sup>183</sup>

$$\frac{\textit{cher}}{\textit{VOITURE}} = \frac{\textit{cher}}{\textit{VOYAGE}} = \dots$$

Fig. 3.6. Criterio di “proporzionalità” tra i *lexies*<sup>184</sup>.

3. condizione delle *mots-clés*: una Funzione Lessicale deve poter essere applicata a un numero significativo di parole diverse (*arguments/mots-clés*).
4. condizione delle *expressions*: la funzione deve produrre molte espressioni diverse.

<sup>181</sup> Mel’čuk 1995, *op. cit.*, p. 127.

<sup>182</sup> *Ibidem.*

<sup>183</sup> *Ibidem.*

<sup>184</sup> *Ibidem.*

Quant aux Conditions 3 et 4, les exemples cités montrent à l'évidence que le sens 'très', qui correspond à une FL d'intensification, est compatible avec un nombre incalculable de mots-clés et que ces expressions sont aussi fort nombreuses.<sup>185</sup>

Parallelamente, Mel'čuk elenca le FL non standard, che, a differenza delle standard, non sono sistematiche e la loro applicazione è limitata a un numero ristretto di basi a causa del loro senso molto specifico e del loro forte legame con il contesto: vengono utilizzate soprattutto per descrivere le collocazioni non standard e/o i pragmatemi<sup>186</sup> :

[...] l'expression « à vendre », qu'on met sur les pancartes, n'est pas libre et doit être listée. En effet, on ne dit pas \**pour vente* ou \**en vente* (cf. anglais *for sale*) ou bien \**X se vend* (cf. espagnol *Se vende caballo* = *Cheval à vendre*). L'expression « à vendre » comme annonce commerciale ne peut être construite par des règles générales du français, et le DEC se doit de la présenter dans l'article de VENDREI.1 comme un semi-phrasème, au moyen d'une FL non standard.<sup>187</sup>

Anche da quest'ultima definizione si evince quanto gli obiettivi complessivi della LEC (e di conseguenza anche del *Dictionnaire explicatif et combinatoire du français contemporain*) siano chiari: organizzare la lingua in modo che il parlante possa sia selezionare la parola giusta all'interno di un dato paradigma semantico (scelta paradigmatica), sia generare combinazioni lessicali appropriate (scelta sintagmatica) utilizzando le Funzioni Lessicali. Queste ultime, di conseguenza, descrivono due tipologie relazioni lessicali, che Polguère distingue in:

1. relazioni paradigmatiche (legate alla sostituzione di termini nel lessico, come iperonimia, iponimia, sinonimia, meronimia) che « [...] connectent les lexies au sein du lexique par des connexions sémantiques, éventuellement accompagnées de connexions morphologiques »<sup>188</sup>;

---

<sup>185</sup> Mel'čuk 1995, *op. cit.*, p. 129.

<sup>186</sup> « Un pragmatème est une expression lexicale contrainte par la situation de son emploi. » (Mel'čuk 2013, p. 11).

<sup>187</sup> Mel'čuk 1995, *op. cit.*, p. 196.

<sup>188</sup> Polguère 2016, p. 130.

2. relazioni sintagmatiche e predicative (legate alla combinazione di termini, attività e comportamenti) che « [...] lient les lexies au sein de la phrase, selon des affinités combinatoires »<sup>189</sup>.

Ne consegue che le Funzioni Lessicali di tipo paradigmatico<sup>190</sup> mettono in relazione un lessema con altri lessemi che condividono un dato legame semantico e/o relazionale, permettendo di interscambiarsi in un determinato contesto senza alterare il significato fondamentale della frase; essi non generano combinazioni lessicali, bensì recuperano un'alternativa lessicale che può essere usata in sostituzione del *lexie* di partenza.

D'altro canto, le Funzioni Lessicali di tipo sintagmatico e derivazionale<sup>191</sup> operano sulla combinazione delle parole specificando quali co-occorrenze sono idiomatiche e corrette per un dato lessema: esse non possono essere dedotte esclusivamente dalla semantica, o dalla sintassi dei singoli elementi, ma riflettono le convenzioni d'uso della lingua.

Tuttavia, è importante sottolineare l'interconnessione tra FL paradigmatiche e sintagmatiche, come evidenziato da Mel'čuk e Polguère:

Les deux classes de FL [...] ne sont pas séparées par une frontière étanche, car toutes les FL, qu'elles soient paradigmaticques ou syntagmaticques, servent à modéliser le même phénomène linguistique : l'expression contextuellement restreinte d'un ajout de sens 'σ<sup>f</sup>'.<sup>192</sup>

Sebbene le FL paradigmatiche trasformino un lessema in un altro e le FL sintagmatiche associno un co-occorrente al lessema di partenza, queste due classi di funzioni non operano in isolamento, spesso la scelta di una FL paradigmatica per

---

<sup>189</sup> Polguère 2016, p. 130. Tuttavia, è noto che i paradigmi della terminologia convenzionale – che pongono il concetto al centro dell'indagine – hanno condotto i terminologi negli anni a focalizzarsi principalmente sulle relazioni paradigmatiche, a scapito di un insieme di altre relazioni considerate essenziali e pertinenti per la rappresentazione delle lingue di specialità. Come afferma L'Homme: "Predicative lexical units have been largely ignored in terminology for a number of reasons: one of them is the focus on entities viewed as the nodes of knowledge representations; another is the lack of linguistic perspective on the data to be represented" (L'Homme 2014, *op. cit.*, p.1).

<sup>190</sup> Mel'čuk *et al.* 1995, *op. cit.*, p. 129.

<sup>191</sup> Mel'čuk *et al.* 1995, *op. cit.*, p. 136.

<sup>192</sup> Mel'čuk & Polguère 2021, *op. cit.*, p. 15.

una base (*i.e.* un sinonimo) può influenzare le FL sintagmatiche che quella nuova base ammette<sup>193</sup>.

Questa rete di relazioni tra lessemi e le loro combinazioni è il cuore della lessicologia esplicativa e combinatoria, fornendo un quadro robusto per la descrizione del lessico in modo comprensivo e *user-friendly*.

La teoria di Mel'čuk, Clas e Polguère, tramite le Funzioni Lessicali, evidenzia che la combinazione lessicale non può essere spiegata solo da regole sintattiche o semantiche, occorre tener conto delle convenzioni linguistiche specifiche ad ogni lessia e che le FL rappresentano, quindi, uno strumento chiave per descrivere e prevedere queste combinazioni<sup>194</sup>, favorendo una ricca espressività linguistica ed una maggiore precisione, in particolare nei linguaggi specialistici.

---

<sup>193</sup> Questa *parente conceptuelle* si manifesta in almeno 3 modalità distinte: (de) fusione degli elementi di valore, cooperazione nel parafrasaggio, implicazioni mutue (si rinvia a Mel'čuk & Polguère 2021, p. 15).

<sup>194</sup> Nel nostro studio, l'individuazione delle FL attraverso un approccio manuale è stata talvolta arricchita dall'impiego di strumenti informatici volti ad agevolare l'analisi semantico-lessicale dei termini estratti. Nello specifico, si è fatto ricorso a DicoEnviro (L'Homme 2014, *op. cit.*) e SpiderLex (Gaume 2008), i quali offrono sezioni dedicate alla mappatura dei *liens lexicaux* tra i termini. I risultati e le metodologie relative all'utilizzo di tali risorse saranno oggetto di una disamina più approfondita nei capitoli successivi.

## CAPITOLO 4. ESTRAZIONE TERMINOLOGICA DA *ZOOCOR*

### *Abstract*

Il quarto capitolo descrive le scelte metodologiche, le fasi di realizzazione ed i risultati dell'estrazione terminologica automatica condotta dal corpus specialistico *ZooCor*, attraverso il software *TermoStat*.

A seguito di una disamina dei principi teorici, sono state illustrate le fasi interconnesse – tra cui quelle di estrazione (che ha permesso il reperimento di 30.445 unità dal corpus), spoglio, *dépouillement manuel*, classificazione e selezione – volte a garantire un affinamento progressivo dei termini pertinenti al dominio della fauna marina.

Di conseguenza, un campione di 5.000 candidati termini, estratti automaticamente, è stato dapprima sottoposto ad uno spoglio manuale, per poi essere categorizzato in diverse fasce predefinite (*Bons*, *Rumeurs*, *SynScient*, etc.); successivamente è stata condotta un'ulteriore estrazione dalla sezione dedicata al sottodominio PHOQUES DE MER di *ZooCor*, al fine di delineare e validare l'impianto metodologico dell'analisi semantico-lessicale prima su un campione ridotto di termini, e dopo su un numero superiore di unità estratte e selezionate dal corpus *ZooCor* (§ capitolo 5) .

#### 4.1. Estrazione dei termini dal corpus *ZooCor*: criteri metodologici

L'estrazione automatica di termini rappresenta una fase essenziale nella terminologia moderna e può essere definita come il processo computazionale volto ad individuare – all'interno di un testo o di un insieme di testi– parole grafiche o sequenze di parole grafiche che possiedono lo *statut terminologique*<sup>195</sup>.

Nel nostro studio, l'estrazione di termini dalla componente francese del corpus *ZooCor* è stata effettuata tramite il software *TermoStat*<sup>196</sup>: sviluppato dall'« Observatoire de linguistique Sens-Texte » dell'Université de Montréal, lo strumento è progettato per facilitare l'estrazione di termini<sup>197</sup> da un corpus, che deve essere sottoposto al software in formato .txt. Il software provvede alla categorizzazione delle unità estratte secondo caratteristiche definite sulla base di specifici criteri, tra cui predominanza dei termini di natura nominale, complessità dei termini<sup>198</sup>, numero finito di sequenze che possono costituire un termine complesso<sup>199</sup> e soprattutto frequenza e specificità di un'unità rispetto ad un *corpus de référence*<sup>200</sup>.

Nel nostro studio, la selezione dei termini per l'analisi lessicale (Fase 2 § paragrafo 4.1.1.) si è basata principalmente su questi ultimi due criteri principali, che figurano

---

<sup>195</sup> L'Homme 2004, *op. cit.*, p.166.

<sup>196</sup> Drouin 2003. Tuttavia, è ampiamente riconosciuto che l'automazione completa dell'estrazione terminologica è un'impresa complessa, in quanto gli estrattori, nonostante i continui miglioramenti, producono liste di *candidats termes*, ossia parole o sequenze di parole che potrebbero essere unità terminologiche, ma che richiedono un'ulteriore purificazione o arricchimento da parte del terminografo. Tale necessità deriva dalla difficoltà intrinseca per uno strumento informatico di riconoscere la natura terminologica di un'unità in una massa indifferenziata di elementi lessicali, specialmente in domini non familiari.

<sup>197</sup> Questo processo si distingue dal reperimento manuale per l'efficienza e la capacità di gestire quantità di dati altrimenti inaccessibili, in cui il software che automatizza l'estrazione deve « [...] comme le terminographe, prendre des décisions sur la nature des unités lexicales » (L'Homme 2004, *op. cit.*, p.166) basandosi tuttavia su indici di natura computazionale.

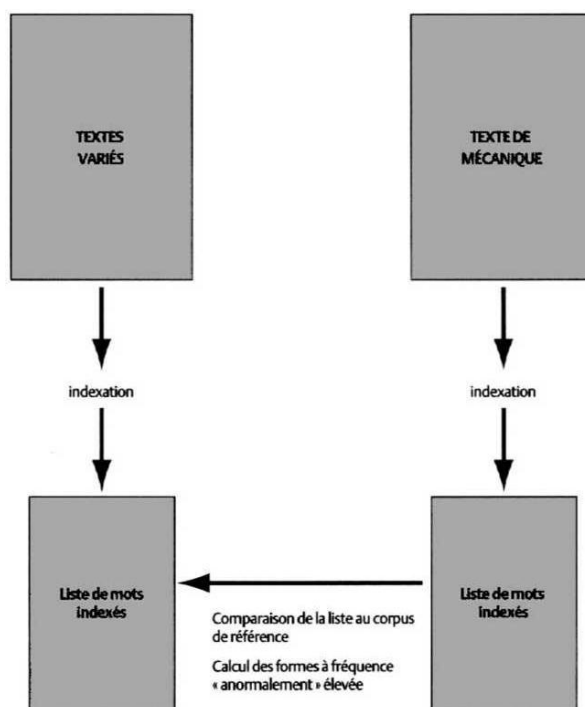
<sup>198</sup> In un testo specialistico, infatti, è solito trovare numerosi termini complessi (*phoque à capuchon, activité de ponté, aire marine protégée*, etc.), di conseguenza molti estrattori si focalizzano sulla ricerca di tali unità, assumendo che esse costituiscano la maggior parte della terminologia di un dominio. Questa tecnica non si basa sul confronto con un corpus di riferimento, ma sulla co-occorrenza interna al corpus stesso (L'Homme, 2004, *op. cit.*, p. 167).

<sup>199</sup> Questo criterio sussiste a causa del numero limitato di sequenze di parti del discorso (*i.e.* nome + aggettivo, nome + nome, nome + preposizione + nome) attraverso cui sono costituiti i termini complessi (L'Homme, 2004, *op. cit.*).

<sup>200</sup> *Ibidem*.

come degli indicatori maggiormente significativi dello status terminologico di un'unità estratta, in quanto, come afferma L'Homme: « Un terme significatif sera normalement utilisé à plusieurs reprises dans un texte spécialisé »<sup>201</sup>.

In particolare, la frequenza indica il numero di occorrenze dei termini all'interno di un corpus specializzato, ma tale misura viene ponderata in relazione alla distribuzione ed alle occorrenze della medesima forma linguistica in un corpus generale più ampio<sup>202</sup>: il principio è che i termini specifici di un dominio mostreranno una frequenza « anormalement élevée »<sup>203</sup> nel testo specializzato rispetto al corpus di riferimento (Fig. 4.1).



**Fig. 4.1.** Elenco prodotto a seguito del confronto tra un corpus di meccanica e un corpus di riferimento<sup>204</sup>.

Ne consegue che tale criterio quantifica la presenza brutta di un termine nel corpus, fornendo un'indicazione della sua rilevanza e centralità nel dominio: un termine con alta frequenza è presumibilmente centrale in un determinato dominio e

<sup>201</sup> L'Homme, 2004, *op. cit.*, p. 167.

<sup>202</sup> Si rinvia agli studi citati per il criterio dimensionale (§ 1.2.).

<sup>203</sup> L'Homme, 2004, *op. cit.*, p. 167.

<sup>204</sup> L'Homme 2004, *op. cit.*, p. 168.

ampiamente utilizzato. Tuttavia, se da un lato la frequenza facilita l'identificazione dei termini più comuni e potenzialmente importanti, dall'altro potrebbe includere termini generici non strettamente specifici del dominio, ma semplicemente molto ricorrenti.

Candidat de regroupement	Fréquence (Spécificité)	Score	Variante orthographiques	Matrice
<b>espèce</b>	21221	198.82	<i>espèce espèces</i>	Nom
<b>marin</b>	10975	147.23	<i>marin marins marines</i>	Adjectif
<b>phoque</b>	6942	119.77	<i>phoque phoques</i>	Nom
<b>corail</b>	6459	115.24	<i>corail coraux</i>	Nom
<b>tortu</b>	6347	114.74	<i>tortu tortue tortues</i>	Adjectif
<b>animal</b>	7344	106.44	<i>animal animaux</i>	Nom
<b>tortue marine</b>	5341	105.54	<i>tortue marine tortue marines tortues marines</i>	Nom Adjectif

Fig. 4.2. Esempi dell'estrazione automatica dal corpus *ZooCor*, tramite il software *TermoStat*: colonna degli indici di frequenza.

Per questo motivo, nel nostro studio ci siamo soffermati anche sul criterio di specificità, che invece, riflette il « degré de précision des descripteurs utilisés pour l'indexation de documents »<sup>205</sup>. Un elevato livello di specificità permette di individuare i termini più rilevanti e distintivi per il dominio in esame: un termine con un alto punteggio di specificità è, quindi, unico o altamente rilevante per il dominio. Dunque, come evidenziato negli « indices sur l'extraction »<sup>206</sup> di L'Homme, la specificità è uno strumento promettente per valutare la qualità dei termini all'interno di un corpus.

<sup>205</sup> Grand Dictionnaire Terminologique 2023, <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8363492/specificite>.

<sup>206</sup> L'Homme 2004, *op. cit.*, p. 167.

Corpus >> **zoocor** HUMAENVI | Aide | Déconnexion

# Résultats

Liste des termes | Nuage | Statistiques | Structuration | Bigrammes

Candidat de regroupement	Fréquence	Score (Spécificité) +	Variante orthographiques	Matrice
<b>espèce</b>	21221	198.82	espèce espèces	Nom
<b>marin</b>	10975	147.23	marin marins marines	Adjectif
<b>phoque</b>	6942	119.77	phoque phoques	Nom
<b>corail</b>	6459	115.24	corail coraux	Nom
<b>tortu</b>	6347	114.74	tortu tortue tortues	Adjectif
<b>animal</b>	7344	106.44	animal animaux	Nom
<b>tortue marine</b>	5341	105.54	tortue marine tortue marines tortues marines	Nom Adjectif

**Fig. 4.3.** Esempi dell'estrazione automatica dal corpus *ZooCor*, tramite il software *TermoStat*: colonna degli indici di frequenza.

In particolare, il calcolo della specificità riportato in *TermoStat* è un adattamento del « calcul de l'indicateur des formes spécifiques »<sup>207</sup> di Lafon:

« Le calcul de spécificité a été proposé par Lafon (1980) afin de cerner le vocabulaire spécifique à un sous-corpus par rapport à l'ensemble d'un corpus »<sup>208</sup>.

Secondo Lafon, infatti, il coefficiente di specificità indica il livello di significatività dello scarto rilevato tra la frequenza della forma (e quindi del termine) nella parte selezionata (ovvero nel corpus), e la frequenza della stessa in altre "subparti". Lafon fa un calcolo in cui il secondo denominatore, la "subparte", non è altro che una parte dello stesso corpus nel quale viene calcolata la frequenza dello stesso termine: calcola quindi lo scarto tra la frequenza del termine nell'intero corpus e la frequenza dello stesso termine in una parte del corpus<sup>209</sup>.

<sup>207</sup> Lafon 1980, Sur la variabilité de la fréquence des formes dans un corpus, in Guide de l'Utilisateur de *TermoStat*, consultato nel 2024, link: [https://TermoStat.ling.umontreal.ca/doc\\_TermoStat/doc\\_TermoStat.html](https://TermoStat.ling.umontreal.ca/doc_TermoStat/doc_TermoStat.html).

<sup>208</sup> Guide de l'Utilisateur de *TermoStat*, consultato nel 2024, link: [https://TermoStat.ling.umontreal.ca/doc\\_TermoStat/doc\\_TermoStat.html](https://TermoStat.ling.umontreal.ca/doc_TermoStat/doc_TermoStat.html)

<sup>209</sup> Lebart et al. 1998.

D'altro canto, il coefficiente di specificità in *TermoStat* indica lo stesso rapporto di scarto, ma, questa volta, in relazione alla frequenza del termine in un corpus di riferimento:

Nous adaptons légèrement la démarche en fusionnant le corpus de référence et le corpus d'analyse afin de vérifier si le lexique de ce dernier se comporte comme le lexique du premier. Le calcul des spécificités conduit à l'obtention d'un score qui facilite le classement des CT les uns par rapport aux autres.<sup>210</sup>

Di conseguenza, mentre Lafon misura lo scarto tra la frequenza di una forma in una porzione selezionata di un corpus rispetto ad altre sue sotto-parti, Drouin adatta questo concetto riportando la frequenza del termine nel corpus analizzato a quella in un corpus di riferimento esterno. Questo consente di verificare se il lessico del corpus specialistico si comporta in modo analogo a quello di un corpus generale, facilitando una classificazione più accurata dei termini.

Sull'impronta di tali indici, nel nostro studio abbiamo adottato un approccio bifasico e bilanciato: inizialmente, una selezione basata sulla frequenza ha garantito l'inclusione di tutti i termini comuni e potenzialmente rilevanti nel sottodominio; successivamente, un raffinamento basato sulla specificità ci ha permesso di escludere i termini troppo generici, concentrandoci su quelli effettivamente distintivi del nostro dominio. Questo metodo consente di ottenere un set di unità terminologiche che sono sia rilevanti per la biologia marina sia altamente specifiche del sottodominio "Phoque de mer", migliorando significativamente la qualità delle reti lessicali e delle funzioni lessicali che saranno oggetto delle fasi successive all'estrazione terminologica.

#### 4.1.1. Estrazione dei termini dal corpus *ZooCor*: fasi automatizzate e manuali

Al fine di condurre un'analisi terminologica approfondita del lessico della componente francese del corpus *ZooCor*, abbiamo segmentato il processo in tre fasi:

---

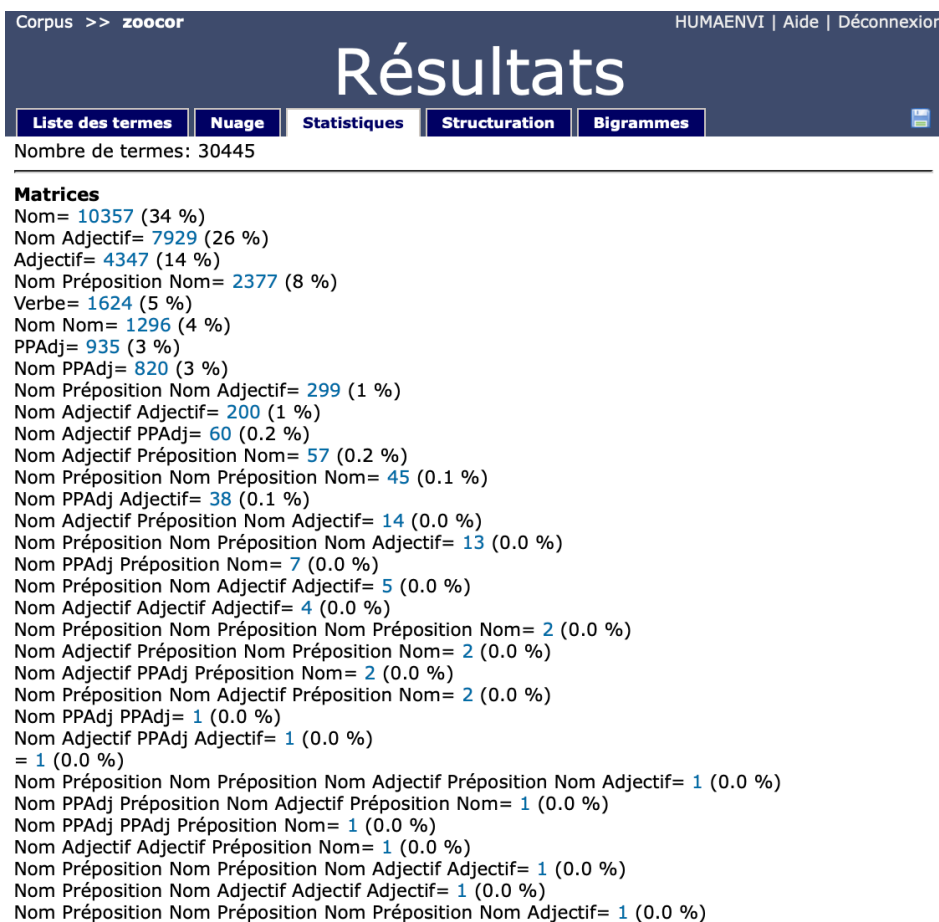
<sup>210</sup> Guide de l'Utilisateur de *TermoStat*, consultato nel 2024, link: [https://TermoStat.ling.umontreal.ca/doc\\_TermoStat/doc\\_TermoStat.html](https://TermoStat.ling.umontreal.ca/doc_TermoStat/doc_TermoStat.html).

1. estrazione automatica: il corpus francese è stato caricato su *TermoStat*, consentendo l'estrazione automatica di un totale di 30.445 unità;
2. spoglio manuale e classificazione: un campione di 5.000 termini, selezionato tra i primi estratti e ordinato alfabeticamente, è stato sottoposto ad uno spoglio manuale ed una "pulitura". I termini sono stati poi classificati in categorie predefinite;
3. selezione e scelta dei termini: selezione semi-automatica di 30 termini per i quali sono stati estratti contesti d'uso, frequenza, specificità e varianti ortografiche.

Nel dettaglio, la Fase I del processo di estrazione terminologica ha rappresentato il primo passo fondamentale per la gestione del corpus *ZooCor*: il corpus, nella sua componente interamente in francese e di notevoli dimensioni (350 testi riuniti in un unico file in formato .txt) è stato quindi caricato nel software di estrazione automatica *TermoStat*<sup>211</sup>. Al termine di questa operazione, *TermoStat* ha generato un elenco complessivo di 30.445 termini, che costituiscono il punto di partenza per le successive fasi di analisi, fornendo un inventario esaustivo del lessico specialistico presente nel corpus.

---

<sup>211</sup> La scelta di questo software è stata dettata dalla sua comprovata efficacia nell'identificazione e nell'estrazione di unità terminologiche da grandi volumi di testo.



**Fig. 4.4.** Esempi dell'estrazione automatica dal corpus *ZooCor*, tramite il software *TermoStat*: sezione « Statistiques ».

Successivamente, la Fase 2 ha avuto come scopo il raffinamento e la categorizzazione dei termini estratti automaticamente: un campione di 5.000 termini candidati – selezionati in ordine alfabetico (A-Z) tra quelli generati nella Fase 1 – è stato sottoposto ad uno spoglio ed una pulizia manuali<sup>212</sup>, che hanno garantito un'analisi approfondita di ogni singolo termine.

Per la classificazione dei 5.000 termini, è stato utilizzato un file Excel, dove i termini sono stati suddivisi in categorie specifiche attraverso un sistema di colorazione delle caselle, indicante una metodologia di classificazione interamente

<sup>212</sup> Ogni termine è stato osservato e classificato in una categoria in base alla sua entità (segue nel testo del paragrafo la spiegazione), ma prima di tutto è stato effettuato un controllo ortografico di ciascuno dei 5.000 termini, il che ha permesso di effettuare una successiva disamina coerente ed equilibrata.

manuale. Le categorie e la relativa codifica cromatica sono state definite come segue:

— *Bons de ETS*<sup>213</sup> (caselle bianche): termini che si riferiscono inequivocabilmente al dominio della fauna marina;

Candidat de regroupement	Fréquence	Spécificité	Variantes orthographiques	Matrice
abattage	70	4.17	abattage__abattages	Nom
abyssal	52	4.11	abyssal__abyssale__abyssaux	Adjectif
acanthaster	10	4.26	acanthaster	Verbe
acanthaster	41	9.1	acanthaster	Nom
acanthasters	286	24.37	acanthasters	Nom
acanthasters adultes	10	4.26	acanthasters adultes	Nom Adjectif
acclimation	124	14.55	acclimation	Nom
acclimatée	29	6.22	acclimatée	PPAdj
acclimater	56	7.56	acclimater__acclimater__acclimater__acclimaterait__acclimaterait__acclimaterait	Verbe
acclimatization	6	3.14	acclimatization	Nom
accouplé	14	3.42	accouplé	PPAdj
accouplement	203	18.58	accouplement__accouplements	Nom
accoupler	66	9.52	accoupler__accouple__accouple__accouplerait__accoupleraient__accoupleraient	Verbe
acropores	10	4.26	acropores	Nom
acte de braconnage	18	5.53	acte de braconnage__actes de braconnage__actes de braconnages	Nom Préposition Nom
action de communication	42	9.54	action de communication__actions de communication__actions de communications	Nom Préposition Nom
action de conservation	49	9.97	action de conservation__actions de conservation__actions de conservations	Nom Préposition Nom
action de gestion	19	6.07	action de gestion__actions de gestion	Nom Préposition Nom
action de protection	14	4.56	action de protection__actions de protection__actions de protections	Nom Préposition Nom
action de sensibilisation	62	10.1	action de sensibilisation__actions de sensibilisation__actions de sensibilisations	Nom Préposition Nom
action de surveillance	12	4.72	action de surveillance__actions de surveillance	Nom Préposition Nom
actions anthropiques	13	4.94	actions anthropiques	Nom Adjectif
activation neutronique	21	6.41	activation neutronique	Nom Adjectif
activité biologique	21	5.91	activité biologique__activités biologiques	Nom Adjectif
activité de chasse	25	7.03	activité de chasse__activités de chasse	Nom Préposition Nom
activité de conservation	15	4.77	activité de conservation__activités de conservation	Nom Préposition Nom
activité de nage	12	4.72	activité de nage__activités de nage	Nom Préposition Nom
activité de nidification	59	10.97	activité de nidification__activités de nidification	Nom Préposition Nom
activité de pêche	162	18.02	activité de pêche__activités de pêche__activités de pêches	Nom Préposition Nom
activité de ponte	216	21.16	activité de ponte__activité de pontes__activités de pontes	Nom Préposition Nom
activité métabolique	12	4.72	activité métabolique__activités métabolique	Nom Adjectif
activité photosynthétique	43	9.32	activité photosynthétique	Nom Adjectif
activité physiologique	6	3.14	activité physiologique__activités physiologiques	Nom Adjectif
activité touristique	29	5.89	activité touristique__activités touristiques	Nom Adjectif
activité vitale	8	3.74	activité vitale__activités vitales	Nom Adjectif

Fig. 4.5. Rappresentazione del file Excel contenente la categoria di classificazione dell'estrazione terminologica *Bons de ETS*.

— *SynScient*<sup>214</sup> de ETS (caselle azzurre): termini latini che designano il primo o il secondo nome scientifico (nomenclatura binomiale) di una specie marina, o qualsiasi altra denominazione di livello tassonomico;

<sup>213</sup> ETS: anagramma di *Extraction Terminologique Spécifique*.

<sup>214</sup> SynScient si rifà ad una nostra definizione di sinonimia scientifica, approfondita nel capitolo successivo in merito alle funzioni lessicali, come vedremo nel capitolo 6.

Candidat de regroupement	Fréquence	Spécificité	Variantes orthographiques	Matrice
abyssicola	6	3.14	abyssicola	Verbe
abyssicola	8	3.74	abyssicola	Nom
acanthotetra	16	5.53	acanthotetra	Nom
acanthurus	19	6.07	acanthurus	Verbe
acropora	7	3.45	acropora	Verbe
acropora	24	6.88	acropora	Nom
acrosmlia	6	3.14	acrosmlia	Verbe
acrosmlia	12	4.72	acrosmlia	Nom
aculeata	14	5.14	aculeata	Nom
acutidens	9	4.01	acutidens	Nom
acutorostrata	47	9.76	acutorostrata	Nom
aeglefinus	6	3.14	aeglefinus	Nom
aegyptica	6	3.14	aegyptica	Nom
agaricia	11	4.5	agaricia	Nom
agariciformis	9	4.01	agariciformis	Verbe
agaricites	31	7.87	agaricites	Nom
agassizii	69	11.88	agassizii	Nom
alveolata	18	5.9	alveolata	Nom
ambigua	15	5.34	ambigua	Nom
ambiguus	7	3.45	ambiguus	Nom
amblyrhynchos	7	3.45	amblyrhynchos	Nom
ampullatus	15	5.34	ampullatus	Adjectif
amygdalophyllides	7	3.45	amygdalophyllides	Nom
amygdalophylloide	8	3.74	amygdalophylloide	Adjectif
anasterias	10	4.26	anasterias	Nom
angélus	14	3.83	angélus	Nom
angustrostris	6	3.14	angustrostris	Adjectif
annularis	47	9.76	annularis	Nom
annularis	6	3.14	annularis	Adjectif

Fig. 4.6. Rappresentazione del file Excel contenente la categoria di classificazione dell'estrazione terminologica *SynScient de ETS*.

— *À contextualiser de ETS* (caselle arancioni): termini di natura generica il cui significato, se contestualizzato appropriatamente o abbinato a termini specifici, potrebbe risultare rilevante per lo studio del dominio in oggetto;

Candidat de regroupement	Fréquence	Spécificité	Variantes orthographiques	Matrice
abdomen	219	18.32	abdomen	Nom
abdominal	333	24.33	abdomina__abdominale__abdominales	Adjectif
abiotique	89	13.39	abiotique__abiotiques	Adjectif
abondance de nourriture	8	3.36	abondance de nourriture	Nom Préposition Nom
abondance de pétoncles	6	3.14	abondance de pétoncles__abondances de pétoncles	Nom Préposition Nom
abondance de proie	6	3.14	abondance de proie__abondance de proies__abondances de proies	Nom Préposition Nom
abondance moyenne	10	4.26	abondance moyenne	Nom Adjectif
abondance relatif	123	15.82	abondance relatif__abondance relative__abondances relatives	Nom Adjectif
abondance totale	17	5.72	abondance totale	Nom Adjectif
aboucher	19	6.07	aboucher__abouche__abouchent	Verbe
abritée	92	6.2	abritée	PPAdj
absence de dents	6	3.14	absence de dents	Nom Préposition Nom
absence de donnée	30	6.91	absence de donnée__absence de données	Nom Préposition Nom
absence de lumière	6	3.14	absence de lumière	Nom Préposition Nom
absence de plage	6	3.14	absence de plage__absence de plages	Nom Préposition Nom
absence de plaques	7	3.45	absence de plaques	Nom Préposition Nom
absorbance	18	5.9	absorbance__absorbances	Nom
absorption	446	26.31	absorption	Nom
absorption d' ammonium	16	5.53	absorption d' ammonium	Nom Préposition Nom
absorption d' azote	19	6.07	absorption d' azote	Nom Préposition Nom
absorption d' azote inorganique	6	3.14	absorption d' azote inorganique	Nom Préposition Nom Adjectif
absorption de nitrate	10	4.26	absorption de nitrate__absorption de nitrates	Nom Préposition Nom
absorption de phosphore	33	8.13	absorption de phosphore	Nom Préposition Nom
absorption de phosphore inorganique	10	4.26	absorption de phosphore inorganique	Nom Préposition Nom Adjectif
absorption de phosphore organique	9	4.01	absorption de phosphore organique	Nom Préposition Nom Adjectif
abyssalis	6	3.14	abyssalis	Nom
accidentel	1220	46.03	accidentel__accidentels__accidentelle__accidentelles	Adjectif
accolée	160	13.98	accolée	PPAdj

Fig. 4.7. Rappresentazione del file Excel contenente la categoria di classificazione dell'estrazione terminologica *À contextualiser de ETS*.

- *Termes spécifiques de ETS* (caselle grigie): termini afferenti a contesti tecnici, scientifici, chimici, medici, o anatomici, relativi alla Fauna Marina, o che possono essere di interesse per lo studio;

Candidat de regroupement	Fréquence	Spécificité	Variantes orthographiques	Matrice
abactinal	105	14.71	abactinal	Nom
abactinal compact	6	3.14	abactinal compact	Nom Adjectif
abactinal formé de plaques polygonales ± circulaires	8	3.74	abactinal formé de plaques polygonales ± circulaires	Nom PPAdj Préposition Nom Adjectif Préposition
abactinal réticulé	6	3.14	abactinal réticulé	Nom Adjectif
abactinal réticulé	10	4.26	abactinal réticulé	Nom PPAdj
abactinale	63	11.34	abactinale	Adjectif
abactinale	39	8.86	abactinale	Nom
abactinales	214	21.06	abactinales	Nom
abactinales	41	9.1	abactinales	Adjectif
abactinales disposées	18	5.9	abactinales disposées	Nom PPAdj
abactinales élargies	10	4.26	abactinales élargies	Nom PPAdj
abactinales latérales	7	3.45	abactinales latérales	Nom Adjectif
abactinales pourvues	14	5.14	abactinales pourvues	Nom Adjectif
abactinaux	40	8.98	abactinaux	Nom
abaisseur	15	5.34	abaisseur__abaisseurs	Adjectif
abaisseur	25	7.03	abaisseur__abaisseurs	Nom
abajoue	44	9.43	abajoue__abajoues	Nom
abcès	51	5.31	abcès	Nom
abcès pulmonaire	8	3.74	abcès pulmonaire__abcès pulmonaires	Nom Adjectif
ablation	29	3.52	ablation	Nom
acellulaire	13	4.94	acellulaire__acellulaires	Adjectif
acéré	27	4.68	acéré	PPAdj
acerosa	6	3.14	acerosa	Nom
acetabulum	14	5.14	acetabulum	Nom
acétabulum	10	4.26	acétabulum	Nom
acétique	53	10.21	acétique	Adjectif
acétone	19	4.31	acétone	Nom
acheuléen	15	5.34	acheuléen__acheuléens__acheuléenne	Adjectif

**Fig. 4.8.** Rappresentazione del file Excel contenente la categoria di classificazione dell'estrazione terminologica *Termes spécifiques de ETS*.

- *Rumeurs* (caselle rosse): questa categoria include elementi non pertinenti o da scartare, come segni di punteggiatura o lettere singole, termini con errori grammaticali, termini in lingue diverse dal francese, termini con riferimenti chiaramente fuori contesto, termini eccessivamente generici (*i.e. absence, activité*) e termini che si riferiscono a specie non target, ovvero non appartenenti alla fauna marina<sup>215</sup>.

<sup>215</sup> Come già anticipato nel paragrafo precedente, questa fase ha integrato un approfondimento del quadro teorico, essenziale per definire i criteri di selezione. In particolare, durante questa fase, è stata esaminata la distinzione tra *term et lexie*, e la differenza tra termini complessi e semplici (si rinvia alla teoria di L'Homme 2020, *op. cit.*, pp. 1 e 58), come riportato nel § capitolo 3.

Candidat de regroupement	Fréquence	Spécificité	Variante orthographiques	Matrice
	11	4.5		Nom
	18	5.9		PPAd
varez	7	3.45	varez	Verbe
varez-fem-ndez	9	4.01	varez-fem-ndez	Verbe
	29	7.6		Nom
	49	9.97		PPAd
ge 1 maturité sexuelle	5	3.14	ge 1 maturité sexuelle	Nom Préposition Nom Adjectif
ge adulte	95	4.64	ge adulte	Nom Adjectif
ge de maturité	5	3.14	ge de maturité _ge de maturité	Nom Préposition Nom
ge différent	14	3.63	ge différent _ge différents _ges différents	Nom Adjectif
ge estimé	9	3.64	ge estimé	Nom PPAd
ge réservoir	18	5.9	ge réservoir _ges réservoir _ges réservoirs	Nom Nom
ge viséen	9	3.74	ge viséen	Nom Adjectif
ne	222	19.64	ne _nes	Nom
ne domestique	13	4.94	ne domestique _nes domestiques	Nom Adjectif
ne sauvage	8	3.74	ne sauvage _nes sauvages	Nom Adjectif
nessa	25	5.58	nessa _nesses	Nom
nn	25	7.03	nn	Nom
strea	11	4.5	strea	Nom
ne	75	12.4	ne	Nom
ne	10	4.26	ne	Nom
ne	20	6.24	ne	PPAd
t-dire	62	7.64	t-dire	Nom
t-fait	10	4.26	t-fait	Nom
gge	7	3.45	gge	Nom
ll	6	3.14	ll	Nom
lla	29	7.58	lla	Nom
ltn	11	4.5	ltn	Nom
lstraa	12	4.72	lstraa	Nom
lstréides	9	3.74	lstréides	Nom
tune	1	4.01	tune	Nom
	12	4.72		Nom
	13	4.94		PPAd
	10	4.26		Nom
	22	6.57		PPAd

Fig. 4.9. Rappresentazione del file Excel contenente la categoria di classificazione dell'estrazione terminologica *Rumeurs*.

I processi di spoglio e di classificazione manuale hanno garantito un'accurata organizzazione del lessico estratto, ponendo le basi per le successive fasi di selezione e di analisi.

Infine, la Fase 3 ha rappresentato l'apice del processo di raffinazione terminologico, focalizzandosi sulla selezione mirata dei termini più rilevanti dal corpus *ZooCor*.

Per la selezione dei termini candidati all'analisi è stata adottata una metodologia semi-automatica, combinando l'analisi di indici quantitativi con una scelta qualitativa basata su criteri teorici: infatti, tra i vari indici considerati (come già discusso nel paragrafo precedente), la frequenza e la specificità si sono rivelati i più rilevanti, invece la "complessità dei termini" (cioè la differenza tra un termine semplice ed uno complesso), sebbene considerata, non è stata un criterio di selezione primario.

Di conseguenza, in linea con una metodologia integrata e semi-automatica, sono stati eseguiti i seguenti steps: nella categoria *Bons de ETS*, comprendente circa 820 termini già validati per il dominio della fauna marina (derivanti dai primi 5.000 termini analizzati), i termini sono stati ordinati per frequenza crescente, tra questi

820 termini, sono stati selezionati<sup>216</sup> i 100 con il più alto tasso di frequenza; a questo punto è stata effettuata una normalizzazione dei dati, sostituendo il punto decimale con la virgola negli score di specificità, per garantirne il corretto riconoscimento numerico, ed infine, tra i 100 termini selezionati per la loro frequenza, sono stati identificati i 50 più specifici, sulla base del valore crescente dello score di specificità. Quest'ultima classificazione è stata nominata *Sélection de Bons de ETS*, come mostrato dalle seguenti figure:

Candidat de regroupement	Fréquence	Spécificité	Variantes orthographiques	Matrice
animal	7344	106,44	animal__animaux	Nom
capture	2669	71,87	capture__captures	Nom
chasse	3259	63,43	chasse__chasses	Nom
adulte	1789	50,86	adulte__adultes	Adjectif
cirre	1120	48,31	cirre__cirres	Nom
capturée	1224	45,69	capturée	PPAdj
algue	1156	45,61	algue__algues	Nom
cétacé	1025	45,23	cétacé__cétacés	Nom
carapace	941	43,13	carapace__carapaces	Nom
ambulacraire	891	43,08	ambulacraire__ambulacraires	Adjectif
capture accidentelle	875	42,61	capture accidentelle__captures accidentelle__captures accidentelles	Nom Adjectif
climatique	1242	40,07	climatique__climatiques	Adjectif
baleine	829	38,4	baleine__baleines	Nom
caudal	685	37,67	caudal__caudale__caudaux	Adjectif
baie	1049	37,11	baie__baies	Nom
branche	1746	36,47	branche__branches	Nom
buccal	615	35,39	buccal__buccale__buccales	Adjectif
benthique	556	33,96	benthique__benthiques	Adjectif
biodiversité	740	33,88	biodiversité	Nom
atoll	560	33,17	atoll__atolls	Nom
bec	668	32,82	bec__becs	Nom
blanchissement	470	30,99	blanchissement__blanchissements	Nom
bourelet	461	30,63	bourelet__bourelets	Nom
captivité	528	30,1	captivité__captivités	Nom
chalut	444	29,09	chalut__chaluts	Nom
capsule	452	28,94	capsule__capsules	Nom
cil	430	28,64	cil__cils	Nom
changement climatique	541	28,05	changement climatique__changement climatiques__changements climatiques	Nom Adjectif

Fig. 4.10. Rappresentazione del file Excel contenente la categoria di classificazione dell'estrazione terminologica *Sélection de Bons de ETS*.

<sup>216</sup> Potremmo parlare piuttosto di “scelta”, in quanto la selezione avviene in maniera semi-automatica: in primis i termini BONS sono frutto già di una prima selezione manuale; in secondo luogo, la selezione dei 100 termini è avvenuta, appunto, escludendo – tra i termini BONS più frequenti – quelli non ritenuti rilevanti, quindi soprattutto dove il significato dell’UT non fosse legato al dominio in senso ampio (L’Homme 2004, *op. cit.*); infatti l’esclusione è stata effettuata anche a seguito dell’identificazione dei “Contesti d’uso”: se il termine era troppo tecnico o non presentava un contesto d’uso strettamente inerente, è stato scartato in favore di altri termini (ad esempio *adulte* nome è stato scartato in presenza già di *adulte* aggettivo; *calcium* era stato selezionato in quanto relativo ai coralli, ma il suo contesto d’uso faceva riferimento solo alla sostanza chimica; *ambulacre* presentava un contesto d’uso troppo tecnico, è stato sostituito anche a causa della presenza più specifica, dettagliata e inerente di *ambulacraire*, anche se qui sorge il dubbio sulla derivazione morfologica.

Legenda Sélection de Bons de ETS_5K	
Verde scuro	50 termini + <b>specifici</b>
Verde chiaro	100 termini + <b>frequenti</b> , di cui 50 sono diventati verdi scuri
Nessun Riempimento	tutti i termini <b>Bons</b> , di prima selezione, di cui 100 sono diventati verdi chiari e successivamente 50 sono diventati verdi scuri

**Fig. 4.11.** Sezione del file Excel contenente la legenda della categoria di classificazione dell'estrazione terminologica *Sélection de Bons de ETS*.

L'identificazione finale di queste ultime 50 UT è stata pertanto, ancora una volta, validata da una selezione manuale secondo i criteri teorici proposti dallo studio. In particolare, l'attenzione si è focalizzata sul legame semantico dei termini con il dominio esaminato, in particolare:

- sui legami paradigmatici (es. *baleine – mammifère; phoque – peau de phoque; braconnage – chasse; blanchissement – blanchir*);
- sui legami sintagmatici (es. *allaiter les bébé-phoques; chasser – hakapik; capturer les phoques – huile*)<sup>217</sup>.

Dunque, secondo i criteri già menzionati, abbiamo ottenuto i 30 termini più frequenti e specifici dei primi 5.000 termini estratti, e sulla base dei legami semantici, paradigmatici, e derivazionali, li abbiamo suddivisi e schematizzati riproducendo la classificazione offerta dalla principale interfaccia del software *TermoStat* con l'aggiunta dei *Contextes d'usage*:

Candidat de regroupement	Fréquence	Spécificité	Variantes orthographiques	Contexte d'usage (Structuration)	Matrice
--------------------------	-----------	-------------	---------------------------	----------------------------------	---------

**Fig. 4.12.** Categorie di classificazione dei termini estratti e selezionati dal corpus.

**Fig. 4.13.** Categorie di classificazione su *TermoStat* dei termini estratti dal corpus.

#### 4.1.1. Selezione dei termini correlati a specifici campi semantici

<sup>217</sup> Per le prospettive future di questo studio, si prevede di integrare ulteriori criteri, in particolare nell'ambito delle unità predicative (meno esplorate in questa fase iniziale), con un focus sulla presenza di riferimenti ad attanti correlati ai termini già selezionati, includendo affinità paradigmatiche, legami morfologici e strutture attanziali specifiche, andando ad escludere le unità predicative i cui attanti non rimandano ad UT già selezionate (L'Homme 2014a, *op. cit.*).

Alla luce del processo metodologico esposto nei paragrafi precedenti (in particolare delle fasi preliminari di estrazione automatica dal corpus *ZooCor* tramite il software *TermoStat*, del *dépouillement manuel*, della classificazione dei termini secondo criteri di frequenza e specificità), la presente sezione si focalizza sulla selezione dei 30 termini più rappresentativi e del loro raggruppamento in campi semantici distinti. La motivazione sottostante a tali scelte è duplice: da un lato, organizzare il lessico in domini concettuali omogenei; dall'altro, identificare un sotto-dominio di particolare interesse per lo studio, al fine di delineare un *fil rouge* tematico che consenta un'analisi semantico-lessicale di un sottoinsieme coerente di termini. I termini sono stati categorizzati in quattro campi semantici principali, che riflettono le aree tematiche emergenti dal corpus:

⇒ Campo Semantico 1: Habitat e ambienti marini

Questo campo include termini che descrivono gli ecosistemi e gli spazi in cui vivono le specie marine, di cui riportiamo alcuni esempi:

Candidat de regroupement		Fréquence	Spécificité	Variante orthographiques	Contexte d'usage (Structuration)	Matrice
1.	aire marine	192	19,94	aire marine ; aires marines	aire marine protégée - aires marines protégées - création d'aires marines - réseau d'aires marines	Nom Adjectif
2.	aquarium	356	24,48	aquariums	aquarium d'eau	Nom
3.	algueraie	23	6,72	algueraies ; algueraie	algueraies à sargasses	Nom
4.	atoll	560	33,17	atoll ; atolls		Nom
5.	baie	1049	37,11	baie ; baies	baie longue - fond de baie - baie blanche - plage de baie	Nom
6.	canyon	201	18,47	canyon ; canyons	canyons sous-marins - canyons méditerranéens	Nom

**Fig. 4.14.** Termini estratti dal corpus *ZooCor*: “Habitat e ambienti marini”

⇒ Campo Semantico 2: Morfologia biologica – “parti di”

Questa categoria raggruppa termini relativi alle strutture anatomiche delle specie marine, tra cui:

Candidat de regroupement		Fréquence	Spécificité	Variante orthographiques	Contexte d'usage (Structuration)	Matrice
7.	aileron	65	7,86	aileron ; ailerons	aileron dorsal - aileron droit - ailerons latéraux	Nom
8.	bec	668	32,82	bec ; becs	bec étroit - bec noir - forme de bec - bec de cane - bec corné - dauphin à bec	Nom

					- bec d'oie - bec supérieur - bec brun - bec jaune - bec blanc - baleine bec	
9.	branche	1746	36,47	branche ; branches	branches épaisses - branche postérieure - branche montante - branches divergentes - branche ventrale - branches secondaires - branche externe - branche interne - branche supérieure - branche utérine - branches maîtresses - branches grosses - branche de corail - branche ascendante - branche nerveuse - branche terminale - branche cylindrique - branche antérieure - branche maxillaire - branche latérale - branches grêles - branche principale	Nom
10.	branchie	186	19,45	branchie ; branchies	branchie interne - branchie externe - branchies lymphatiques - branchie en formation - branchies operculaires	Nom
11.	calotte	49	6,13	calotte ; calottes	calotte colorée - calotte postérieure - calotte glaciaire - calotte antarctique - calottes polaires - calottes sphérique	Nom
12.	buccal	615	35,39	buccal ; buccale ; buccales	cils buccaux - appareil buccal - mamelon buccal - vestibule buccal - anneau buccal - bord buccal - tentacules buccaux - armature buccale - extrémité buccale - structure buccale - région buccale - plancher buccal - cavité buccale - segment buccal - bouclier buccal - orifice buccal - pièce buccale - buccal - plaque buccale - ouverture buccale - membrane buccale - fente buccale - papille buccale - angle buccal	Adjectif
13.	carapace	941	43,13	carapace ; carapaces	carapace osseuse - carapace largeur - longueur droite de carapace - carapace moyenne - cm de carapace - carapace length - longueur de carapace - carapace de tortue	Nom
14.	cartilagineux	331	26,16	cartilagineux ; cartilagineuse ; cartilagineuses	pièce cartilagineuse - crâne cartilagineux - anneau cartilagineux - poissons cartilagineux - partie cartilagineux - état cartilagineux - lame cartilagineuse - origine cartilagineux - ceinture cartilagineuse - cartilagineux - plaque cartilagineux - ceinture scapulaire cartilagineuse	Adjectif
15.	caudal	685	37,67	caudal ; caudale ; caudaux	palettes caudales - veine caudale - prolongement caudal - dernières caudales - capuchon caudal - extension caudale - nageoire caudale - région caudale - appendice caudal - caudal - vertèbre caudale - partie caudale - extrémité caudale - pédoncule caudal	Adjectif
16.	cil	430	28,64	cil ; cils	rangée de cils - cils buccaux - cil long - cils moteurs - cils vibratiles	Nom

Fig. 4.15. Termini estratti dal corpus *ZooCor*: "Morfologia biologica"

⇒ Campo Semantico 3: Attività compiute dalle specie marine

Questo campo include i termini che si riferiscono ad azioni, processi biologici o funzioni svolte attivamente dagli organismi marini, ad esempio:

Candidat de regroupement	Fréquence	Spécificité	Variantes orthographiques	Contexte d'usage (Structuration)	Matrice
17. accouplement	203	18,58	accouplement ; accouplements	habitat d'accouplement	Nom
18. activité de ponte	216	21,16	activité de ponte ; activité de pontes ; activités de pontes	activité de ponte de la saison - activité de ponte enregistrée - activité de ponte observe - activités de pontes de tortue verte	Nom Préposition Nom
19. adaptation génétique	6	3.14	adaptation génétique	phénomènes d'adaptation génétique - potentiel d'adaptation génétique	Nom Adjectif
20. asexué	111	13.22	asexué ; asexués ; asexuéés	mode de reproduction asexué - manière asexuéée - reproduction asexué - asexué	Adjectif
21. activité de nidification	59	10.97	activités de nidification	activité de nidification de la tortue olivâtre - activités de nidification recensées	Nom Préposition Nom
22. alimentation	1655	49.37	alimentations	alimentation côtiers - alimentation favorables - alimentation important - source d'alimentation - alimentation de tortues vertes - alimentation estivale - temps d'alimentation - lieu d'alimentation - alimentation océanique - type d'alimentation - alimentation hivernale - alimentation majeur - phase d'alimentation - site d'alimentation - zone d'alimentation - aire d'alimentation - habitat d'alimentation - alimentation benthique	Nom
23. circulation océanique	64	10.84	circulation océaniques ; circulations océaniques	modèles de circulation océanique - variations de la circulation océanique - évolution des circulations océaniques - circulation océanique de la zone	Nom Adjectif

**Fig. 4.16.** Termini estratti dal corpus *ZooCor*: Attività compiute dalle specie marine

Infine, partendo dai termini con maggiore indice di frequenza e specificità di questa prima estrazione da *ZooCor*, come *chasse* [FR.: 3259; SPÉC.: 63.43], ed *espèce*<sup>218</sup>

<sup>218</sup> Sebbene il termine *espèce*, benché estratto ovviamente dal corpus, non figuri tra i termini sottoposti direttamente a pulizia, spoglio e selezione manuali (in quanto la lettera iniziale "E" non rientrava ancora nella selezione in ordine alfabetico effettuata), esso costituisce un nucleo concettuale fondamentale all'interno del nostro corpus focalizzato sulla fauna marina. Inoltre, la sua

[FR.: 21221; SPÉC.: 198.82] abbiamo focalizzato l'indagine in maniera più approfondita sulle UT correlate al sottodominio delle minacce per le specie marine e per la biologia della conservazione, che quindi denotano una condizione di rischio o un'azione lesiva. Tale approfondimento ha condotto alla creazione del “Campo Semantico 4: Attività subite, rischi e minacce per le specie marine”, in cui abbiamo inserito i seguenti termini estratti da *ZooCor*:

⇒ Campo Semantico 4: Attività subite, rischi e minacce per le specie marine  
Questo campo include i termini che si riferiscono ai pericoli, agli impatti negativi, al degrado ed alle azioni (umane o naturali) che gli organismi subiscono:

Candidat de regroupement	Fréquence	Spécificité	Variante orthographiques	Contexte d'usage (Structuration)	Matrice
24. abattage	70	4.17	abattages	technique d'abattage - méthode d'abattage	Nom
25. appauvrissement	45	3.19	appauvrissements	appauvrissement de la diversité biologique - appauvrissement de la biodiversité des récifs - le rythme d'appauvrissement de tous les habitats naturels	Nom
26. blanchissement	470	30,99	blanchissements	ornement de blanchissement - blanchissement adaptatif - épisode de blanchissement - phénomène de blanchissement - blanchissement corallien - blanchissement massif - événement de blanchissement	Nom
27. braconnage	376	27,27	braconnages	braconnage de tortue - braconnage intensif - braconnage de tortue marine - acte de braconnage	Nom
28. capture	2669	71,87	captures	indices de capture - capture d'individus - mode de capture - capture d'oiseaux - formulaire de capture - captures annuels - traces de capture - méthode de capture - capture de phoques - lieu de capture - capture d'électrons - opération de capture - capture d'espèce - cas de capture - zone de capture - nombre total de captures - tentacules de capture - fréquence de capture - captures de dauphins - point de capture - dynamique de capture - capture accessoire - nombre de capture -	Nom

pervasiva presenza nelle colonne di *Contexte d'usage* (come evidenziato nella fase di Structuration di *TermoStat*) ne sottolinea la centralità e la forte correlazione semantica con la maggior parte degli altri termini analizzati.

					capture de tortue marine - capture accidentelle - taux de capture - capture de tortue - site de capture - risque de capture - tentative de capture - donnée de capture - capture intentionnelle - capture de proies - capture totale	
29.	chalut	444	29,09	chaluts	chaluts à perche - chalut de fond - chalut pélagique - trait de chalut - chalut benthique	Nom
30.	chasse	3259	63,43	chasses	cheval de chasse - chasse de phoques gris - forme de chasse - chasse sous-marine - lieu de chasse - chasse de subsistance - scène de chasse - saison de chasse - technique de chasse - chasse réglementée - localisations de chasse - excursion de chasse - stratégies individuelles de chasse - chasse benthique - chasse intensif - habitat de chasse - succès de chasse - chasse probable - stratégie de chasse - chasse commercial - chasse traditionnel - effort de chasse - méthode de chasse - zone de chasse - comportement de chasse - type de chasse - clusters de chasse - pratique de chasse - permis de chasse - plongées de chasse - activité de chasse - chasse individuel	Nom

**Fig. 4.17.** Termini estratti dal corpus *ZooCor*: Attività subite, rischi e minacce per le specie marine

I termini afferenti al Campo Semantico 4. Minacce per le specie hanno costituito la base per la successiva analisi lessicografica, per la quale ci siamo avvalsi della metodologia della LEC (*Lexicographie Explicative et Combinatoire*).

In particolare, il termine *chasser*<sub>2</sub> è stato qui inserito con la pendice “2” (riferito all’attività antropica nei confronti delle specie marine), al fine di differenziarlo da *chasser*<sub>1</sub> con pendice “1”, presente nel Campo Semantico 3 (relativo alle attività generiche di ricerca o inseguimento delle prede da parte degli animali)<sup>219</sup>. A valle di tale distinzione vi è un controllo incrociato dei dati nel nostro corpus che ha rivelato una forte correlazione positiva tra il lemma *chasse* e il lemma *phoque*:

<sup>219</sup> È importante sottolineare che questa distinzione semantica non è stata rilevata dal software di estrazione; è stato invece il controllo manuale umano e l’analisi approfondita dei contesti d’uso a permetterci di identificare tale differenza, come verrà illustrato nel dettaglio nel § Capitolo 5.

l'espressione *chasse au phoque* (Fig. 4.18.) rappresenta una delle espressioni più frequenti riscontrate nell'intero corpus.

Contextes	
Concordance	
de Sherbrooke Avril 2008 Mots clés :	<b>chasse</b> au phoque , développement durable , bien-être
La	<b>chasse</b> au phoque dans l' Atlantique est une
est d' évaluer la soutenabilité de la	<b>chasse</b> au phoque dans l' est du Québec
information circule via différents médias sur la	<b>chasse</b> au phoque et il est difficile pour
une part , les groupes contre la	<b>chasse</b> au phoque montrent des images très brutales
images très brutales et sanguinaires de la	<b>chasse</b> et d' autre part , le gouvernement
part , le gouvernement soutient que cette	<b>chasse</b> est effectuée sans cruauté .
? SOMMAIRE La	<b>chasse</b> au phoque dans l' Atlantique est une
diffusent largement des images sensationnelles de la	<b>chasse</b> avec des porte-parole bien connus , comme
information circule via différentes sources sur la	<b>chasse</b> au phoque et il est difficile pour
une part , les groupes contre la	<b>chasse</b> au phoque montrent des images très brutales
images très brutales et sanguinaires de la	<b>chasse</b> et d' autre part le gouvernement soutient
autre part le gouvernement soutient que cette	<b>chasse</b> est effectuée sans cruauté .
est d' évaluer la soutenabilité de la	<b>chasse</b> au phoque dans l' est du Québec
la compréhension de la problématique de la	<b>chasse</b> au phoque est présentée .
Ensuite , la description de la	<b>chasse</b> éclaire tout l' aspect historique relié à
règlementation exhaustive qui s' applique à la	<b>chasse</b> , des plans de gestion établis par
qui sont commercialisés ainsi que de la	<b>chasse</b> au phoque exercée ailleurs dans le monde
La	<b>chasse</b> au phoque est une activité qui est
Îles-de-la-Madeleine , là où la plus importante	<b>chasse</b> au phoque est réalisée au Québec .
On comprend aussi que la	<b>chasse</b> au phoque permet de retirer des revenus
le bien-être des animaux lors de la	<b>chasse</b> au phoque sont résumés .
2002. Le premier rapport indique que la	<b>chasse</b> au phoque provoque une souffrance considérable et
et le deuxième rapport indique que la	<b>chasse</b> au phoque pratiquée dans l' Atlantique canadien
discute de l' amélioration des pratiques de	<b>chasse</b> dans le but de la rendre plus
Canadiens pour améliorer la situation de la	<b>chasse</b> au phoque sont indiquées .
compromettre la survie des troupeaux , la	<b>chasse</b> au phoque est soutenable .
arriver à conclure avec certitude que la	<b>chasse</b> ne provoque pas de souffrance inutile pour
d' études sur certains aspects de la	<b>chasse</b> au phoque .
à l' Association de la	<b>chasse</b> au Canada

**Fig. 4.18.** Rappresentazione dell'espressione *chasse au phoque* nella sezione *Concordance* di *TermoStat*.

Una tale correlazione significativa, insieme con altri elementi che verranno elencati di seguito, hanno indirizzato la nostra attenzione in modo specifico su questa specie.

#### 4.2. Studio pilota: Estrazione terminologica dalla sezione dedicata al sottodominio PHOQUES DE MER

Dopo aver illustrato la composizione del nostro corpus *ZooCor* (§ capitolo 2) e le principali fasi di estrazione terminologica condotte su di esso, abbiamo sviluppato uno studio pilota dedicato all'estrazione ed alla successiva analisi di unità terminologiche da un sottoinsieme più contenuto del corpus di riferimento: l'obiettivo di tale sperimentazione era quello di definire e testare un percorso metodologico chiaro ed efficace per l'analisi lessicale approfondita presentata nel

§ capitolo 5. A tal fine, è stata scelta una sezione specifica della componente francese di *ZooCor*, interamente dedicata al sottodominio PHOQUES DE MER<sup>220</sup>.

La nostra attenzione verso questa specifica Famiglia di animali deriva dalle seguenti motivazioni<sup>221</sup>: in primo luogo, lo studio è spinto dai risultati del paragrafo precedente sulla prima estrazione di *ZooCor*, che hanno messo in luce la centralità del termine *chasse* – i cui alti indici di frequenza e specificità ne hanno confermato la rilevanza come concetto chiave nel dominio delle minacce per le specie – e la sua alta correlazione con il termine *phoque*. Questa evidenza statistica – dimostrata anche dall’espressione *chasse au phoque* che rappresenta una delle espressioni più frequenti riscontrate nell’intero corpus (§ Fig. 4.18.) – ha elevato *phoque* (foca) a termine *pivot* per una successiva analisi, dando luogo alla nostra decisione di concentrare lo studio lessicografico e tematico su questa specifica specie nel presente studio pilota. In secondo luogo, anche lo stesso termine *phoque* costituisce una delle UT di maggior interesse in *ZooCor*<sup>222</sup>, secondo gli indici di frequenza e specificità [FRÉQUENCE : 6942; SCORE DE SPÉCIFICITÉ : 119.77], maggiore anche allo stesso termine *chasse* (§ Fig. 4.19.);

---

<sup>220</sup> Tale sezione è costituita da 13 testi raccolti durante la ricerca inerente al sottodominio PHOQUES DE MER, in quanto si rammenta che, in fase di costituzione del corpus, i testi sono stati raccolti per categoria di specie, ovvero secondo il criterio tematico di suddivisione in “Famiglie” (si rinvia ai paragrafi § 1.2. e § 2.2.2.).

<sup>221</sup> Tra queste anche una ragione di natura scientifica, trattandosi prevalentemente di testi accademici e di ricerca.

<sup>222</sup> Infatti, dal punto di vista extralinguistico ed in ottica di *citizen science* – considerato il recente incremento degli avvistamenti della Foca Monaca, sottospecie a rischio d’estinzione della foca, lungo le coste campane .. (Punta Campanella, 25.04.2025, *il Sole 24h*; Capri, 10.05.2025, *Corriere della Sera*; Ischia, 12.05.2025, *Il Mattino*; Ischia, 16.07.2025, *La Repubblica*; Cilento, 9.08.2025, *FanPage*; Capri, 3.09.2025, *FanPage*) – si è resa necessaria un’azione di divulgazione scientifica dei comportamenti appropriati nei confronti di questa specie, un aspetto che supporta direttamente gli obiettivi del progetto di ricerca legati alla trasmissione delle conoscenze, alla sensibilizzazione ed alla protezione del territorio.

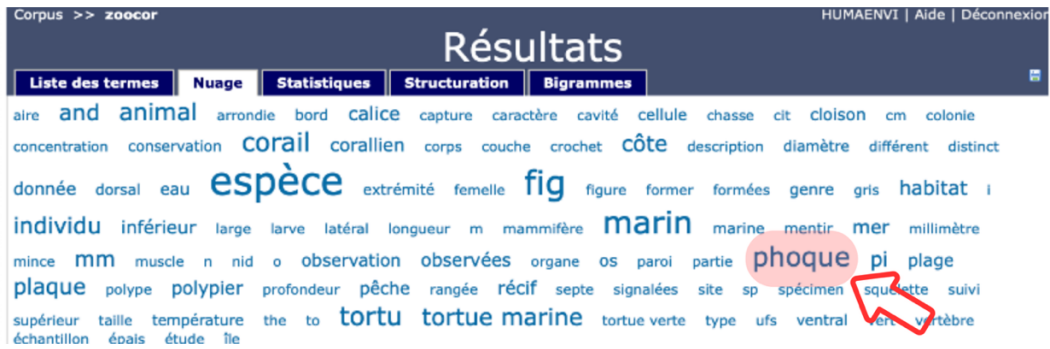


Fig. 4.19. Rappresentazione in una *word cloud* proposta dal software *TermoStat*<sup>223</sup>.

Dopo aver sottoposto la sottosezione del corpus al software di estrazione automatica *TermoStat*<sup>224</sup>, abbiamo identificato sia termini semplici (nomi, verbi, aggettivi) sia termini complessi nominali (cioè composti da più entità grafiche, separate da uno spazio bianco o da un trattino o da un apostrofo) (§ capitolo 3).

Tra i termini che si sono rivelati di maggiore interesse per la nostra analisi lessicale/terminologica, spiccano:

- Termini semplici: *phoque* (frequenza 2724, specificità 404.82), *espèce* (frequenza 1225, specificità 186.75), *plongée* (frequenza 625, specificità 175.25), *chasse* (frequenza 832, specificità 150.26), etc.;
- Termini complessi: *mammifère marin* (frequenza 134, specificità 89.08), *phoque commun* (frequenza 82, specificità 71.76), *chasse commercial* (frequenza 58, specificità 60.41), etc.

<sup>223</sup> Nella *word cloud* dell'estrazione terminologica nel corpus *ZooCor* la frequenza dei termini è direttamente proporzionale alla loro grandezza, di conseguenza si evince che il termine *phoque* è tra i più frequenti dell'intero corpus (insieme con *espèce*, *tortue marine*, *corail*, *animal*, etc.).

<sup>224</sup> In particolare, i 13 testi riguardati il sottodominio PHOQUES DE MER (già sottoposti a pulizia e formattazione in .txt, come descritto nel paragrafo § 2.2.2.) sono stati uniti in un unico file – proprio come avvenuto per *ZooCor* – che è stato poi caricato nel software.

Candidat de regroupement	Fréquence	Score (Spécificité)	Variantes orthographiques	Matrice
phoque	2596	404.82	phoque phoques	Nom
gris	1005	205.84	gris grise grises	Adjectif
marin	865	196.95	marin marine marines	Adjectif
espèce	1225	186.75	espèce espèces	Nom
harbour	500	178.51	harbour	Nom
veaux-marins	483	175.81	veaux-marins	Adjectif
plongée	625	175.25	plongée plongées	Nom
phoques veaux-marins	479	175.08	phoques veaux-marins	Nom Adjectif
zone de chasse	369	153.4	zone de chasse zones de chasse	Nom Préposition Nom
chasse	832	150.26	chasse chasses	Nom

Fig. 4.20. Rappresentazione della distribuzione dei termini estratti nella sezione dedicata al sottodominio PHOQUES DE MER con *TermoStat*.

Matrices
Nom= 2763 (32 %)
Nom Adjectif= 1804 (21 %)
Adjectif= 1130 (13 %)
Nom Nom= 982 (11 %)
Nom Préposition Nom= 610 (7 %)
Verbe= 393 (5 %)
PPAdj= 205 (2 %)
Nom PPAdj= 194 (2 %)
Nom Adjectif Adjectif= 151 (2 %)
Nom Préposition Nom Adjectif= 147 (2 %)
Nom Adjectif PPAdj= 76 (1 %)
Nom Adjectif Préposition Nom= 41 (0.5 %)
Nom Préposition Nom Préposition Nom= 23 (0.3 %)
Nom Adjectif Préposition Nom Adjectif= 11 (0.1 %)
Nom Préposition Nom Préposition Nom Adjectif= 10 (0.1 %)
Nom PPAdj Adjectif= 7 (0.1 %)
Nom Préposition Nom Adjectif Adjectif= 4 (0.0 %)
Nom Adjectif Adjectif PPAdj= 4 (0.0 %)
Nom Préposition Nom Préposition Nom Préposition Nom= 3 (0.0 %)
Nom PPAdj Préposition Nom= 3 (0.0 %)
Nom Préposition Nom Adjectif Préposition Nom= 3 (0.0 %)
Nom Préposition Nom Adjectif Préposition Nom Adjectif= 2 (0.0 %)
Nom Adjectif Préposition Nom Préposition Nom Adjectif= 1 (0.0 %)
Nom PPAdj Préposition Nom Adjectif= 1 (0.0 %)
Nom Adjectif PPAdj Adjectif= 1 (0.0 %)

Fig. 4.21. Rappresentazione della distribuzione dei termini di output nella sezione dedicata al sottodominio PHOQUES DE MER con *TermoStat*.

Anche per questa sezione specifica del corpus, l'identificazione delle UT è avvenuta secondo criteri di selezione automatica (indici di frequenza e specificità),

combinati ad un processo di scelta manuale basato sui criteri proposti nel quadro teorico in cui ci inseriamo<sup>225</sup>.

---

<sup>225</sup> Tra cui: definizione di *term* e differenza con *lexie* (l'Homme 2004, *op. cit.*, p.1) e differenza tra termini complessi e termini semplici (L'Homme 2004, *op. cit.*, p. 58) (§ capitolo 3);