

Université Fédérale



Toulouse Midi-Pyrénées

THÈSE



UNIVERSITÉ
DE GENÈVE

FACULTÉ DE TRADUCTION
ET D'INTERPRÉTATION

En vue de l'obtention du DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 2 - Jean Jaurès

Cotutelle internationale : Université de Genève

Présentée et soutenue par

Julie HUMBERT-DROZ

Le 7 septembre 2021

**Définir la déterminologisation : approche outillée en corpus
comparable dans le domaine de la physique des particules**

Ecole doctorale : **CLESCO - Comportement, Langage, Education, Socialisation,
Cognition**

Spécialité : **Sciences du langage**

Unité de recherche :

CLLE - Unité Cognition, Langues, Langage, Ergonomie

Thèse dirigée par

Anne CONDAMINES et Aurélie PICTON

Jury

M. Patrick Drouin, Rapporteur

Mme Agnès Tutin, Rapporteuse

Mme Mathilde Fontanet, Examinatrice

Mme Amélie Josselin-Leray, Examinatrice

Mme Anne Condamines, Directrice de thèse

Mme Aurélie Picton, Directrice de thèse

Faculté de traduction et d'interprétation
Département de traitement informatique
multilingue

Université de Genève

École doctorale CLESCO
Unité de recherche CLLE (UMR 5263)

Université Toulouse – Jean Jaurès

Définir la déterminologisation : approche outillée en corpus comparable dans le domaine de la physique des particules

Thèse présentée pour l'obtention du
Doctorat en Traitement informatique multilingue
et du
Doctorat en Sciences du langage

Julie Humbert-Droz

Soutenue le 7 septembre 2021

JURY

Aurélie Picton, Professeure, Université de Genève, co-directrice
Anne Condamines, Directrice de recherche, CNRS, co-directrice
Agnès Tutin, Professeure, Université Grenoble Alpes, rapporteure
Patrick Drouin, Professeur, Université de Montréal, rapporteur
Amélie Josselin-Leray, Maître de conférence, Université Toulouse – Jean Jaurès, examinatrice
Mathilde Fontanet, Professeure, Université de Genève, examinatrice et présidente du jury

Déclaration attestant le caractère original du travail effectué

J'affirme avoir pris connaissance des documents d'information et de prévention du plagiat émis par l'Université de Genève et la Faculté de traduction et d'interprétation (notamment la *Directive en matière de plagiat des étudiant-es*, le *Règlement d'études de la Faculté de traduction et d'interprétation* ainsi que l'*Aide-mémoire à l'intention des étudiant-es préparant un mémoire de Ma en traduction*).

J'atteste que ce travail est le fruit d'un travail personnel et a été rédigé de manière autonome.

Je déclare que toutes les sources d'information utilisées sont citées de manière complète et précise, y compris les sources sur Internet.

Je suis consciente que le fait de ne pas citer une source ou de ne pas la citer correctement est constitutif de plagiat et que le plagiat est considéré comme une faute grave au sein de l'Université, passible de sanctions.

Au vu de ce qui précède, je déclare sur l'honneur que le présent travail est original.

Nom et prénom : Humbert-Droz Julie

Lieu / date / signature : Genève, le 17 juin 2021



« The meanings of scientific and technical words are likely to change quickly and abruptly with the evolution of the science or the technique. This would seem to be particularly true of scientific words, whose meanings exist only temporarily, possibly even only within a particular theory: they are basically unstable. Their meanings are apt to change every time the current scientific research modifies the theory, particularly in certain branches of science, particle physics for example. »

(Henri Béjoint (1988), « Scientific and Technical Words in General Dictionaries », *International Journal of Lexicography* (1)4, p. 358)

« L’ancrage dans la réalité de certaines notions de la physique actuelle n’est peut-être pas plus solide que celui des licornes et des sirènes de la mythologie. »

(Christine Portelance (1991), « Fondements linguistiques de la terminologie », *META* (36)1, p. 65)

Résumé

Cette thèse porte sur la question de la déterminologisation dans le domaine de la physique des particules. La déterminologisation peut s'envisager à la fois comme un processus de passage de termes d'une langue de spécialité à la langue générale et comme le résultat de ce processus, c'est-à-dire le fonctionnement de termes dans la langue générale. Face au manque de travaux traitant spécifiquement le processus, notre travail vise à combler cette lacune en abordant cette question de manière systématique et approfondie. Nous nous intéressons particulièrement aux phénomènes sémantiques qui se produisent au cours du processus. Par la description de ces phénomènes, nous cherchons à mieux cerner la notion de déterminologisation et ses manifestations dans les textes.

Dans ce but, nous avons développé une méthode de linguistique outillée pour l'analyse des termes dans des textes représentant différentes étapes du processus de déterminologisation. Cette méthode repose sur un corpus comparable et sur deux indices. Les données sont organisées en cinq sous-corpus, constitués de textes qui relèvent de différents genres textuels et degrés de spécialisation. Ces textes sont sélectionnés dans le but d'approcher le continuum entre langue de spécialité et langue générale. En outre, afin de caractériser les différents fonctionnements des termes dans le corpus, deux indices basés sur les contextes distributionnels des termes sont définis.

L'exploration de ces indices dans le corpus révèle la diversité des fonctionnements des termes dans la presse, en comparaison avec les textes spécialisés. Les phénomènes repérés permettent d'alimenter la réflexion sur la nature des changements sémantiques résultant de la déterminologisation et montrent que ces changements reflètent des différences de points de vue ou de conceptualisation, inhérents à la multidimensionnalité des concepts. Les données mettent également en lumière les mécanismes à l'œuvre dans la création des emplois métaphoriques des termes, qui se fondent non seulement sur les usages des termes dans la presse, lorsque de nouvelles connotations sont véhiculées, et sur la coexistence dans la langue générale des composants du terme ou d'unités de la même famille morphologique.

Nos observations contribuent ainsi à définir la déterminologisation comme un processus complexe et non linéaire de passage de termes dans la langue générale, qui fait intervenir de nombreux intermédiaires et qui est influencé par au moins cinq facteurs, aussi bien linguistiques qu'extralinguistiques. Plus largement, nous montrons l'ampleur des questions liées au processus d'intégration et au fonctionnement des termes dans la langue générale. En ce sens, notre thèse participe à reconnaître la transversalité de la déterminologisation dans les questions de circulation des termes, de diffusion des connaissances et de néologie.

Abstract

This thesis deals with the notion of determinologisation in the field of particle physics. This notion can be understood both as the process by which terms move into general language and its result, i.e. the use of terms in general language. Given the lack of research focusing on the process, our thesis aims to fill this gap by systematically analysing determinologisation as the progressive integration of terms into general language. We particularly concentrate on different kinds of semantic changes that are likely to occur in the process. By identifying and describing these changes, we aim to gain a better understanding of determinologisation and the ways in which it is manifested in textual data.

In order to analyse terms in texts that represent different stages of the determinologisation process, we developed a method in corpus linguistics, which relies on a comparable corpus and two linguistic clues. The corpus consists of five sub-corpora that are composed of texts from different textual genres and levels of specialisation. The texts were selected for the corpus to reflect the continuum between language for special purposes and general language. Two clues based on the distribution of terms were defined in order to analyse the different ways in which terms are used in the corpus.

These clues reveal numerous differences regarding the use of terms in newspaper and specialised texts. These phenomena allow us to further discuss the nature of the semantic changes that result from determinologisation. They mainly refer to points of view or different ways to conceptualise the same concept, which lead to the notion of multidimensionality. Results also show how the possibility of using terms as metaphors is created in the corpus. It relies on both the new connotations that are added in newspaper texts and the coexistence in general language of terms and term components or lexical items from the same derivational family.

Our research enables us to define determinologisation as a complex and non-linear process of integration of terms into general language. This process involves different types of intermediaries and is influenced by at least five factors, either linguistic or extralinguistic. From a more general point of view, we show the extent of the issues related to the migration and the use of terms in general language. In this regard, our thesis acknowledges the transversality of determinologisation in phenomena related to term circulation, knowledge transfer and neology.

Remerciements

J'ai eu beaucoup de chance d'avoir été soutenue par de nombreuses personnes pendant mes « années thèse ». J'aimerais exprimer ici toute ma gratitude envers toutes les personnes qui ont répondu présentes lorsque j'en éprouvais le besoin et qui m'ont laissée à mon ordinateur lorsque je me suis plongée dans la rédaction.

Tout d'abord, je souhaite remercier mes directrices de thèse, Aurélie Picton et Anne Condamines, pour leur patience, leur bienveillance, leur disponibilité et leurs encouragements tout au long de ce parcours. Elles ont toujours su trouver les mots pour me faire avancer, me redonner confiance et surmonter mes (trop nombreux) doutes ; cette thèse n'aurait jamais pu aboutir sans leur soutien indéfectible.

Je tiens ensuite à remercier sincèrement Agnès Tutin et Patrick Drouin pour avoir accepté d'être les rapporteurs de cette thèse. Je remercie également Amélie Josselin-Leray et Mathilde Fontanet, qui ont accepté de faire partie de mon jury.

J'aimerais exprimer toute ma gratitude envers Mojdeh Banafsheh et Mercedes Paniccia, qui m'ont initiée à la physique des particules et qui ont joué leur rôle d'expertes à la perfection. Merci pour la richesse de leurs interventions, pour leur implication sans faille et pour avoir toujours répondu sans sourciller à mes questions de terminologie, parfois un peu alambiquées.

Je tiens à remercier l'ensemble des membres du laboratoire CLLE, qui m'ont toujours réservé un accueil des plus chaleureux, et particulièrement Anne Condamines, Cécile Fabre, Mai Ho-Dac, Amélie Josselin-Leray, Anne Le Draoulec, Fabio Montermini, Anne Przewozny, Josette Rebeyrolle, Ludovic Tanguy pour leur gentillesse et les moments passés dans la salle de convivialité. Un merci tout particulier à Amélie, pour sa rigueur et sa bienveillance, ainsi qu'à Anne, qui a toujours su me pousser juste ce qu'il fallait et grâce à qui tout a fini par « jouer ».

Mes séjours toulousains n'auraient pas été aussi inoubliables sans les doctorantes et doctorants CLLE(-ERSS) : Aleksandra, Bénédicte, Camilla, Chiara, Daniele, Émilie, Filip, Giusi, Julie, Karla, Léa, Lena, Lison, Marc-Philippe, Marine, Maxime, Natacha, Nataly, Océane, Pavel, Silvia. Grâce à elles et eux, chacun de mes passages à Toulouse a été une vraie bouffée d'oxygène. Merci pour nos interminables discussions, autour de la linguistique et du reste, les pauses café, les soirées passées à refaire le monde (académique), et les soirées raclette.

Côté genevois, j'aimerais adresser mes plus sincères remerciements à toutes et tous mes collègues « du couloir », et les autres bien entendu. Sans pouvoir nommer tout le monde (la liste

serait sans fin !), je souhaite remercier tout particulièrement Véronique, ma constante à la FTI, pour tous ses conseils avisés et sa bonne humeur inébranlable, Imen, qui n'a cessé de me rappeler que pour terminer cette thèse il fallait surtout que je prenne soin de moi, Emma, pour son énergie inépuisable, son aide précieuse et sa motivation sans faille, Donatella, qui a su me transmettre sa passion pour la terminologie dès le premiers cours (de lexicologie !), qui m'a enseigné tant de choses sur la vraie vie et grâce à qui *terminologie* rimera toujours avec *gastronomie*.

Je souhaite également remercier Estefania, qui a été une collègue et une amie sans pareille, qui a toujours trouvé les mots pour me rebooster et dont le soutien m'a été indispensable dans l'accomplissement de ce travail. Merci d'avoir supporté sans ciller toutes mes crises existentielles ! J'aimerais témoigner toute ma reconnaissance à Aurélie, qui a cru en moi dès le début (et même avant !) et qui aura été bien plus qu'une directrice de thèse. Dire que je lui dois beaucoup serait un bel euphémisme.

Je tiens enfin à remercier ma famille et mes proches, qui ont toujours suivi cette aventure de près ou de loin, se demandant ce que pouvait bien signifier déterminoooo... quoi ? et quel rapport cela pouvait bien avoir avec le boson de Higgs. Une mention spéciale à ma sœur, Marie, qui, pour être passée par là avant moi, sait bien ce que représente l'achèvement d'une thèse, ainsi qu'à ma cousine, Chloé, qui n'a eu de cesse de m'encourager, particulièrement ces derniers mois. Le dernier mais non des moindres, celui sans qui je ne serais jamais arrivée au bout, celui qui m'a toujours apporté un soutien inconditionnel, celui qui a toujours été présent au gré de mes humeurs sur mes montagnes russes émotionnelles : merci Séb d'être mon roc. Je n'aurais jamais pu y arriver sans ton soutien, ton amour, ton rire, ta cuisine, ton écoute, ta gentillesse, et ton inébranlable confiance en moi.

Table des matières

Résumé	ix
Abstract	xi
Remerciements	xiii
Table des matières	xvii
Liste des tableaux	xxiii
Liste des figures	xxiv
Introduction	1
PREMIÈRE PARTIE – APPROCHE DE LA DÉTERMINOLOGISATION : ANCRAGE THÉORIQUE	7
Chapitre 1 Quand le terme sort de la sphère spécialisée	9
1.1. Circulation des termes dans les langues de spécialité et la langue générale.....	9
1.1.1. Origine et évolution de la notion de circulation des termes	9
1.1.2. Langue de spécialité et langue générale : les deux extrêmes d'un continuum	14
1.1.3. La relativité du terme	18
1.1.3.1. Entre <i>terme</i> et <i>mot</i>	18
1.1.3.2. Le terme comme objet d'étude dans le continuum entre langue de spécialité et langue générale.....	20
1.2. Les termes du point de vue de la transmission des connaissances spécialisées vers le grand public	22
1.2.1. Perspective en vulgarisation scientifique	22
1.2.2. Discours de transmission des connaissances et rôle des médias généralistes	25
1.3. Les termes dans les dictionnaires généraux.....	28
1.3.1. Sélection des termes à décrire dans les dictionnaires	29
1.3.2. La définition des termes dans les dictionnaires, un défi pour la lexicographie	32
1.4. Perception néologique des termes en langue générale	34
1.4.1. Intégration des termes comme « nouveaux mots » dans la langue générale.....	35
1.4.2. Changements sémantiques des termes et néologismes sémantiques en langue générale.....	38
1.5. Bilan	42
Chapitre 2 La déterminologisation : les enjeux d'une définition.....	45
2.1. Pluralité des points de vue sur le passage des termes dans la langue générale45	
2.1.1. La proposition de Meyer	45
2.1.1.1. Les termes métaphoriques à la base de la réflexion sur la déterminologisation	46
2.1.1.2. La société du savoir, un accélérateur de déterminologisation ?.....	47
2.1.1.3. Déterminologisation et changements sémantiques	49
2.1.2. La banalisation lexicale selon Galisson	53
2.1.3. La déspecialisation : une notion aux contours flous.....	57
2.1.4. La dédomanialisation : une approche sémique du passage des termes dans la langue générale.....	58
2.2. Dans la continuité des propositions de Meyer	60
2.2.1. Reprise des propositions de Meyer.....	61
2.2.2. Trois axes de développement de la notion de déterminologisation	64
2.2.2.1. Description affinée du fonctionnement des termes dans un contexte non spécialisé.....	64

2.2.2.2.	Prise en compte de la dimension diachronique	69
2.2.2.3.	Approche de la continuité de la déterminologisation, entre langue de spécialité et langue générale.....	73
2.3.	Pour une double définition de la déterminologisation	74
2.3.1.	La déterminologisation comme un résultat	74
2.3.2.	La déterminologisation comme un processus.....	77
2.3.2.1.	Première dimension : le rôle du temps	77
2.3.2.2.	Deuxième dimension : le continuum entre langue de spécialité et langue générale.....	78
2.4.	Bilan	79
DEUXIÈME PARTIE – CONSTRUIRE UNE MÉTHODOLOGIE D’ANALYSE OUTILLÉE EN CORPUS.....		83
Chapitre 3 Outiller l’analyse de la déterminologisation.....		85
3.1.	Pour une démarche en linguistique outillée.....	85
3.1.1.	De la linguistique de corpus à la linguistique outillée	86
3.1.2.	Les outils de la linguistique : quelques précisions.....	89
3.1.3.	Intérêt d’une approche outillée en corpus	92
3.2.	Spécificités de la terminologie outillée	96
3.2.1.	Évolution des théories de la terminologie	96
3.2.2.	Les propositions de la terminologie textuelle.....	100
3.2.2.1.	De la constitution de ressources à la prise en compte de différentes situations de variation terminologique.....	100
3.2.2.2.	Principes d’analyse en terminologie textuelle	102
3.3.	Cadre d’analyse	105
3.3.1.	Construction d’un cadre d’analyse adéquat	105
3.3.2.	Choix des outils	107
Chapitre 4 Constitution de l’objet d’étude		119
4.1.	Choix du domaine de la physique des particules.....	119
4.1.1.	Pertinence de la physique des particules	119
4.1.1.1.	Médiatisation du domaine	119
4.1.1.2.	Proximité avec les locuteurs.....	123
4.1.2.	Précisions sur la physique des particules.....	125
4.2.	Constituer un corpus adéquat pour l’analyse de la déterminologisation	127
4.2.1.	Représentativité et équilibre : deux principes à nuancer.....	128
4.2.2.	Trois principes de constitution pour garantir la pertinence du corpus	132
4.2.2.1.	Représenter un continuum en corpus par l’inclusion de genres textuels diversifiés.....	133
4.2.2.1.1.	Identifier des genres textuels pertinents.....	133
4.2.2.1.2.	Approcher le continuum par l’organisation des données en plusieurs sous-corpus.....	139
4.2.2.2.	Définir une fenêtre temporelle pertinente	141
4.2.2.3.	Assurer la pertinence du corpus par rapport au domaine grâce à une procédure de sélection des textes systématique et objective	143
4.2.2.3.1.	Sélectionner les textes constitutifs des sous-corpus <i>Spécialisé, Communiqués, Rapports</i> et <i>Vulgarisation</i>	143
4.2.2.3.2.	Adapter la procédure au sous-corpus <i>Presse</i>	145
4.3.	Méthode endogène de sélection des termes.....	148
4.3.1.	Repérer des candidats-termes de manière semi-automatique	148
4.3.2.	Filtrer les résultats et constituer un échantillon pertinent	149

Chapitre 5	Indices distributionnels	157
5.1.	Analyse de la distribution en corpus	157
5.1.1.	Définition et principe de l'analyse distributionnelle	157
5.1.2.	Contextes distributionnels	159
5.1.3.	Automatisation de l'analyse distributionnelle	161
5.2.	Définition d'une méthode d'analyse de la distribution adaptée à notre étude	164
5.2.1.	Limites de l'analyse distributionnelle automatique	165
5.2.1.1.	Limites générales des méthodes automatiques actuelles	165
5.2.1.2.	Spécificités de notre objet d'étude pour l'analyse distributionnelle	167
5.2.2.	Définition d'indices distributionnels pertinents	169
5.2.2.1.	Mesurer les différences lexicales dans la distribution des termes : repérage de <i>keywords</i>	169
5.2.2.2.	Affiner l'analyse sémantique des termes : prise en compte des dépendances syntaxiques	173
5.2.3.	Définition d'une stratégie de comparaison en deux temps	176
5.3.	Mise en œuvre des indices	178
5.3.1.	Méthodologie pour le repérage de <i>keywords</i>	178
5.3.1.1.	Extraction des contextes distributionnels des termes	179
5.3.1.2.	Établir le score de <i>keyness</i> : enjeu pour l'identification de <i>keywords</i> pertinents	181
5.3.2.	Méthodologie pour l'analyse des contextes syntaxiques	186
5.3.2.1.	Contrainte de fréquence sur le choix des termes à analyser	186
5.3.2.2.	Repérage et extraction des contextes syntaxiques en corpus	188
TROISIÈME PARTIE – APPRÉHENDER LA DÉTERMINOLOGISATION PAR LA DISTRIBUTION DES TERMES EN CORPUS		195
Chapitre 6	Exploration des <i>keywords</i> en corpus	197
6.1.	Observation globale des <i>keywords</i>	197
6.1.1.	Regroupement des <i>keywords</i>	197
6.1.1.1.	Repérage de catégories homogènes	197
6.1.1.2.	Catégories retenues pour notre travail	202
6.1.2.	Limitation de la subjectivité par le recours à un accord inter-annotateurs	206
6.1.2.1.	Tâche de classification	207
6.1.2.2.	Calcul de l'accord inter-annotateurs et résultats	209
6.2.	Identification de différents types de contextes dans le SC-P	210
6.2.1.	Contextes explicitant l'utilité ou l'utilisation pour d'autres domaines	210
6.2.2.	Contextes évoquant d'autres domaines	213
6.2.3.	Contextes insistant sur des aspects financiers	215
6.2.4.	Contextes mettant en évidence des aspects culturels	216
6.2.5.	Contextes évoquant des aspects extraordinaires	216
6.2.6.	Contextes évoquant des aspects mystérieux	217
6.2.7.	Contextes dénotant des emplois métaphoriques des termes	217
6.3.	Exploration des <i>keywords</i> dans les sous-corpus intermédiaires	220
6.3.1.	Répartition des unités des catégories dans les sous-corpus	220
6.3.2.	Prise en compte des contextes	227
6.3.2.1.	Importance de la thématique financière	228
6.3.2.2.	Interdisciplinarité de la recherche	229
6.3.2.3.	Importance de certains termes dans plusieurs domaines connexes	231
6.3.2.4.	Aspects extraordinaires, hors norme	231
6.3.2.5.	Aspects mystérieux, imaginaires	233
6.4.	Bilan	235

Chapitre 7	Exploration des contextes syntaxiques en corpus	239
7.1.	Point de vue quantitatif : mesure objective des différences de distribution...	239
7.2.	Observation des différences principales entre les SC-P et SC-S	246
7.2.1.	Différents fonctionnements des termes selon le sous-corpus.....	246
7.2.1.1.	Modifieurs entrant dans la composition de termes complexes	246
7.2.1.2.	Adjectifs axiologiques.....	249
7.2.1.3.	Adjectifs dérivés de pays ou de régions.....	250
7.2.1.4.	Verbes faisant partie de collocations spécialisées	251
7.2.2.	Apparition de thématiques différentes selon le sous-corpus	254
7.2.2.1.	Thématiques autour des instruments de recherche.....	254
7.2.2.2.	Thématique autour des particules	257
7.3.	Exploration dans les sous-corpus intermédiaires	260
7.3.1.	Modifieurs entrant dans la composition de termes complexes	260
7.3.2.	Adjectifs axiologiques	263
7.3.3.	Répartition des différents aspects des concepts dans les sous-corpus intermédiaires ...	269
7.4.	Bilan	271
Chapitre 8	Vers une meilleure compréhension du processus de	
	déterminologisation.....	275
8.1.	Saisir la richesse des phénomènes sémantiques en jeu	275
8.1.1.	Préciser la place de la métaphore dans la déterminologisation.....	276
8.1.1.1.	Création du potentiel métaphorique.....	276
8.1.1.2.	Difficultés spécifiques à l'analyse des métaphores	279
8.1.2.	Aborder la déterminologisation à travers la multidimensionnalité	281
8.1.2.1.	Définition de la multidimensionnalité et intérêt pour notre étude	281
8.1.2.2.	Multidimensionnalité à travers les domaines	285
8.1.2.3.	Multidimensionnalité à travers les genres et les degrés de spécialisation.....	288
8.1.3.	Affirmer la place de la déterminologisation dans la néologie	289
8.2.	Comprendre la diversité des intermédiaires	291
8.2.1.	Interroger la continuité du processus par la prise en compte des sous-corpus	
	intermédiaires.....	292
8.2.1.1.	Potentiel métaphorique.....	292
8.2.1.2.	Multidimensionnalité.....	294
8.2.2.	Clarifier la place de la vulgarisation	296
8.2.3.	Définir le double statut de la presse	298
8.2.4.	Identifier d'autres types d'intermédiaires.....	299
8.3.	Redéfinir la déterminologisation comme un processus non linéaire	303
8.3.1.	Reconsidérer la continuité du processus.....	303
8.3.2.	Définir des facteurs de déterminologisation	305
Conclusion		313
Bibliographie		319
Index des auteurs		351
Index thématique		355
ANNEXES		359
Annexe A – Constitution du corpus		361
Annexe B – Liste des termes extraits		365

Liste des tableaux

Tableau 3.1 : Extrait des candidats-terms extrait du sous-corpus Spécialisé avec TermoStat	110
Tableau 3.2 : Exemple d'une phrase lemmatisée et étiquetée avec TreeTagger.....	112
Tableau 3.3 : Exemple d'une phrase analysée syntaxiquement en dépendances par Talismane...	113
Tableau 4.1 : Composition du corpus : genres et degrés de spécialisation	139
Tableau 4.2 : Composition des sous-corpus en termes de genres et de degrés de spécialisation.	141
Tableau 4.3 : Taille du corpus en nombre d'occurrences	147
Tableau 5.1 : Fréquences et scores de keyness pour les unités avoir et immense.....	183
Tableau 5.2 : Extrait des résultats bruts issus de la relation sujet-verbe pour le terme particule .	191
Tableau 5.3 : Extrait des résultats présentés sous forme de tableau	191
Tableau 6.1 : Exemple de la tâche de classification soumise aux participants	208
Tableau 7.1 : Similarité de Jaccard entre les SC-P et SC-S	243
Tableau 7.2 : Similarité de Jaccard entre le SC-S et les sous-corpus intermédiaires	243
Tableau 7.3 : Similarité de Jaccard entre le SC-P et les sous-corpus intermédiaires.....	244
Tableau 7.4 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de boson de Higgs.....	261
Tableau 7.5 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de collisionneur.....	261
Tableau 7.6 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de muon.....	261
Tableau 7.7 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de collisionneur.....	263
Tableau 7.8 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de boson de Higgs.....	264
Tableau 7.9 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de particule	264
Tableau 7.10 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de neutrino	264
Tableau 7.11 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de verbes lorsque particule est objet.....	266
Tableau 7.12 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de détecteur.....	266
Tableau 7.13 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs d'accélérateur.....	267
Tableau 7.14 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de collisionneur.....	267
Tableau 7.15 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de verbes lorsque particule est objet.....	270
Tableau A.1 : Source des textes constitutifs du corpus	362
Tableau A.2 : Répartition des occurrences selon les sources et les périodes considérées	364

Liste des figures

Figure 4.1 : Illustration d'un atome.....	126
Figure 4.2 : Première étape de sélection des termes – extraction.....	148
Figure 4.3 : Seconde étape de sélection des termes – filtres	149
Figure 6.1 : Nombre d'unités attestées dans les sous-corpus intermédiaires, par catégorie.....	222
Figure 6.2 : Fréquence relative de l'ensemble des unités de chaque catégorie, dans les sous-corpus intermédiaires et le SC-P	224
Figure 6.3 : Fréquence relative des unités de la catégorie « sensationnalisme » dans les sous-corpus intermédiaires et dans le SC-P	225
Figure 6.4 : Fréquence relative des unités de la catégorie « médical » dans les sous-corpus intermédiaires et dans le SC-P	226
Figure 6.5 : Fréquence relative des unités <i>dollar</i> , <i>euro</i> et <i>franc</i> dans les SC-P, SC-C et SC-R	228
Figure 8.1 : Représentation schématique du passage de termes dans la presse à partir de textes de domaines variés.....	302

Introduction

L'intégration dans la langue générale de termes spécifiques à un domaine spécialisé est un phénomène assez courant, qui témoigne des dynamiques linguistiques à l'œuvre entre les langues de spécialité et la langue générale. Ce phénomène soulève cependant de nombreuses questions, à propos des termes, de leur sens ou encore de leur statut dans la langue générale. En terminologie, l'étude de ces questions est rendue possible par l'émergence et le développement de propositions théoriques qui se distinguent de la Théorie Générale de la Terminologie (TGT) et qui visent un rapprochement avec la linguistique, d'un point de vue tant théorique que méthodologique.

Un changement majeur au niveau théorique concerne le terme. Dans ces propositions, en effet, le point de vue sur le terme se démarque de celui de la TGT par le fait que celui-ci n'est plus considéré comme la seule étiquette du concept. Au contraire, les approches plus récentes de la terminologie reconnaissent au terme un statut d'unité lexicale, possédant un sens spécialisé à l'intérieur d'un domaine. Dans cette perspective, le terme est soumis à la variation, comme toute autre unité de la langue, et fonctionne dans différents types de discours. Le terme est donc une entité « mobile » (Dury, 2008 : 231), dont l'usage s'affranchit des frontières traditionnellement dressées entre les langues de spécialité et la langue générale. Le rapprochement avec la linguistique sur le plan méthodologique conduit par ailleurs les chercheurs à tirer profit des méthodes de linguistique de corpus et de linguistique outillée pour aborder plusieurs problématiques terminologiques. Ces méthodes rendent désormais possible l'observation des termes dans des données textuelles attestées, qui sont analysées à l'aide de différents outils informatiques. La prise en compte des usages réels des termes permet alors de reconnaître pleinement et d'explorer la variabilité des termes et de leur sens, selon les discours dans lesquels ils fonctionnent. Lorsque les termes « s'émancipent » des discours des experts et intègrent la langue générale, leur fonctionnement peut être appréhendé par le prisme de la déterminologisation.

Ce phénomène décrit les changements sémantiques qui se produisent lorsque les termes « attire[nt] l'attention du public » (Meyer et Mackintosh, 2000a : 199) et sont employés dans un contexte non spécialisé, par des locuteurs qui ne sont pas nécessairement experts du domaine. Deux tendances concernant ces changements sont généralement observées : la première tendance concerne des changements plutôt peu conséquents, qui résultent de l'emploi des termes par des locuteurs ne possédant pas la connaissance du domaine, dans des situations de communication non spécialisées. Dans ce cas, les termes ont un sens moins précis et de nouvelles connotations peuvent apparaître dans un contexte non spécialisé. La seconde tendance rend compte de changements plus « profonds », qui s'éloignent dans une plus large mesure de l'usage des termes dans une situation

de communication spécialisée et qui peuvent donner lieu à de nouveaux sens par métaphore ou par métonymie.

La déterminologisation décrit également le processus de passage des termes dans la langue générale. Ce processus est double : les termes s'intègrent dans la langue générale progressivement dans le temps et cette intégration s'effectue par le biais de différents canaux, qui se situent entre langue de spécialité et langue générale. Ces canaux peuvent être de différentes natures ; il peut s'agir par exemple des médias généralistes, qui jouent un rôle reconnu dans la diffusion des termes, ou de textes de vulgarisation, dont l'objectif est de transmettre des connaissances spécialisées auprès de locuteurs non spécialistes.

Bien qu'il soit reconnu que la déterminologisation est un phénomène qui s'envisage à la fois comme un processus et comme un résultat, le manque de travaux abordant de manière systématique et approfondie la question du processus est criant. Le besoin de description est particulièrement présent en ce qui concerne les différents canaux par lesquels les termes passent et leur rôle dans les fonctionnements observés dans la langue générale. À ce jour, en effet, les études existantes de la déterminologisation ne tiennent que très peu compte de cet aspect, pourtant crucial si l'on souhaite parvenir à une meilleure compréhension du processus de déterminologisation dans sa globalité.

C'est dans ce contexte que s'inscrit notre thèse. À travers une analyse outillée et systématique des termes dans des textes rédigés dans des situations de communication dont nous faisons l'hypothèse qu'elles jouent un rôle dans la déterminologisation, nous cherchons à définir ce phénomène de manière plus complète et en tenant compte de sa dualité. En outre, afin de saisir la diversité des changements sémantiques qui sont susceptibles de se produire au cours du processus, et non uniquement les plus saillants, nous nous focalisons sur les termes d'un domaine précis, la physique des particules.

Plus précisément, c'est par la description détaillée du fonctionnement des termes dans un contexte non spécialisé que nous serons à même de mieux cerner le processus de déterminologisation. Mais notre approche de la déterminologisation se démarque de la majorité des travaux portant sur cette question, puisque nous l'abordons par l'étude d'un corpus comparable contenant des textes non spécialisés et des textes spécialisés, mais également des textes semi-spécialisés et vulgarisés. La prise en compte de la déterminologisation à la fois comme résultat et comme processus nous permet ainsi de développer notre réflexion sur les dynamiques du changement sémantique entre langue de spécialité et langue générale.

Axer la réflexion tant sur le résultat que sur le processus d'intégration des termes dans la langue générale fait cependant surgir plusieurs questionnements théoriques et méthodologiques. D'un point de vue théorique, un premier défi consiste à circonscrire notre objet d'étude, en clarifiant les rapports que la déterminologisation entretient avec d'autres notions apparentées et d'autres approches qui s'intéressent aux termes dans la langue générale. Parallèlement, le manque de définition précise de la déterminologisation en tant que passage d'une langue de spécialité à la langue générale engendre un second défi, qui s'articule essentiellement autour de l'identification des canaux par lesquels les termes sont susceptibles d'atteindre la langue générale. Par ailleurs, si notre point de vue est résolument terminologique, considérer le fonctionnement des termes dans leur passage d'une langue de spécialité à la langue générale amène nécessairement à tenir compte des spécificités à la fois des langues de spécialité et de la langue générale. Tout au long de notre thèse, nous mobilisons ainsi des outils théoriques et méthodologiques qui permettent de concilier ces différents points de vue, afin de construire une interprétation des phénomènes relevés qui soit valide théoriquement.

Une seconde série de questionnements, d'ordre méthodologique, se pose ensuite. Afin d'interroger le fonctionnement des termes dans le processus de déterminologisation, nous mettons à profit les acquis de la linguistique outillée et de la terminologie textuelle, et basons notre réflexion sur une analyse de corpus. Notre démarche s'articule alors autour de la constitution et de l'exploration d'un corpus adéquat pour le phénomène étudié et de la définition d'indices pertinents observables en corpus. Ce type d'approche implique plusieurs défis : premièrement, puisque rassembler des données textuelles en corpus impose de figer les dynamiques linguistiques en une représentation statique, la constitution d'un corpus servant à l'observation d'un processus doit répondre à des exigences particulières, afin non seulement d'approcher la continuité qui caractérise le processus de déterminologisation, mais également de permettre son étude. Deuxièmement, l'analyse de ces données requiert la mise en œuvre d'une méthodologie pertinente, qui tient compte des spécificités du corpus. Dans la perspective d'une analyse basée sur la distribution des termes, nous définissons deux indices, pensés pour permettre l'analyse fine de l'environnement des termes et la mise au jour de phénomènes que nous pouvons interpréter en relation avec la déterminologisation. Afin de garantir la pertinence de ces indices pour notre objectif, leur définition passe par une importante réflexion sur les manifestations possibles de la déterminologisation, telles qu'elles sont observables en corpus. Parallèlement, l'élaboration et la mise en œuvre de ces indices dans le corpus reposent sur des critères objectifs et transparents, satisfaisant par conséquent une exigence de reproductibilité de notre démarche.

L'exploration outillée d'un corpus conçu spécifiquement pour l'étude du processus de déterminologisation, à l'aide de deux indices interrogeant la distribution des termes dans le corpus, nous fournit ainsi les éléments nécessaires pour mener une réflexion approfondie sur ce phénomène et sur les changements sémantiques des termes susceptibles de se produire.

Notre réflexion s'articule autour de trois axes, organisés en trois parties dans ce manuscrit. Dans la première partie, nous exposons les bases théoriques qui sous-tendent notre travail. Dans le Chapitre 1, nous dressons un panorama des travaux abordant les questions du passage et du fonctionnement des termes dans la langue générale, afin de situer notre thèse dans la problématique globale de la circulation des termes. Nous insistons particulièrement sur la pluralité de ces approches et précisons la place de la déterminologisation parmi les phénomènes traités. Le Chapitre 2 se focalise plus spécifiquement sur la déterminologisation. Nous contextualisons notre réflexion en clarifiant les rapports qu'elle entretient avec d'autres notions apparentées, puis nous tirons profit des travaux existants qui développent cette question pour en proposer une définition actualisée. Cette définition met l'accent sur la continuité du processus, entre langue de spécialité et langue générale, et constitue la base de notre travail de constitution et d'analyse de données.

La deuxième partie de notre thèse est consacrée à l'élaboration d'un cadre méthodologique adéquat pour observer le processus de déterminologisation en corpus comparable, à l'aide d'une méthode de linguistique outillée. La construction de ce cadre passe par trois étapes principales, détaillées dans trois chapitres. Dans le Chapitre 3, nous posons les bases méthodologiques de notre réflexion. Dans ce but, nous montrons l'intérêt de la linguistique outillée pour l'exploration de corpus textuels et précisons les spécificités de l'analyse du fonctionnement des termes en corpus, à travers les méthodes de terminologie outillée. Nous reprenons ensuite à notre compte les principes de la terminologie textuelle afin de construire notre démarche globale. Le Chapitre 4 se concentre sur la constitution de notre objet d'étude, à travers la délimitation d'un domaine pertinent, la compilation d'un corpus comparable adéquat et la sélection des termes à analyser à partir d'une méthode endogène. Le corpus est au cœur de notre démarche et sa constitution nécessite de fixer de nombreux paramètres, afin d'assurer sa pertinence pour l'étude du processus de déterminologisation, pour le domaine et pour les indices mis en œuvre. La définition de ces indices fait l'objet du Chapitre 5. Afin d'appréhender les changements sémantiques des termes au cours de la déterminologisation, nous mettons en œuvre une démarche d'analyse de la distribution des termes dans le corpus, à travers deux indices. Nous insistons sur la complémentarité de ces indices et justifions la nécessité d'aborder l'analyse de la distribution de manière outillée, et non automatique. Ce choix vise alors à saisir finement les différents fonctionnements des termes dans

le corpus, dans le but d'aboutir à une description détaillée des changements possibles au cours du processus de déterminologisation.

La dernière partie de notre thèse est dédiée à l'exploration des indices en corpus et à la mise au jour d'éléments contribuant à une redéfinition de la déterminologisation. Dans le Chapitre 6, nous détaillons la mise en œuvre du premier indice distributionnel, à partir de l'ensemble des unités apparaissant dans la distribution des termes à l'étude et à l'aide d'une méthode de repérage de keywords. Dans le Chapitre 7, nous nous focalisons sur l'analyse des unités qui entretiennent une relation de dépendance syntaxique avec les termes. Dans ces deux chapitres, nous nous concentrons également sur l'exploration de la dimension continue de la déterminologisation entre langue de spécialité et langue générale. La richesse de nos observations dans le corpus nous permet d'argumenter, dans le Chapitre 8, en faveur d'une redéfinition de la déterminologisation, en tant que processus complexe et non linéaire de passage de termes dans la langue générale, qui s'effectue par le biais de différents types d'intermédiaires et qui se manifeste par une grande diversité de changements susceptibles de se produire dans les fonctionnements des termes, selon plusieurs facteurs, aussi bien linguistiques qu'extralinguistiques.

Première partie
Approche de la
déterminologisation : ancrage théorique

Chapitre 1 **Quand le terme sort de la sphère spécialisée**

Dans ce premier chapitre, nous cherchons à situer la question de la déterminologisation par rapport aux différents champs de recherche qui abordent le passage et/ou le fonctionnement des termes dans la langue générale. Dans ce but, nous expliquons d'abord les raisons qui nous amènent à considérer que la déterminologisation s'inscrit dans la problématique globale de la circulation des termes, puis nous exposons les différentes situations dans lesquelles les termes peuvent circuler. Envisager les termes comme des objets qui circulent soulève par ailleurs plusieurs questions fondamentales en terminologie, qui remettent en question les frontières habituellement dressées entre les langues de spécialité et la langue générale.

À travers un point de vue que nous souhaitons large, nous montrons ensuite comment les aspects qui nous intéressent sont abordés dans des travaux qui se situent dans divers champs de recherche. Nous dressons ainsi un portrait du terme et des problématiques qu'il soulève lorsqu'il est utilisé dans des textes qui visent à transmettre des connaissances spécialisées auprès du grand public, par le biais de la vulgarisation scientifique ou non (section 1.2), lorsqu'il est enregistré et décrit dans des dictionnaires de langue générale (section 1.3) et lorsqu'il est considéré du point de vue du renouvellement lexical de la langue (section 1.4). Dans chacune de ces sections, nous présentons les enjeux des différents travaux pris en compte, qui touchent directement à notre objet d'étude, la déterminologisation.

1.1. Circulation des termes dans les langues de spécialité et la langue générale

De manière générale, la circulation des termes renvoie au fait que les termes ne sont pas toujours utilisés exclusivement par des experts, dans des situations de communication qui relèvent d'un domaine en particulier, mais qu'ils « voyagent », au contraire, dans différentes langues de spécialité ou dans la langue générale. Cette perspective suppose de considérer les langues de spécialité et la langue générale comme des ensembles à travers lesquels des échanges sont permis. Mais en réalité, le phénomène de circulation des termes est plus complexe, interroge plusieurs notions fondamentales de la terminologie et pose de nouvelles questions.

1.1.1. Origine et évolution de la notion de circulation des termes

La question de la circulation des termes entre les langues de spécialité et la langue générale n'est pas nouvelle, comme le rappelle notamment Delavigne (2020). Les travaux de Guilbert dans

les années 1970 sont, à ce titre, fondateurs¹. Sous l'angle de la créativité lexicale, l'auteur développe en effet une abondante réflexion sur la « mobilité » des unités lexicales, qui passent continuellement de la langue générale aux langues de spécialité, des langues de spécialité à la langue générale et d'une langue de spécialité à une autre langue de spécialité.

Dans les deux premiers cas, Guilbert évoque ainsi un « mouvement d'emprunt constant entre le vocabulaire général et les vocabulaires techniques » (Guilbert, 1973a : 6-7) ou « un double mouvement », qui passe « par la spécialisation dans les vocabulaires particuliers de termes de la langue commune et par la migration de termes techniques et scientifiques dans le vocabulaire général » (Guilbert, 1971 : 49), ou encore un « mouvement de va-et-vient et d'emprunt mutuel » (Guilbert, 1975 : 80). Dans le troisième cas, Guilbert décrit plus particulièrement l'idée d'un passage de termes d'une sphère spécialisée à une autre : « le terme sort du groupe professionnel spécialisé où l'équivalence définitoire de la dénomination demeure » et « entre dans un autre groupe » (Guilbert, 1975 : 82). L'idée de mouvement, ou de passage, est omniprésente dans les travaux de cet auteur, comme l'illustrent les deux citations suivantes :

« [o]n a pu constater qu'un grand nombre de lexèmes, qui fonctionnent comme signes à l'intérieur de ce champ, appartiennent en même temps à une autre sphère conceptuelle et qu'ils sont passés d'une sphère à l'autre par un transfert de signification. » (Guilbert, 1965 : 68)

« le passage d'un terme d'un vocabulaire spécialisé à un autre vocabulaire spécialisé le charge d'une signification nouvelle qui ne provient pas seulement du référent nouveau mais aussi du milieu professionnel des locuteurs. » (Guilbert, 1973b : 23)

L'utilisation des unités *mouvement* et *passage* dans ces extraits dénote en particulier le caractère « mobile » des termes et des unités lexicales à l'intérieur d'une langue et, bien que n'abordant pas cette question en ces termes exactement, Guilbert souligne ici les dynamiques existant entre les langues de spécialité et la langue générale, qui sous-tendent la création de nouveaux termes ou de nouvelles unités. Cette idée sera reprise et étoffée tout au long de notre travail.

Plusieurs auteurs, se revendiquant d'une approche sociolinguistique de la terminologie et se situant dans la continuité de Guilbert², adoptent le terme *circulation* pour rendre compte de ces

¹ Bien que Meillet ait développé des idées tout à fait similaires dans son célèbre article *Comment les mots changent de sens* (Meillet, 1905), et bien que Guilbert se soit largement inspiré des idées de Meillet, c'est essentiellement à partir des travaux de Guilbert que les réflexions sur la circulation des termes se sont par la suite développées en socioterminologie.

² D'ailleurs, du fait qu'il évoque explicitement l'intérêt sociologique et sociolinguistique des mécanismes complexes de créativité lexicale (Guilbert, 1975 : 81), Guilbert est parfois considéré comme l'un des précurseurs de la socioterminologie (par Gaudin (1993) ou Humbley (2018b) notamment).

mouvements et de leur complexité. Différents syntagmes apparaissent alors dans la littérature, notamment *circulation terminologique*, chez Guespin (1991) et Gambier (1991b), *circulation langagière* (Gaudin, 1993) et, le plus répandu, *circulation des termes* (Gaudin, 1990, 1991, 2003, 2005 ; Gambier, 1991a ; Delavigne, 2001, 2020), que nous reprenons ici. Les recherches dans ce domaine s'attachent ainsi dans une large mesure à la description du fonctionnement réel des termes et de leur sens lorsqu'ils circulent en dehors des textes de spécialistes. À travers une approche sémantique des termes (par exemple Delavigne et Bouveret, 2000a ; Delavigne, 2001 ; Gaudin, 2003), c'est dans la diversité des situations, des textes et des objectifs dans lesquels les termes sont employés que leur sens est appréhendé et que les modifications de sens entraînées par ces mouvements sont décrites. Les travaux en socioterminologie visent ainsi à mieux comprendre les enjeux de la circulation des termes sur un plan sociolinguistique (Gambier, 1991b : 13 ; Gaudin, 2005 : 82)³.

L'idée que les termes circulent entre plusieurs langues de spécialité et entre une langue de spécialité et la langue générale est par ailleurs tout à fait répandue en terminologie. Comme le dit Delavigne (2020 : §1), « [o]n a beau faire, les termes jouent l'indiscipline et ne restent guère cantonnés à l'intérieur des sphères d'activité qui les voient naître ». De ce point de vue, la circulation des termes en dehors de leur langue de spécialité « d'origine » serait inévitable. Il n'est donc pas étonnant de retrouver ces questions dans nombre de travaux traitant de terminologie, dans une perspective de néologie (à l'instar des travaux de Guilbert) ou dans d'autres perspectives, que nous développons *infra*. Citons à titre d'exemple Varantola (1986), Boulanger et L'Homme (1991), Gentilhomme (1994a), Cabré (1994), Dury (2000, 2008), Depecker (2002), Celotti et Musacchio (2004), Oliveira (2005, 2009), Ungureanu (2006), Nicolae et Delavigne (2013), Botta (2013), Vicari (2018) ou encore L'Homme (2020).

Dans ces travaux, l'idée que les termes « voyagent » entre les langues de spécialité et la langue générale est exprimée de plusieurs manières, en reprenant les termes utilisés en socioterminologie ou en utilisant des formulations alternatives, dont certaines sont d'ailleurs plus proches de l'idée de mouvement exprimée par Guilbert. Par exemple, dans un article portant sur l'apport de la diachronie dans l'enseignement des langues de spécialité, Celotti et Musacchio évoquent le « nomadisme des mots » (Celotti et Musacchio, 2004 : 267). Les autrices⁴ expliquent que

³ Les développements de la terminologie permettant la prise en compte de ces problématiques plus variées sont traités dans le Chapitre 3.

⁴ Dans ce cas, *langue commune* est entendu dans le même sens de *langue générale*.

« [s]uivre des mots à travers leur histoire permet aussi d'appréhender leur "nomadisme" : leur voyage d'une discipline à l'autre, d'une discipline à la langue commune, de la langue commune à la spécialisée. » (*Ibid.*)

De la même manière, Ledouble traite de l'instabilité conceptuelle qui résulte de la circulation, en dehors des discours des spécialistes, des termes du domaine de l'agriculture et de la protection de l'environnement, plus précisément dans le contexte de la protection des plantes par différents moyens (naturels ou synthétiques) et de la perception de ces moyens par la société, notamment les conséquences néfastes des pesticides sur l'environnement et sur la santé (Ledouble, 2019, 2020). L'autrice évoque alors des problématiques liées au « glissement terminologique d'un domaine spécialisé vers la langue générale » (Ledouble, 2019 : 60), auxquelles elle fait également référence en parlant de « migration des termes d'un domaine scientifique ou technique vers le grand public » (*Ibid.* : 63).

Cependant, en dépit d'une certaine richesse lexicale autour de la notion de circulation des termes, il semble que l'idée de mouvement soit celle qui ait retenu l'attention d'un plus grand nombre de chercheurs. Cabré explique ainsi qu'il « se produit un mouvement permanent de termes du lexique général au lexique de spécialité, du lexique des spécialités au lexique commun, et des lexiques des spécialités entre eux » (Cabré, 1994 : 591). Citons également Durieux, qui « observe que les mouvements entre langue usuelle et langue spécialisée se font à double sens » (Durieux, 1996-1997 : 91-92), Botta, qui parle de « "mouvements migratoires" des termes entre les langues de spécialité (LSP) et la langue courante, et vice-versa » (Botta, 2013 : 277) ou encore Meyer et Mackintosh, qui, en mettant l'accent sur les aspects sémantiques, parlent de « mouvements de flux et de reflux du sens entre la langue générale et les langues spécialisées » (Meyer et Mackintosh, 2000a : 214).

Enfin, notons que, dans des ouvrages plus généraux sur la terminologie tels que celui de Rondeau (1981), l'idée de transferts d'unités de la langue générale à une langue de spécialité ou d'une langue de spécialité à la langue générale est également évoquée :

« en principe, rien n'interdit à une forme linguistique de se retrouver à la fois dans les deux zones [zone de la langue commune et zone des langues de spécialité], ce qui peut se produire de deux manières, soit dans le sens Lc → Lsp, soit dans le sens Lsp → Lc. » (Rondeau, 1981 : 26)

Avec ces quelques observations, nous voulons montrer que, comme le dit Josselin-Leray,

« il est maintenant bien reconnu qu'il s'opère de constants mouvements de va-et-vient ou d'interférence entre la langue générale – ou langue commune – et la langue spécialisée, dans un sens comme dans l'autre. » (Josselin-Leray, 2005 : 33)

Remarquons toutefois qu'en dépit de ce consensus, tous les phénomènes liés à la circulation des termes ne sont pas nécessairement pris en compte, selon les auteurs. Nous distinguons ainsi des cas où les mouvements sont envisagés uniquement entre une langue de spécialité et la langue générale, indépendamment de la direction de ces mouvements (comme l'illustrent les citations de Durieux, Meyer et Mackintosh, Rondeau ou encore Josselin-Leray ci-dessus), des cas où ces mouvements se produisent d'une langue de spécialité à la langue générale ou à une autre langue de spécialité et enfin des cas où toutes ces possibilités sont envisagées. C'est notamment le choix d'Ungureanu, qui inscrit son étude dans le champ de la néologie sémantique et cherche à dévoiler « la dynamique interne du système lexical » et à interroger « le rôle de la langue commune en tant que source de matière première pour la terminologie » (Ungureanu, 2006 : 18-19)⁵. Dans ce but, l'auteur explore les « trois processus de base [qui] amorcent et maintiennent cette circulation [de termes] » (*Ibid.* : 121), c'est-à-dire :

- le phénomène de **terminologisation**, qui renvoie aux mécanismes par lesquels le sens des unités de la langue générale se spécialise afin de désigner des concepts du domaine de l'informatique, dans le cas de cette étude (*Ibid.* : 62-63) ;
- le phénomène de **déterminologisation**, qui renvoie au transfert de termes des langues de spécialité vers la langue générale ainsi qu'aux changements de sens que les termes subissent nécessairement (*Ibid.*) ;
- et enfin la **nomadisation** des termes entre plusieurs langues de spécialité, qui renvoie au passage de termes d'une langue de spécialité à une autre et à la création de termes dans une langue de spécialité à partir de termes appartenant à une autre langue de spécialité (*Ibid.*).

L'idée de circulation langagière (ou de mouvements entre langue de spécialité et langue générale) implique donc ces trois phénomènes différents. Cependant, dans notre travail, nous nous focalisons particulièrement sur les mouvements de termes qui se produisent d'une langue de spécialité vers la langue générale⁶, que nous désignons, à l'instar de Meyer et Mackintosh et d'Ungureanu notamment, par le terme *déterminologisation*. Ce phénomène amène différents

⁵ Entendu ici comme *langue générale*.

⁶ Nous renvoyons néanmoins les lecteurs aux travaux de Sager (notamment 1997), Cabré (1994, 2005, 2016 notamment), Calberg-Challot (2007, 2015), Dury (2008) ou encore à la norme ISO 1087:2019 (ISO, 2019) à propos de la terminologisation ; aux travaux de Sager (1997), Celotti et Musacchio (2004), Stengers (1987) ou à la norme ISO 1087:2019 (sous *emprunt transdisciplinaire*, 2019 : 12) sur la nomadisation des termes ; et aux travaux d'Assal (1995), Rossi (2009, 2016), Gaudin (2003), Humbley (2005, 2018a) ou Temmerman (2000) lorsque la création terminologique s'effectue à partir d'unités de la langue générale ou de termes d'autres domaines par le biais de métaphores.

questionnements théoriques et méthodologiques, et remet en cause en particulier deux distinctions habituellement considérées comme fondamentales en terminologie, entre *langue de spécialité* et *langue générale*, d'une part, et entre *terme* et *mot* d'autre part. Afin de mieux situer notre propos et de définir la perspective que nous adoptons dans notre travail, nous discutons ces dichotomies dans les deux sections suivantes.

1.1.2. Langue de spécialité et langue générale : les deux extrêmes d'un continuum

La première distinction fondamentale que nous souhaitons traiter est celle qui est traditionnellement opérée entre les langues de spécialité et la langue générale⁷, car c'est précisément cette distinction entre les ensembles formés par les langues de spécialité et par la langue générale qui permet de conceptualiser les « déplacements » des termes ou des unités lexicales de l'un à l'autre. Parallèlement, il est reconnu que ces deux ensembles ne sont pas totalement distincts l'un de l'autre et forment plutôt un continuum, avec des situations de communication plus ou moins spécialisées, qui relèvent plus ou moins des langues de spécialité ou de la langue générale. Dans cette section, nous cherchons donc à contextualiser la réflexion qui permet de considérer ce continuum, à expliciter l'intérêt de cette conception et à préciser les choix que nous faisons dans notre thèse.

De manière générale, il est admis que les langues de spécialité se distinguent de la langue générale à l'intérieur d'une langue donnée par le fait que les langues de spécialité renvoient à des usages caractéristiques de situations de communication spécialisées. Dans cette perspective, Cabré, par exemple, précise que les langues de spécialité sont utilisées dans des situations où les sujets sont spécialisés et où les locuteurs sont des spécialistes de ces sujets (Cabré, 1998 : 124-125). Par « sujets spécialisés », l'auteur entend les sujets « dont les contenus notionnels ne sont généralement pas partagés par l'ensemble des locuteurs d'une langue, et qui nécessitent un apprentissage particulier » (*Ibid.* : 125). À l'inverse, la langue générale se constitue d'un « ensemble de règles, d'unités et de restrictions qui font partie des connaissances de la majorité des locuteurs » (*Ibid.* : 118).

Ainsi, même si « la différence entre la langue générale [...] et les langues de spécialité est difficile à résoudre » (*Ibid.*: 134), la distinction généralement opérée entre ces deux ensembles

⁷ Précisons qu'il existe différentes dénominations, qui sont le reflet de plusieurs nuances selon l'auteur et le contexte dans lequel elles sont développées. Sans prétendre à une liste exhaustive, et sans vouloir entrer dans les débats existants sur les nuances entre les différents choix, les dénominations suivantes peuvent être observées en français : *langue spécialisée* (Lerat, 1995 ; Cabré, 1998 ; Delavigne, 2001 ; Depecker, 2002), *vocabulaire spécialisé* (Dury, 2000), *vocabulaire de spécialité* (Calberg-Challot, 2007) pour *langue de spécialité* et *langue commune* (Cabré, 1998 ; Thoiron, 1998), *langue usuelle* (Durieux, 1996-1997), *vocabulaire général* (Guilbert, 1973a ; Gaudin, 1993), *langue courante* (Petit, 2001 ; Josselin-Leray, 2005 ; Péraldi, 2011) pour *langue générale*. Dans notre travail, en revanche, nous choisissons de toujours parler de *langue de spécialité* et de *langue générale*.

repose essentiellement sur le besoin de distinguer certaines situations de communication spécialisées, associées à des connaissances spécialisées pour lesquelles il existe des experts, identifiables et reconnus comme tels, d'autres situations de communication, qui ne sont pas spécialisées. Cette idée est largement partagée et se retrouve dans nombre de propositions différentes, tant pour définir les langues de spécialité que pour définir la langue générale (Rondeau, 1981 : 26 ; Varantola, 1986 : 11 ; Lerat, 1995 : 20 ; Kocourek, 1991 : 21 ; Delavigne, 2001 : 172 ; Bowker et Pearson, 2002 : 25*sqq.* ; Ungureanu, 2006 : 37*sqq.* parmi d'autres). La notion de spécialisation permet dans ce cas de contraster entre ce qui relève d'une langue de spécialité, c'est-à-dire lorsque l'on peut observer une situation de communication spécialisée, et ce qui relève de la langue générale, c'est-à-dire lorsque la situation de communication n'est pas spécialisée. Formulée ainsi, cette distinction est donc tout à fait claire.

Mais cette clarté n'est qu'apparente et renvoie plutôt, selon nous, au fait que « we are dealing with two intuitively correct assumptions that are good as working concepts but which resist clear-cut definition and delimitation » (Varantola, 1986 : 10). Dans cette perspective, nous pensons en effet que le travail sur les langues de spécialité et sur les terminologies qui y sont associées amène à développer une idée prototypique, ou une intuition, de ce que sont les langues de spécialité et de ce qu'est la langue générale. Si cette intuition peut parfois s'avérer suffisante, dès lors que nous nous intéressons à un phénomène tel que celui de la circulation des termes ou celui de la déterminologisation, celle-ci ne suffit plus, et ce pour plusieurs raisons.

Premièrement, l'un des défauts de ces définitions concerne leur circularité et, par extension, le manque de caractéristiques intrinsèques permettant de délimiter clairement les langues de spécialité et la langue générale. De fait, *langue de spécialité* se définit alors par opposition avec *langue générale* et *langue générale* par opposition avec *langue de spécialité* (Rondeau, 1981 : 26 ; Delavigne, 2001 : 174 ; Bowker et Pearson, 2002 : 25 ; Gaudin, 2003 : 47). De la même manière, lorsque des caractéristiques sont précisées, celles-ci restent difficiles à saisir. Par exemple, pour ce qui est de la langue générale, il semble particulièrement difficile d'identifier ce qui fait « partie des connaissances de la majorité des locuteurs » (Cabré, 1998 : 118) ou ce qui constitue la partie de la langue « that we use every day to talk about ordinary things in a variety of common situations » (Bowker et Pearson, 2002 : 25).

C'est également ce que Thoiron⁸ met en évidence :

« [q]uand on parle de langues de spécialités [...], on cherche souvent à les distinguer de la langue commune. Il apparaît de plus en plus clairement maintenant qu'une telle distinction est certes intéressante et fondamentalement

⁸ L'auteur utilise ici *langue commune* dans le sens de *langue générale*.

utile mais qu'elle fait surgir une série inquiétante de difficultés dès qu'on veut établir des listes de critères pertinents. » (Thoiron, 1998 : 621)

Deuxièmement, la notion de spécialisation servant à caractériser les langues de spécialité est plus complexe qu'il n'y paraît de prime abord (Cabré, 1998 : 123 ; Pearson, 1998 : 26-28 ; Delavigne, 2001 : 170*sqq.* ; Condamines, 2003 : 26). Une difficulté non négligeable consiste à déterminer les caractéristiques qui permettent de décider qu'une situation de communication est véritablement spécialisée. Par exemple, Cabré considère que « [l]e sujet spécialisé d'une communication donnée peut alors être le critère qui détermine la spécialisation d'un texte ou d'une langue de spécialité » (Cabré, 1998 : 121), mais l'auteurice reconnaît également que ce critère seul ne suffit pas, notamment « parce que la vie de tous les jours comprend pour chaque individu une quantité d'activités spécialisées, même si on ne les perçoit pas comme telles » (*Ibid.*). Ce dernier point semble par ailleurs contradictoire avec l'idée que la langue générale serait utilisée « to discuss ordinary things in everyday situations » (Bowker et Pearson, 2002 : 39), puisque « toute activité humaine nécessite [...] une certaine spécialisation » (Cabré, 1998 : 121).

Par ailleurs, même si, à l'instar de Cabré, nous ajoutons au critère du sujet spécialisé « des conditions de nature pragmatique, comme le type de situation et de locuteurs » (*Ibid.* : 123), il existe en réalité une pluralité de situations de communication et de types de locuteurs, qui rendent cette distinction particulièrement inopérante, comme le suggère notamment Pearson (1998 : 35*sqq.*). Pour ces raisons, c'est une conception des langues de spécialité et de la langue générale formant un continuum qui semble désormais dominer (Thoiron, 1991 : 629 ; Gaudin, 1993 : 130 ; Durieux, 1996-1997 : 91 ; Barona, 1998-1999 : 109 ; Delavigne et Bouveret, 2000a : 10 ; Dury, 2000 : 26 ; Delavigne, 2001 : 179 ; Depecker, 2002 : Partie 1, §106 ; Josselin-Leray, 2005 : 34 ; Poibeau, 2005 ; Ungureanu, 2006 : 62-63 ; Calberg-Challot, 2007 : 74 ; Vicari, 2018 : 448) et que nous adoptons dans notre travail. Bien que cette conception n'aille pas sans poser certains problèmes, nous pensons qu'elle présente surtout plusieurs avantages.

Un premier avantage concerne le fait que cette conception est plus en adéquation avec la réalité, puisque la langue n'est pas divisée en deux états uniquement, qui correspondraient aux langues de spécialité d'un côté et à la langue générale de l'autre. En effet, suivant les propositions de Pearson (1998 : 35*sqq.*), affinées ensuite par Bowker et Pearson (2002 : 27-28), nous considérons qu'il existe différentes situations de communication, où les besoins de la communication font appel à différents degrés de spécialisation, selon le statut des locuteurs notamment⁹. C'est également ce qu'explique Cabré :

⁹ Nous y reviendrons au Chapitre 4 (4.2.2.1).

« chaque langue de spécialité peut être actualisée à des niveaux différents de spécialisation. Le sommet de la pyramide correspond à la communication entre spécialistes et la base, à la communication de vulgarisation destinée au grand public. » (Cabré, 1998 : 124)

La diversité des productions langagières et de leurs différents niveaux de spécialisation semble donc couverte par la conception d'un continuum entre langue de spécialité et langue générale.

Un autre avantage de cette conception consiste en la possibilité d'élargir la réflexion sur les phénomènes impliquant les langues de spécialité et la langue générale. En effet, nous pensons que l'utilisation des termes *langue de spécialité* et *langue générale* peut amener à conceptualiser deux ensembles uniquement, à mettre en regard deux états de langue uniquement, et donc à contraindre la réflexion en fonction des ensembles théoriques représentés par les langues de spécialité et la langue générale. Or, puisque la réalité ne se classe pas si facilement, la conception du continuum permet de s'émanciper de la vision réductrice imposée par cette dichotomie langue de spécialité/langue générale et d'aborder certaines problématiques sous un angle différent, comme nous proposons de le faire dans notre thèse.

Parallèlement, même si, d'un point de vue théorique, concevoir une continuité entre les langues de spécialité et la langue générale (sans rupture nette entre les deux ensembles, mais plutôt une diminution progressive de la spécialisation) offre une représentation plus proche de la réalité, le problème du flou autour des concepts de langue de spécialité et de langue générale persiste. Autrement dit, ce continuum reste peu « maniable » d'un point de vue pratique ou méthodologique et la distinction langue de spécialité/langue générale reste paradoxalement nécessaire théoriquement et méthodologiquement (Thoiron, 1998 : 621).

En effet, cette distinction, bien qu'imparfaite, offre un cadre théorique permettant de situer les réflexions autour d'un phénomène en particulier dans les champs de la terminologie et des langues de spécialité. Comme le dit Depecker (2002 : Partie 1, §106-108), cette distinction est « opératoire » pour « penser les dynamiques qui traversent la langue et la manière dont y fonctionnent les terminologies » et, plus généralement, pour « établir des repères utiles dans des questions souvent complexes ». Nous rejoignons également Boulanger (1996 : 138), qui considère que « du point de vue scientifique les compartimentations n'ont pour finalité que de faciliter les analyses et de favoriser des classements ».

C'est donc dans une optique opérationnelle que nous faisons le choix, dans ce travail, de parler de *langue générale* pour désigner l'extrémité la moins spécialisée (et donc générale) du continuum et de *langue de spécialité* pour en désigner l'extrémité la plus spécialisée. Ce choix répond

à un double objectif : le premier est essentiellement de cadrer notre réflexion et notre travail autour de la question de la déterminologisation. Plus précisément, ce choix nous permet de conceptualiser les mouvements de termes qui se produisent d'une langue de spécialité à la langue générale, et leur analyse, dans un cadre théorique défini et connu, tout en restant consciente que ces mouvements ne s'effectuent pas d'un ensemble fermé à un autre ensemble fermé, mais se produisent à l'intérieur du continuum formé par les langues de spécialité et la langue générale. Nous détaillons les enjeux et les défis de ce parti pris dans le prochain chapitre.

Notre second objectif est d'adopter un cadre théorique qui soit compatible avec la méthode que nous mettons en place. En effet, comme nous l'expliquons aux Chapitres 3 et 4, nous basons notre analyse principalement sur des données textuelles rassemblées en corpus. Ce choix impose de figer les phénomènes dans une représentation statique, qui est nécessaire dans une approche en corpus, mais qui réduit fortement les possibilités d'analyser un continuum. Distinguer clairement différents degrés de spécialisation entre langue de spécialité et langue générale permet alors de constituer des ensembles cohérents et d'approcher la question du continuum dans les données. Nous détaillons cette réflexion particulièrement au Chapitre 4 (4.2.2.1).

Situer notre travail dans le continuum entre langue de spécialité et langue générale et considérer le passage des termes des langues de spécialité dans la langue générale amènent cependant un second questionnement, à propos de la dichotomie traditionnellement admise en terminologie entre les termes et les mots.

1.1.3. La relativité du terme

1.1.3.1. Entre *terme* et *mot*

L'analyse des termes lorsqu'ils sortent de la sphère spécialisée a des implications directes sur la manière de les envisager, tant d'un point de vue théorique que d'un point de vue méthodologique¹⁰. En réalité, les travaux qui se focalisent sur les mouvements des termes vers la langue générale se sont d'abord intéressés au statut du terme lorsqu'il sort des discours des spécialistes et est en usage dans la langue générale. Une problématique récurrente dans ce cas concerne le passage du terme vers le lexique général et son statut de terme. En effet, différents auteurs se demandent si, lorsqu'il est utilisé dans la langue générale, le terme doit toujours être considéré comme un terme ou s'il doit, au contraire, être considéré comme un mot de la langue générale, voire si son statut s'est « transformé » au cours du processus, faisant ainsi passer une unité

¹⁰ Les aspects méthodologiques seront traités dans le Chapitre 4.

du statut de terme au statut de mot lorsqu'il fonctionne dans la langue générale (Mortureux, 1995 : §4 ; Dury, 2000 : 26 ; Delavigne, 2003 : 85-87 ; Ungureanu, 2006 : 63).

Cette question renvoie en réalité au débat de longue date sur la différence entre un terme et un mot. Néanmoins, dans la mesure où ce débat a été largement repris et discuté dans la littérature¹¹, nous insistons ici uniquement sur les enjeux qui concernent directement le passage de termes et leur fonctionnement dans la langue générale, afin d'exposer la conception du terme que nous adoptons dans notre travail. Nous retenons ainsi des travaux qui abordent ce point depuis la perspective de la circulation des termes qu'un

« terme ne devient pas un mot par-delà une frontière imaginaire : un "piston" par exemple n'est pas tantôt un mot, tantôt un terme – il entre, avec des degrés divers de technicité, dans des discours contraints différemment, émis et reçus selon des finalités différentes. » (Gambier, 1991a : 36)

La position défendue par Gambier l'est également par d'autres chercheurs, notamment dans les approches de la terminologie qui, d'un point de vue tant théorique que méthodologique, se rapprochent de la linguistique. Dans ce cas, le terme est entendu comme un mot qui a un sens spécialisé dans un domaine particulier, qui fonctionne comme toute autre unité de la langue et qui s'insère donc dans des discours variés (Rondeau, 1981 : 21 ; Depecker, 2002 : Partie 1, §15 ; L'Homme, 2004 : Chapitre1, §2 ; Gaudin, 2005 : 82 ; Ungureanu, 2006 : 70 ; Cabré, 2016 : 78). De ce point de vue, si les termes sont considérés comme des termes, c'est « **uniquement** par le fait que leur maîtrise linguistique est associée à la maîtrise d'un domaine de connaissance donné » et non en raison « d'une hypothétique nature linguistique propre » (L'Homme et Polguère, 2007).

Les auteurs insistent ici sur le fait que rien ne distingue véritablement un terme d'un mot, si ce n'est la connaissance extralinguistique qui est projetée sur le terme, dans une situation de communication en particulier. Ce dernier élément renvoie en particulier au constat que « the difficulty in distinguishing between two types of units stems from the fact that formally terms are indistinguishable from words » (Sager 1998: 41, cité dans L'Homme, 2020 : 55), constat qui est largement partagé (Mortureux, 1995 ; Bouveret, 1998 : 11 ; Pearson, 1998 : 8 ; Petit, 2001 : 71 ; Péraldi, 2011 : 131 parmi d'autres). Pour Béjoint et Thoiron d'ailleurs,

« il semble bien qu'il y ait consensus pour dire que le terme n'est pas radicalement différent du mot, même si l'on maintient généralement qu'il y a des différences, qui ne sont d'ailleurs pas nécessairement essentielles, mais peuvent être

¹¹ Voir par exemple à ce propos Dury (1997 : 38-44), Delavigne (2001 : 207-218), Adelstein et Cabré (2002) ou encore Thoiron et Béjoint (2010).

pragmatiques ou même simplement méthodologiques. » (Béjoint et Thoiron, 2000a : 16)

Ainsi, si l'on adhère à cette vision du terme, il semble alors que rien n'empêche un terme de fonctionner comme un mot de la langue générale et vice versa. C'est par ailleurs ce qui soutient les processus de terminologisation et de déterminologisation. Mais cela implique également que la même unité lexicale peut être considérée tour à tour comme un terme ou comme un mot selon le contexte dans lequel elle est utilisée. C'est ce qu'interrogent Béjoint et Thoiron :

« [p]eut-on encore dire qu'une unité lexicale est un terme lorsque le concept auquel elle renvoie est du domaine de la langue générale, qu'il est maîtrisé par tout un chacun dans la communauté linguistique ? Quand un terme cesse-t-il d'être un terme ? Peut-on dire qu'*aspirine*, par exemple, est un terme lorsqu'il est employé dans la langue générale par des locuteurs ignorant tout ou presque de ce qui constitue le concept spécialisé, type de molécule, moyen d'action, indications thérapeutiques, effets secondaires, etc. ? » (Béjoint et Thoiron, 2000a : 11)

Ces questions restent par ailleurs ouvertes et il semble qu'aucune réponse entièrement satisfaisante ne puisse être apportée. Certains auteurs vont même jusqu'à affirmer que l'opposition entre terme et mot n'aurait désormais plus lieu d'être (Alexandru et Gaudin, 2006 : 59).

Cela dit, les problématiques soulevées par la circulation des termes et par l'opposition traditionnelle entre termes et mots s'enrichissent mutuellement. En effet, c'est en partie grâce aux réflexions sur les possibilités pour les termes de circuler entre langue de spécialité et langue générale que la réflexion sur les différences – ou plutôt les ressemblances – entre termes et mots peut se développer. De la même manière, le fait de considérer les termes essentiellement comme des unités lexicales concrétise la possibilité de les envisager comme des entités qui ne restent pas strictement enfermées dans les mêmes types de textes, dans des situations de communication où les interlocuteurs sont des spécialistes du domaine en question. En résulte alors une conception plus nuancée du terme, que nous adoptons sans réserve dans notre travail et qui est formulée de la manière suivante par Dury :

« it seems more appropriate to consider terms as lexical entities which transcends the boundaries of expert language and can also be used by the general public in non-specialized communication. » (Dury, 2005 : 38)

1.1.3.2. Le terme comme objet d'étude dans le continuum entre langue de spécialité et langue générale

Bien que les positions exposées *supra* – et auxquelles nous adhérons – considèrent clairement les termes comme des unités lexicales, ce choix reste peu opérationnel si l'on souhaite

différencier, dans l'analyse des données, les unités sur lesquelles porte notre travail (les termes de physique des particules) et les autres unités. C'est donc essentiellement dans un souci de clarté que nous faisons le choix de continuer à appeler *termes* les unités sur lesquelles portent nos observations, indépendamment du fait qu'elles peuvent se situer à l'intérieur du continuum (c'est-à-dire dans des textes spécialisés comme dans des textes non spécialisés).

Une seconde raison sous-tend toutefois ce choix. Comme nous l'expliquons *infra* (dans la section 1.3), l'usage des termes dans la langue générale est également appréhendé du point de vue de la lexicographie, avec des enjeux tels que le choix des termes à intégrer dans les dictionnaires généraux et la manière de les définir. Si ce dernier point revêt une certaine importance en lexicographie, c'est notamment du fait que le sens des termes qui est véhiculé en langue générale diffère toujours dans une certaine mesure du sens qui est véhiculé en langue de spécialité (cf. sections 1.3, 1.4 et Chapitre 2). Les définitions des termes dans les dictionnaires généraux doivent donc être adaptées. Néanmoins, malgré les différences de sens, il semble que les termes, lorsqu'ils fonctionnent dans la langue générale, expriment toujours dans une certaine mesure le sens spécialisé lié au domaine auquel ils appartiennent. Meyer et Mackintosh (2000a : 203) parlent à ce propos du « concept terminologique sous-jacent » qui reste ainsi visible en langue générale et qui justifierait alors le choix de continuer à parler de *termes* en lexicographie (Meyer et Mackintosh, 2000a : 203 ; Ungureanu, 2006 : 108). En témoignent nombre de travaux (Boulanger et L'Homme, 1991 ; Cottez, 1994 ; Bigras et Simard, 1997 ; Thoiron, 1998 ; Roberts, 2004 ; Josselin-Leray, 2005, Josselin-Leray et Roberts, 2013), où il est bel et bien question de termes.

Ainsi, à l'instar de Delavigne, notamment, nous considérons dans notre analyse que « le terme reste un terme » (Delavigne, 2001 : 230) d'un bout à l'autre du continuum existant entre langue de spécialité et langue générale¹² et nous parlerons par conséquent dans ce travail de *termes* lorsque nous nous référerons à notre objet d'étude, c'est-à-dire les termes de physique des particules (et occasionnellement d'autres domaines – nous y reviendrons dans les Chapitres 6 et 7) et d'*unités lexicales*, ou simplement d'*unités*, lorsque nous nous référerons aux autres unités apparaissant dans les données que nous analysons.

En bref, il est donc admis que les termes circulent dans le continuum entre langue de spécialité et langue générale et qu'ils y fonctionnent comme toute autre unité lexicale de la langue. Il est également reconnu que les termes apparaissent dans une diversité de textes et de discours différents. Le déplacement des préoccupations des chercheurs du débat terme/mot vers le

¹² Ce qui n'est pas en contradiction avec la définition du terme comme une unité lexicale.

fonctionnement des termes dans une pluralité de situations différentes permet le développement d'une réflexion autour du rôle des termes dans la transmission des connaissances spécialisées auprès du grand public. Dans la section suivante, nous dressons ainsi un portrait des différentes approches visant à caractériser le fonctionnement des termes lorsqu'ils servent à transmettre des connaissances, dans des situations de communication différentes (vulgarisation scientifique, presse généraliste, textes didactiques, etc.) et à partir de multiples points de vue (construction et négociation du sens, stratégies de reformulation, réception des termes par les non-spécialistes, etc.). Une revue de ces différents enjeux permettra alors de situer de manière plus détaillée notre thèse parmi les nombreux travaux portant sur des questions relatives au mouvement des termes d'une langue de spécialité vers la langue générale.

1.2. Les termes du point de vue de la transmission des connaissances spécialisées vers le grand public

Comme le dit Gambier (1991a : 36), lorsque le terme circule en dehors des discours de spécialistes, il « entre, avec des degrés divers de technicité, dans des discours contraints différemment, émis et reçus selon des finalités différentes ». L'enjeu des travaux qui s'intéressent à ces questions consiste alors à décrire les fonctionnements des termes dans ces différents types de discours et en fonction des objectifs poursuivis. Parmi ces objectifs se trouve la transmission des connaissances spécialisées vers le grand public. Il est alors admis que les termes sont utilisés dans des textes dont le but est d'expliquer des connaissances spécialisées auprès de locuteurs non experts. Ces textes peuvent être classés selon qu'ils entrent sous la dénomination *vulgarisation scientifique* ou non. Dans cette section, deux points de vue complémentaires sur ces questions nous permettent d'aborder la problématique du terme lorsqu'il s'intègre dans des discours visant à transmettre des connaissances spécialisées.

1.2.1. Perspective en vulgarisation scientifique

Le champ de la vulgarisation scientifique est vaste et peut être appréhendé à partir de multiples disciplines, par exemple en sociologie, psychosociologie, linguistique, philosophie, didactique, sémiotique, selon Jacobi (1984 : 484), en terminologie et socioterminologie (Delavigne, 2001 ; Gaudin, 2005). Nous retenons deux points de vue, pertinents pour notre travail, celui de la socioterminologie et celui de la linguistique, à travers des travaux se revendiquant d'une approche en analyse du discours. Ces points de vue sont complémentaires, car les objectifs poursuivis se rapprochent sensiblement et concernent de manière générale la transmission et la construction du

sens dans des textes destinés à des non-experts. Les approches diffèrent néanmoins dans une certaine mesure.

En effet, par la question de la vulgarisation scientifique, les travaux en socioterminologie interrogent particulièrement « [t]he role that terminology plays in permitting non specialists to gain some insights into specialised knowledge » (Humbley, 2018b : 481). Pour Gaudin, la diffusion des termes en dehors des cercles de spécialistes amène nécessairement à appréhender des problématiques liées à la vulgarisation scientifique : « l'étude de la **circulation sociale** des termes implique également des pratiques langagières telles que celles que l'on désigne du nom de *vulgarisation* » (Gaudin, 2005 : 90-91)¹³. La dimension sociale de la vulgarisation est en réalité inhérente à sa fonction de transmission des connaissances (Mortureux, 1985 ; Vicari, 2018 ; Delavigne, 2020).

Ainsi, dans une perspective socioterminologique, Delavigne considère que le terme représente le point d'entrée dans ces problématiques et dans l'analyse des textes de vulgarisation scientifique. L'autrice cherche alors à

« repérer les termes et à examiner la façon dont ces termes *attestés* sont utilisés, à déterminer pourquoi ils sont utilisés et les moyens qu'offrent les discours aux récepteurs de se les approprier. » (Delavigne, 2001 : 229)

En revanche, les travaux en analyse de discours¹⁴ s'intéressent particulièrement aux stratégies discursives de reformulation qui participent à transmettre et à construire le sens dans les discours de vulgarisation. Celles-ci sont parfois divisées entre des stratégies de reformulation lexicale, qui concernent les différents termes et syntagmes utilisés pour désigner le même concept, et des stratégies de ré-énonciation, plus proches de la paraphrase (Moirand, 2014). Ces travaux interrogent également la « démultiplication des intervenants convoqués dans l'exposition de la science » (Reboul-Touré, 2004 : 195), du fait des facilités offertes désormais par Internet. Cependant, comme nous l'expliquons dans cette section, il existe plusieurs points de rencontre entre ces approches.

Dans ces travaux, il est généralement admis que la vulgarisation scientifique s'inscrit dans une perspective non seulement de diffusion des connaissances, mais aussi de leur appropriation par des locuteurs non experts¹⁵ (Jacobi, 1986 : 71 ; Mortureux, 1988 : 118-120 ; Delavigne,

¹³ Les termes indiqués en gras et en italique le sont dans la citation originale.

¹⁴ Et particulièrement les travaux du CEDISCOR (Cercle de recherche sur les discours), <https://journals.openedition.org/cediscor/1554> (page consultée le 3 février 2021).

¹⁵ En réalité, comme le souligne Jacobi (1984 : 484), « [i]l serait imprudent de définir la vulgarisation de manière univoque » tant les pratiques sont diversifiées. Nous nous permettons toutefois de simplifier

2001 : 81 ; Moirand *et al.*, 2016 : 146). Il est par ailleurs également reconnu que c'est par le biais d'une importante « reformulation du discours scientifique » (Jacobi, 1986 : 15) que ces objectifs de diffusion et d'appropriation sont achevés (Loffler-Laurian, 1984 ; Jacobi, 1988 ; Gaudin, 2003 : 107*sqq.* ; Reboul-Touré, 2004). Comme le dit par exemple Delavigne, « [a]fin de rétablir une intercompréhension, les discours destinés aux "non-experts" s'attachent à redonner du sens aux mots » (Delavigne, 2020 : §2).

Plusieurs stratégies de reformulation des idées exprimées et des termes utilisés sont alors mises en œuvre dans ce but dans les textes de vulgarisation scientifique. Par ailleurs, une importante part des travaux qui portent sur la vulgarisation s'attache à caractériser et à analyser ces stratégies (Jacobi, 1984, 1988 ; Mortureux, 1988 ; Ciapuscio, 1997, 2003 ; Delavigne, 2001 : 515*sqq.*, 2020 ; Reboul-Touré, 2004 ; Moirand *et al.*, 2016 par exemple), qui sont désormais bien décrites. En outre, Mortureux propose d'aborder le repérage et l'analyse de ces reformulations à travers ce qu'elle appelle les *paradigmes désignationnels* et les *paradigmes définitionnels* (Mortureux, 1993) (qui sont notamment repris par Moirand (1997, 2007), Delavigne (2001, 2003, 2020) ou encore Reboul-Touré (2004)). Ces deux outils d'analyse regroupent respectivement les « syntagmes à valeur désignative » et les « périphrases définitionnelles » qui « fonctionn[ent] en coréférence avec un vocable initial dans un discours donné » (Mortureux, 1993 : §4). Dans le premier cas, il peut s'agir de la mise en relation d'un terme et de son synonyme, de leur simple juxtaposition, de l'utilisation de certains signes typographiques, par exemple. À ce propos Delavigne précise que « [l]es discours de vulgarisation associent termes scientifiques et termes supposés appartenir au vocabulaire du destinataire afin de construire du sens » (Delavigne, 2001 : 229). Dans le second cas, il s'agit le plus souvent de paraphrases visant à définir, du moins en partie, les concepts dont il est question.

C'est donc à travers cette importante activité métalinguistique qui caractérise les textes de vulgarisation que s'illustre la volonté de la vulgarisation scientifique de faire passer des connaissances spécialisées auprès du grand public. C'est par les différentes reformulations que le sens se construit et que les locuteurs non spécialistes s'approprient les notions. Il ne nous paraît toutefois pas utile de développer davantage ces questions, car notre travail ne porte pas spécifiquement sur ces stratégies de reformulation¹⁶. En revanche, nous retenons de ces travaux, et notamment de l'abondante réflexion sur le rôle de ces stratégies dans la production du sens et la transmission des connaissances spécialisées, que les textes de vulgarisation font largement usage du

quelque peu, pour les besoins de notre travail. Pour une présentation détaillée de ce champ disciplinaire, voir en particulier les Chapitres 1 et 2 de la thèse de V. Delavigne (2001 : 13-87).

¹⁶ En effet, comme nous l'expliquons dans le Chapitre 5, notre méthodologie repose essentiellement sur une analyse de la distribution des termes en corpus et non sur les aspects métalinguistiques de la reformulation.

métalinguistique et sont caractérisés par d'importantes marques de didacticité (Moirand, 1993 ; Beacco et Moirand, 1995 : 39*sqq.* ; Delavigne, 2001 : 70*sqq.*, 2020). Par ailleurs, comme le dit Delavigne (2020 : §65), « [l]a catégorie du métalinguistique est un trait quasi-définitoire des discours de vulgarisation ». Nous montrons l'utilité de cette caractéristique dans le Chapitre 4.

Nous retenons également que, puisque les termes servent à faire passer des connaissances dans les textes de vulgarisation, alors ces textes « favorise[nt] la circulation des termes » (Alexandru et Gaudin, 2006 : 61) et peuvent être considérés comme des canaux privilégiés par lesquels les termes atteignent le grand public. Mais d'autres canaux sont parfois évoqués par certains des auteurs que nous citons ici, et la vulgarisation scientifique n'est pas le seul moyen de transmettre des connaissances spécialisées auprès du grand public.

1.2.2. Discours de transmission des connaissances et rôle des médias généralistes

Certains des auteurs qui s'intéressent à la vulgarisation scientifique se focalisent également sur la dynamique de la transmission des connaissances spécialisées par le biais de textes plus diversifiés, qui ne relèvent pas explicitement de la vulgarisation. Beacco et Moirand (1995) proposent d'appréhender ces textes et cette dynamique sous l'angle de ce qu'ils appellent des « discours de transmission des connaissances ». D'après les deux auteurs, ce concept renvoie à l'idée qu'une multitude de textes participe à la diffusion des connaissances spécialisées et, partant, au transfert des termes vers la langue générale. Certains de ces textes sont qualifiés de discours ordinaires, « parce qu'ils n'ont pas pour vocation première de transmettre des connaissances » (*Ibid.* : 32-33). Aussi, tant « les discours qui seraient strictement informatifs et qui ne manifesteraient aucune intention de rendre l'autre plus compétent » (*Ibid.* : 39) que les discours de vulgarisation ou d'enseignement entreraient dans ce paradigme.

Les discours de transmission des connaissances sont également qualifiés de « discours seconds » (Moirand, 1999 : §1) par opposition avec « les discours de production de connaissances (*discours premiers* ou *discours sources*) » (*Ibid.*) (ou « discours scientifiques sources » (Beacco et Moirand, 1995 : 33)). Ce concept est né du constat qu'un modèle « classique » de la transmission des connaissances n'est en réalité pas apte à représenter les différentes manières dont les connaissances sont effectivement diffusées entre les spécialistes et les non-spécialistes. En effet, ce modèle classique repose sur l'hypothèse que les connaissances sont disséminées linéairement à partir des discours sources jusqu'au grand public, par l'intermédiaire d'abord de discours didactiques puis des médias (Beacco *et al.*, 2002 : 278). Cependant, les communautés scientifiques entretiennent des rapports de plus en plus complexes avec le grand public et la science « is no longer 'enclosed' within

identifiable discursive forms but circulates within many ordinary discourses, whether in the media or not » (*Ibid.* : 279). Afin de rendre compte de la multiplicité de ces discours ordinaires, il est nécessaire d'adapter ce modèle classique, en particulier son caractère linéaire. Selon les auteurs, la diversité des discours de transmission des connaissances et la complexité des mécanismes et des dynamiques en jeu imposent de repenser les représentations de la transmission des connaissances auprès du grand public. Il faut donc s'éloigner d'une vision linéaire et prendre en considération le caractère immédiat et pluriel, voire simultané, de la transmission des connaissances spécialisées¹⁷. À ce propos, et pour rendre compte de cette complexité, les auteurs parlent de « multiform arterial dissemination of science » (*Ibid.* : 280).

Les médias occupent une place privilégiée dans cette représentation et les auteurs insistent même sur le fait que la science¹⁸ « now tends to saturate informative media discourses » (*Ibid.*). De ce point de vue, nous pouvons considérer que les médias, et par conséquent la presse, contribuent largement au phénomène global de transmission des connaissances vers le grand public, et donc également, par extension, à la circulation des termes en dehors des textes des spécialistes (Moirand, 1997, 1999, 2007 ; Beacco, 1999, 2000 ; Beacco *et al.*, 2002).

Parallèlement, d'autres travaux envisagent le rôle de la presse dans la circulation des termes, d'un point de vue toutefois différent de celui de l'analyse de discours. En réalité, l'idée que les médias généralistes participent à diffuser des termes (et des connaissances) vers les non-experts est largement répandue (Meyer, 2000 ; Carras, 2002 ; Condamines et Picton, 2014b ; Cañete *et al.*, 2016 ; Estopà, 2016 ; Ledouble, 2019 parmi d'autres). Ainsi, Cabré insiste sur le fait que le mouvement des termes vers la langue générale est rendu possible notamment par le développement des médias de masse :

« [l]a grande extension des moyens de communication de masse et la démocratisation de l'enseignement ont favorisé cette divulgation des matières spécialisées, et, par conséquent, la diffusion de la terminologie la plus employée. » (Cabré, 1994 : 593)

De la même manière, Gaudin souligne que les médias s'intéressent particulièrement à la recherche scientifique, qui est délibérément mise en avant :

« [c]'est qu'il y a dans le domaine des discours fortement médiatisés une exploitation quasi systématique du succès de domaines de recherche. Il en découle alors une forte diffusion de mots presque vides de sens, pour lesquels le

¹⁷ Soulignons que des supports non écrits tels que des images ou des vidéos peuvent également entrer en jeu.

¹⁸ Les auteurs entendent ici la science dans un sens général. Mais il est reconnu que certains sujets scientifiques sont privilégiés par rapport à d'autres, notamment lorsque ces sujets impliquent des enjeux politiques ou sociaux (Moirand, 2014).

public construit des concepts, nécessairement flous, coupés de la pratique d'où ils tirent, non seulement leur légitimité, mais leur utilité même. » (Gaudin, 1993 : 131)

Gaudin suggère ici que l'apparition de termes dans les médias, où ils sont probablement repris de discours scientifiques par des journalistes, contribue à leur passage dans la langue générale et, par extension, à leur appropriation par les locuteurs non spécialistes.

Pearson, de son côté, évoque des idées proches de celles de Gaudin :

« many terms describing the symptoms and treatment of AIDS have seeped into general language as a result of widespread media coverage of the subject and people use AIDS related terms in ordinary language as if they really understood the precise meaning of these terms. » (Pearson, 1998 : 27)

Ces deux dernières citations mettent par ailleurs en évidence un phénomène parallèle à la circulation des termes, celui des changements sémantiques qui peuvent se produire lorsque les termes sortent de leur langue de spécialité « d'origine » et fonctionnent dans la langue générale. Ainsi, avec la question de la circulation des termes dans les médias se pose inévitablement la question de leur sens. De la même manière qu'une des préoccupations des auteurs travaillant sur la vulgarisation scientifique consiste à observer et décrire les mécanismes de construction, de négociation et de validation sociale du sens (Gaudin, 1997, 2003, 2005 ; Delavigne, 2013, 2020), les auteurs s'intéressant au fonctionnement des termes dans la presse constatent certains glissements sémantiques.

Par exemple, Le Draoulec *et al.* (2014) observent l'émergence et la diffusion d'usages dépréciatifs du terme *sémantique* dans un corpus journalistique et dans un corpus de textes issus du Web, qui coexistent et contrastent fortement avec l'usage qu'en font les linguistes. De manière similaire, Rouillard s'intéresse à « l'appropriation des termes relatifs à la psychiatrie par les non-spécialistes » (Rouillard, 2016 : 26) et analyse en particulier les usages qu'elle qualifie de « spécialisés » et de « courants » des termes *hystérie* et *autisme* dans un corpus de presse. Notons que, dans ces deux cas, les observations des autrices mettent essentiellement en lumière l'apparition de connotations négatives. Leurs observations sont cependant à rapprocher des recherches de Moirand, où le rôle des médias dans ces changements sémantiques est mis en avant. En effet, à propos d'« événements scientifiques et techniques » (Moirand, 2007 : 19) ayant eu une certaine ampleur dans la presse, l'autrice explique que

« le traitement de ces événements par les médias transforme les notions diffusées en les banalisant à leur manière, et ces notions continuent leur route avec des colorations sémantiques nouvelles dans les différentes communautés concernées par ces événements avant de revenir dans le circuit de la communication

médiatique. [...] Les médias ordinaires vont jouer un rôle incontestable dans la diffusion et la banalisation de ces nominations. » (*Ibid.* : 20)

Remarquons que les propos de Moirand se rapprochent de ceux de Gaudin pour ce qui est des glissements sémantiques des termes. Des constats similaires se retrouvent d'ailleurs dans les travaux de Varantola (1986), Guilbert (1973a, 1975), Mortureux (1995), Delavigne et Bouveret (2000a), Ungureanu (2006) ou encore Botta (2013), par exemple. Il semblerait même que les mouvements des termes vers la langue générale ne puissent pas s'envisager sans que le sens de ces termes soit altéré dans une certaine mesure au cours du processus¹⁹.

Par ailleurs, dès lors que les termes ont intégré la langue générale, avec diverses modifications au niveau de leur sens, la question de leur description dans des dictionnaires de langue générale se pose inévitablement. Ainsi, c'est à travers le point de vue de la lexicographie et de la métaléxicographie que nous poursuivons notre réflexion à propos de la circulation des termes et de leur fonctionnement dans la langue générale.

1.3. Les termes dans les dictionnaires généraux

Comme le dit Josselin-Leray,

« les termes ne sont pas cantonnés aux seuls discours spécialisés (ou communication entre experts) : ils circulent [...]. Et c'est justement parce que les termes circulent qu'ils se retrouvent dans les dictionnaires dits de langue générale. » (Josselin-Leray, 2010 : 65)

Les questions liées à l'inclusion et à la description des termes dans les dictionnaires généraux sont nombreuses (Béjoint, 1988 : 355 ; Bigras et Simard, 1997 : 97), mais toutes ne sont pas forcément en rapport avec notre objet d'étude²⁰. D'autres ont déjà été discutées *supra*, du point de vue de la circulation des termes (par exemple, le traitement des termes dans des dictionnaires généraux, qui décrivent traditionnellement les unités de la langue générale (Roberts, 2004 : 121), interpelle nécessairement la notion de langue générale, la délimitation entre langue générale et langue de spécialité ainsi que la distinction habituelle entre mot et terme – sur lesquelles nous ne revenons pas dans cette section).

La nécessité de décrire certains termes dans les dictionnaires généraux semble indéniable (Boulanger et L'Homme, 1991 : 26 ; Josselin-Leray, 2005 : 88-118 ; Alonso Campo, 2008 : 929),

¹⁹ Nous reviendrons sur ces questions *infra* (dans la section 1.4 et au Chapitre 2).

²⁰ Pour une présentation détaillée, voir notamment le Chapitre 2 de la thèse d'A. Josselin-Leray (2005 : 71-170)

non seulement du fait que les termes passent effectivement des langues de spécialité à la langue générale, comme l'illustre la citation de Josselin-Leray, mais également du fait que leur description dans les dictionnaires généraux répond aux besoins et aux attentes des utilisateurs de ces dictionnaires (Boulanger et L'Homme, 1991 : 28-29 ; Bigras et Simard, 1997 : 97 ; Thoiron, 1998 : 626 ; Roberts et Josselin-Leray, 2005 : 328 ; Bowker, 2017 : 145). En outre, et en dépit du fait que « [l]'entrée de termes dans un dictionnaire général est une question qui a été longuement et diversement débattue » (Cabré, 1994 : 589), il semble qu'une grande diversité de traitement des termes dans les dictionnaires généraux perdure (Depecker, 1994 : 152 ; Alonso Campo, 2008 : 929). Les choix divergent essentiellement au niveau de la sélection des termes et de la manière de les décrire (Cabré, 1994 : 595 ; Alonso Campo, 2008 : 929 ; Rossi, 2014 : 48).

Nombre de travaux visent ainsi à analyser le traitement des termes dans les dictionnaires généraux et à en présenter les enjeux, que ce soit pour améliorer la représentation des différents domaines de spécialité dans les dictionnaires ou pour améliorer les descriptions des termes dans les dictionnaires généraux. Ces questionnements sont abordés du point de vue de la macrostructure (en interrogeant le plus souvent les critères de sélection des termes à inclure) ou de la microstructure des dictionnaires (en se focalisant essentiellement sur les définitions des termes). Par ailleurs, l'analyse se concentre généralement sur les termes d'un domaine, par exemple la volcanologie (Josselin-Leray, 2005, 2010), l'œnologie (Roberts, 2004 ; Roberts et Josselin-Leray, 2005), l'informatique (L'Homme et Polguère, 2007 ; Josselin-Leray et Roberts, 2013 ; Ortego Antón et Fernández Nistal, 2015a, 2015b), l'environnement (Alonso Campo, 2008) ou encore l'alpinisme (Tetet, 1994), ou sur un échantillon de termes correspondant à une certaine caractéristique, par exemple les termes qui figurent à l'intérieur d'une sélection alphabétique précise (Mazière, 1981 ; Bigras et Simard, 1997 ; Thoiron, 1998), les termes officiels (Depecker, 1994), les termes métaphoriques (Rossi, 2014) ou encore les termes en lien avec l'actualité (Tolédano et Candel (2006) sur les termes concurrents *mondialisation* et *globalisation*).

1.3.1. Sélection des termes à décrire dans les dictionnaires

La littérature fait état d'un certain nombre de critères mis en œuvre pour sélectionner les termes à décrire dans les dictionnaires. Ces critères sont nécessaires, car il n'est bien évidemment pas envisageable d'inclure tous les termes qui se retrouvent utilisés dans la langue générale. En effet, la nomenclature des dictionnaires généraux étant nécessairement limitée, seule une faible proportion de termes est susceptible d'y être intégrée. Une manière de les sélectionner consiste par exemple à tenir compte de leur degré de spécialisation, les termes avec un degré de spécialisation moindre étant privilégiés (Thoiron, 1998 : 622 ; Roberts, 2004 : 127). Ce choix peut s'expliquer par

le fait que ces termes sont les plus susceptibles d'apparaître dans la vie quotidienne des locuteurs (Roberts et Josselin-Leray, 2005 : 337). Une autre possibilité consiste à privilégier les termes simples par rapport aux termes complexes, essentiellement du fait que les dictionnaires généraux recensent le plus souvent des unités simples (Béjoint, 1988 : 362 ; Depecker, 1994 : 152 ; Thoiron, 1998 : 642 ; Roberts et Josselin-Leray, 2005 : 337).

Cependant, le critère qui semble déterminant est celui que Josselin-Leray (2005 : 149-155) appelle le critère de la *banalisation lexicale*, c'est-à-dire

« le fait qu'un terme passe de la langue spécialisée à la langue non spécialisée (à la langue courante), phénomène dû à sa diffusion par divers interlocuteurs (dont les principaux représentants sont les médias). » (Josselin-Leray et Roberts, 2013)

Les autrices précisent par ailleurs qu'elles n'utilisent « pas ce terme exactement dans le même sens que Galisson (1978), à qui en revient toutefois la paternité » (Josselin-Leray et Roberts, 2013)²¹.

Différents auteurs (notamment Boulanger et L'Homme, 1991 : 28-29 ; Cabré, 1994 : 595 ; Roberts, 2004 : 123 ; Ortego Antón, 2015 : 129) évoquent également ce critère, ou du moins des idées qui en sont proches. Il ressort ainsi que, parmi l'ensemble des termes d'une langue, il n'est utile d'inclure dans les dictionnaires que ceux que les utilisateurs de dictionnaires sont susceptibles de rencontrer dans leur vie de tous les jours ; autrement dit, les termes qui passent dans la langue générale. Boulanger, par exemple, affirme que

« des vocabulaires établis ou récents réservés au départ à l'apanage des experts sortent rapidement du cercle étroit de la spécialisation pour s'enraciner, en tout ou en partie, dans la langue quotidienne. [...] Les lexicographes doivent absolument récupérer ces vocables et les traiter dans les dictionnaires généraux monolingues (DGM) ou dictionnaires de langue (DL), car ils relèvent de la cohorte de connaissances et d'échanges spécialisés qui rejoignent maintenant le grand public. » (Boulanger, 1996 : 139)

De la même manière, Depecker (1994 : 152) évoque « la frange des vocabulaires techniques ou scientifiques qui touche le plus directement le grand public ». Les termes qui répondent à ce critère restent cependant difficilement identifiables, ce qui est vraisemblablement imputable à la nature du critère et à la difficulté, voire à l'impossibilité, de déterminer clairement à partir de quel moment un terme est passé dans la langue générale²². Ainsi, paradoxalement, il semble que ce critère

²¹ Nous traiterons de la banalisation lexicale au sens de Galisson dans le Chapitre 2.

²² Cela supposerait en effet qu'une délimitation précise de ce que constitue la langue générale soit possible. Or, comme nous l'avons vu *supra* (1.1.2), la question est plus complexe.

figure parmi les plus importants pour ce qui est de l'inclusion des termes dans les dictionnaires généraux mais qu'il reste peu opérationnel dans la pratique.

Parallèlement, nombre d'auteurs qui évoquent le lien entre la diffusion de termes et leur inclusion dans les dictionnaires suggèrent que les médias jouent un rôle important dans cette diffusion auprès du grand public (comme nous le soulignons en 1.2.2). De ce point de vue, la présence de termes dans les médias peut être vue comme un indicateur de l'intérêt supposé que ces termes suscitent chez le grand public et donc comme un moyen d'évaluer la pertinence de l'inclusion de tel ou tel terme dans un dictionnaire général (Boulangier et L'Homme, 1991 : 29 ; Rossi, 2014 : 47-48 ;). C'est dans ce but que Roberts et Josselin-Leray, par exemple, proposent d'exploiter un corpus de presse (Roberts et Josselin-Leray, 2005 : 337). Dans le cadre d'une analyse critique du traitement des termes du domaine de l'œnologie dans des dictionnaires anglais et français, ces autrices proposent en effet d'opérationnaliser l'analyse de la mise en œuvre dans les dictionnaires du critère de la banalisation lexicale en contrastant leurs observations à propos de la présence des termes à l'étude dans les dictionnaires avec leur fréquence dans un corpus de presse comprenant 129 millions d'occurrences (*Ibid.*). La fréquence des termes dans le corpus, ou même simplement leur présence, permet alors non seulement d'estimer la diffusion de ces termes auprès du grand public, mais également de confronter le critère de la banalisation lexicale à des données réelles.

Dans une étude ultérieure, les autrices exploitent par ailleurs cette méthode pour montrer que, sur l'échantillon de 31 termes considérés dans leur analyse, 10 ne satisfont pas au critère de banalisation lexicale. Selon elles, il semble même « jouer un rôle moindre dans les choix des termes faits par les rédacteurs du *NODE* [New Oxford Dictionary of English] », sans pour autant être entièrement négligé (Josselin-Leray et Roberts, 2013).

Le critère de la banalisation lexicale est ainsi étroitement lié à la circulation des termes, notamment lorsque les médias sont impliqués. De plus, comme nous le disons *supra*, lorsque les termes intègrent la langue générale, il est probable qu'ils soient utilisés dans un sens légèrement différent du sens véhiculé par des spécialistes. Dans certains cas, le terme peut même être employé métaphoriquement.

C'est précisément la possibilité d'une utilisation métaphorique des termes dans la langue générale qui donne lieu à un autre critère de sélection des termes à inclure dans les dictionnaires généraux, que Josselin-Leray nomme *métaphorisation*, à partir notamment des propositions de Cottez (1994). En effet, Cottez préconise d'inclure un terme dès qu'il « commence à se prêter à un emploi métaphorique ou figuré » (Cottez, 1994 : 20).

Ainsi, les termes qui deviennent autonomes dans la langue générale, pour reprendre l'idée de Cottez, seraient à sélectionner pour inclusion dans les dictionnaires. Cette autonomie, acquise grâce à un usage métaphorique ou figuré dont nous supposons qu'il est suffisamment stable dans la langue pour être relevé, renvoie par ailleurs aux questionnements que soulève la néologie sémantique (cf. 1.4).

Parallèlement, pour Cottez, le phénomène inverse justifie également une description dans les dictionnaires généraux. Plus précisément, lorsque des unités de la langue générale deviennent des termes et acquièrent un sens particulier dans un domaine, le plus souvent par métaphore, ces nouveaux sens doivent être décrits dans les dictionnaires²³. L'auteur l'explique de la manière suivante :

« [l']utilisation, par les scientifiques, d'une unité du vocabulaire général à laquelle ils donnent un sens spécial, est à enregistrer sans hésitation : c'est le signe d'un pont jeté entre locuteurs de masse et producteurs de textes ésotériques, l'amorce d'un va-et-vient des uns aux autres, d'emprunts mutuels. » (*Ibid.* : 22)

Remarquons cependant que, malgré la nécessité de décrire les sens spécialisés d'unités existant indépendamment dans la langue générale, ceux-ci ne sont que rarement décrits (Josselin-Leray, 2010 : 86). Rossi parvient à une conclusion similaire à partir de l'analyse du traitement de plusieurs termes métaphoriques dans des dictionnaires généraux en français et en italien. L'autrice déplore « un traitement parfois incohérent » de ces termes, qui ont un « statut incertain, qui ne tient compte que de leur nature de sens supplémentaire de mots qui existent déjà dans la langue commune » (Rossi, 2014 : 60). Selon elle, il serait nécessaire d'explicitier davantage le caractère terminologique des nouveaux sens de ces unités existantes (*Ibid.*).

Les constats plutôt négatifs de Rossi mettent en évidence un problème soulevé de manière récurrente dans les travaux portant sur le traitement des termes dans les dictionnaires généraux. Il s'agit du problème de leur définition.

1.3.2. La définition des termes dans les dictionnaires, un défi pour la lexicographie

Les questions liées à la définition des termes dans les dictionnaires généraux sont multiples (Josselin-Leray, 2010 : 92). Par exemple, Rossi remarque un manque de précision dans les définitions de termes métaphoriques (Rossi, 2014 : 60). Ce manque de précision n'est cependant

²³ Notons que ces deux cas décrits par Cottez renvoient aux processus de déterminologisation et de terminologisation tels qu'ils sont nommés du point de vue de la circulation des termes (cf. section 1.1.1).

pas une particularité des termes métaphoriques et se retrouve potentiellement pour tous les termes intégrés dans les dictionnaires généraux (Roberts et Josselin-Leray, 2005 : 341).

Pourtant, de manière générale, il semble admis que la définition d'un terme doit consister en un équilibre entre précision et compréhensibilité, ou, dans les mots de Béjoint, « a balance between scientific precision and general accessibility » (Béjoint, 1988 : 366). De la même manière, selon Josselin-Leray,

« [t]out dépend en réalité du but que s'assigne le dictionnaire : véhiculer une image scientifique la plus juste qu'il soit, ou se mettre à la portée de l'utilisateur. L'idéal serait un juste milieu entre les deux. » (Josselin-Leray, 2010 : 93)

Mais la réalité est autre et ce « juste milieu » semble difficile à atteindre. Les auteurs concluent ainsi le plus souvent sur un manque de satisfaction des définitions (Béjoint, 1988 : 363 ; Depecker, 1994 : 151 ; Roberts et Josselin-Leray, 2005 : 341 ; Rossi, 2014 : 48).

Plusieurs raisons peuvent l'expliquer, parmi lesquelles deux en particulier nous semblent pertinentes pour notre propos²⁴. La première concerne la précision scientifique, évoquée dans les deux citations ci-dessus. En effet, dans la mesure où les utilisateurs de dictionnaires généraux ne sont pas spécialistes des domaines auxquels appartiennent les termes qui y sont décrits, les définitions de ces termes doivent nécessairement être adaptées aux connaissances (scientifiques et linguistiques) supposées de ces utilisateurs. De la même manière, tous les termes d'un domaine ne sont pas susceptibles d'intéresser les utilisateurs. Aussi, afin de définir un terme qui répond aux critères d'inclusion d'un dictionnaire en particulier, il n'est pas envisageable de définir tous les termes nécessaires à la définition de ce terme, ni même de les inclure, comme le remarque Thoiron (1998 : 632) à propos de l'absence des hyperonymes de certains termes. C'est là que les démarches en lexicographie diffèrent des démarches mises en œuvre en terminologie, où les termes sont définis en fonction des relations qu'ils entretiennent avec les autres termes du domaine, notamment la relation d'hyponymie. La simplification des définitions devient alors nécessaire (Béjoint, 1988 : 363 ; Cabré, 1994 : 596 ; Rossi, 2014 : 48).

La seconde raison a trait aux différents changements sémantiques qui se produisent potentiellement lorsqu'un terme passe dans la langue générale. Ceux-ci ne mènent pas systématiquement à l'émergence de sens métaphoriques stables, comme le laisse entendre Cottez (1994), que nous citons *supra*. Ainsi, le manque de stabilité de ces nouveaux sens et la difficulté à cerner précisément les changements compliquent nécessairement la tâche de description des termes

²⁴ Pour une analyse détaillée des définitions des termes dans les dictionnaires, voir par exemple Josselin-Leray (2005 : 325-336).

lors de leur inclusion dans un dictionnaire général. C'est pourquoi ces nouveaux sens sont parfois tout simplement ignorés (Gilbert, 1973 : 41-42).

Par ailleurs, Béjoint nuance la critique de l'imprécision des définitions et souligne les difficultés inhérentes à la définition des termes dans les dictionnaires généraux :

« **the meaning really is fuzzy**: when scientific and technical words have become fairly common in non-specialized use (i.e. when they have moved down our scale), **they tend to be used with slightly different meanings all the time**, and they become as difficult to define as other types of words. » (Béjoint, 1988 : 362)²⁵

Béjoint soulève ici un point important, qui concerne les décalages possibles entre les usages des termes dans la langue générale et les descriptions de ces termes dans les dictionnaires généraux. Selon l'auteur, le sens des termes dans la langue générale peut parfois être flou, instable. Cette instabilité pose évidemment un défi important pour la lexicographie et représente une difficulté pour atteindre l'équilibre entre précision scientifique des définitions et compréhensibilité par les utilisateurs non experts des dictionnaires généraux.

Les changements de sens des termes lorsqu'ils sont utilisés dans la langue générale sont caractéristiques de la circulation des termes en dehors de la sphère spécialisée, que ces changements soient plutôt subtils comme le laisse entendre Béjoint ou qu'ils soient plus conséquents, comme ce qu'évoque le critère de la métaphorisation décrit en 1.3.1, et représentent plusieurs défis pour la lexicographie. Cependant, ces différents changements sémantiques peuvent également être abordés du point de vue de la néologie en langue générale. Nous consacrons donc la dernière section de ce chapitre aux liens entre la circulation des termes et la néologie en langue générale.

1.4. Perception néologique des termes en langue générale

Dans les sections précédentes, nous évoquons brièvement les conséquences de la circulation des termes en dehors de la sphère spécialisée sur le sens de ces termes, avec des modifications plus ou moins importantes selon les auteurs cités (notamment Guilbert, 1975 ; Béjoint, 1988 ; Gaudin, 1993 ; Pearson, 1998 ; Meyer et Mackintosh, 2000a ; Ungureanu, 2006 ; Moirand, 2007 ; Delavigne, 2013). Ces questions touchent plus largement au champ de la néologie. Ainsi, lorsque ces changements sémantiques s'établissent de manière stable et durable, ils peuvent être abordés du point de vue de la néologie sémantique. Plusieurs enjeux sont largement traités dans la littérature, notamment la question du repérage de ces néologismes sémantiques, de

²⁵ Nous soulignons.

leur caractérisation et de leur description dans les ressources lexicographiques. Mais le fonctionnement de termes dans la langue générale interroge également d'autres aspects de la néologie. Nous les détaillons dans cette section²⁶.

1.4.1. Intégration des termes comme « nouveaux mots » dans la langue générale

Le premier point de vue que l'on peut porter sur l'intégration de termes dans la langue générale consiste à les considérer comme de véritables néologismes, où ce n'est pas uniquement la nouveauté sémantique qui est mise en avant. Il faut comprendre ici qu'au moment où certains termes commencent à être utilisés dans la langue générale, dans des contextes non spécialisés, ils peuvent susciter un sentiment de nouveauté (on pourrait parler de « sentiment néologique » (Gardin *et al.*, 1974)). L'apparition d'un terme dans la langue générale (terme qui était jusqu'alors réservé à une langue de spécialité en particulier) correspond alors en tout point à la définition de *néologisme* que donne par exemple Sablayrolles (2012 : 38). Dans ce cas, en effet, le terme peut se « défini[r] comme un nouveau signe avec apparition conjointe d'un nouveau signifiant et d'un nouveau signifié ». Cette affirmation suppose alors qu'aussi bien la forme du terme que son sens étaient inconnus en langue générale jusqu'au moment de son intégration.

Sablayrolles considère ainsi clairement que, lorsqu'ils intègrent la langue générale, les termes fonctionnent comme des néologismes. Il explicite son point de vue de la manière suivante :

« des termes réservés à un petit cercle de spécialistes se diffusent dans des franges plus larges de la population, voire dans la population entière. [...] La détection de l'apparition de termes dans des types de textes où ils n'apparaissaient pas auparavant se fera par l'examen de grands corpus et le repérage pourra aussi être facilité par l'usage de marques linguistiques typographiques ou de gloses explicatives qui accompagnent fréquemment cette migration d'un domaine vers la langue générale, où le terme a alors le statut de néologisme. » (Sablayrolles, 2018 : 35)

D'autres auteurs expriment un point de vue similaire. Guilbert, par exemple, affirme que l'apparition d'un terme dans la langue générale « produit un changement de caractère néologique dans le vocabulaire général » (Guilbert, 1975 : 80). De même, pour Adelstein (1996) et Estopà (2016), les termes peuvent être considérés comme des unités nouvelles lorsqu'ils intègrent la langue générale. Citons également Cañete *et al.* (2016 : 111), Vega Moreno et Llopert Saumell (2017 : 1418)

²⁶ D'autres aspects de la néologie en langue générale, qui ne sont pas directement en lien avec notre objet de recherche, sont discutés notamment par Guilbert (1975), Gaudin et Guespin (2000), Sablayrolles (2000), Renouf (2012, 2014), Cartier (2016), Cartier *et al.* (2018) ou encore Pruvost et Sablayrolles (2019).

et Díaz Hormigo (2020 : 74), qui assimilent la diffusion des termes dans la presse généraliste à un mécanisme d'enrichissement de la langue.

Dans la continuité de ces auteurs, nous considérons donc que la diffusion des termes dans la langue générale contribue au renouvellement lexical d'une langue. Parallèlement, il est reconnu que les médias, et notamment la presse, participent dans une large mesure à la propagation des néologismes dans la langue. Reutenauer affirme par exemple que « certaines sources sont plus à même de refléter une diffusion en français courant ou d'y contribuer » (Reutenauer, 2012b : 47) et met en opposition les néologismes qui sont créés dans des œuvres littéraires, qui n'ont le plus souvent pas vocation à sortir du cadre dans lequel ils ont été créés, et les néologismes qui sont créés ou apparaissent dans la presse. Elle explique ainsi que « le discours journalistique aura tendance à refléter des néologies en cours de diffusion ou à participer à une diffusion relativement large » (*Ibid.*).

La presse peut ainsi être vue non seulement comme le témoin des innovations lexicales d'une langue, mais également comme un lieu privilégié de leur diffusion. De ce point de vue, Gérard propose d'appréhender cette capacité à favoriser la création et à la diffuser parmi les locuteurs d'une langue par le biais de deux propriétés et classe les textes selon s'ils « sont plus *néologènes* que d'autres, leurs normes favorisant voire nécessitant l'innovation lexicale » (Gérard, 2018 : 38) et s'ils sont « *néolophore[s]*, c'est-à-dire incitant à réemployer des néologismes ou à en relayer la diffusion » (*Ibid.*). L'auteur poursuit et montre que la presse est caractérisée par ces deux propriétés (*Ibid.* : 39*sqq.*), ce qui en fait une source de données précieuse pour le repérage et l'analyse de néologismes.

Tirant parti de cette capacité à répandre l'usage de néologismes auprès du grand public, certains chercheurs s'intéressent naturellement à la presse pour développer des systèmes de détection automatique de néologismes. Sablayrolles (2018 : 35) donne à ce propos l'exemple du projet Neoveille. Ce projet de veille néologique consiste en une « plateforme multilingue de veille et de suivi des néologismes » (Cartier, 2016 : 112), où tant les néologismes complets ou formels que les néologismes sémantiques sont visés par le repérage automatique. De manière générale, le système s'appuie sur un corpus de presse, un dictionnaire de référence utilisé comme un corpus d'exclusion et une analyse distributionnelle pour la caractérisation des néologismes sémantiques (*Ibid.* : 113*sqq.*).

Le Logoscope constitue un autre exemple de ce type de système (Gérard *et al.*, 2016), dont les principes de fonctionnement sont semblables au projet Neoveille. Citons également les travaux de Renouf (2012, 2014, 2017), qui se situent dans cette même lignée. Dans ces deux cas, l'identification des néologismes se fonde sur de gros corpus journalistiques (en français et en anglais

respectivement). Par ailleurs, dans ces travaux, des termes sont parfois détectés par les méthodes de repérage mises en œuvre, précisément du fait qu'ils se comportent dans les textes de la même manière que des néologismes. Il semble alors tout à fait légitime de considérer que ces termes sont des néologismes dans la langue générale, ou du moins dans la variété de la langue représentée dans les corpus exploités.

Un dernier point doit être soulevé : nous disons *supra* que les termes, lorsqu'ils apparaissent dans la langue générale, interpellent les locuteurs par leur nouveauté, par leur caractère inhabituel. C'est d'ailleurs ce constat qui permet de considérer ces termes comme des néologismes dans ce cas. Cependant, bien qu'ils semblent nouveaux pour les locuteurs non experts, ils ne sont pas nécessairement néologiques dans la langue de spécialité à partir de laquelle ils sont diffusés. Il semblerait en effet peu probable que seuls les néologismes terminologiques passent dans la langue générale. De ce point de vue, la nouveauté de ces termes dans la langue générale et pour des non-spécialistes doit être nuancée.

Dans ce but, Estopà mobilise la notion de néologicité et affirme que la néologicité d'une unité (ou d'un terme) dépend essentiellement de l'usage que l'on en fait (Estopà, 2016 : 115). Il est donc possible de considérer la même unité (ou le même terme) à la fois comme une unité nouvelle et comme une unité qui n'est pas nouvelle selon si on se place de la perspective d'un locuteur expert ou d'un locuteur non expert (*Ibid.*). Dans ce cas, le degré de néologicité d'un terme dépend non seulement du contexte dans lequel il est utilisé (presse généraliste ou communication entre spécialistes), mais également du statut et des connaissances des locuteurs (non-expert, expert du domaine « d'origine » du terme, expert d'un domaine connexe, etc.).

Reutenauer exprime une idée similaire. Dans un contexte où il est question du repérage de néologismes « en cours d'implantation » (qui est une caractéristique essentielle de la définition de néologie sémantique adoptée par l'autrice (Reutenauer, 2012b : 45-47)), l'accent est porté sur la diffusion des néologismes, par le biais de la presse par exemple, qui est qualifiée de « circulation du dire » selon une formulation empruntée à Sablayrolles (2000 : 200, cité dans Reutenauer, 2012b : 46) :

« [l]a circulation du dire reste une notion relative et floue. Elle dépend de la communauté relativement à laquelle on se positionne. Ainsi, une lexie pourra être stabilisée et intégrée dans un vocabulaire de spécialité, mais elle aura un caractère de nouveauté en langue générale. Par exemple, avec la crise financière de 2008, les *subprimes* ont fait massivement leur apparition dans le vocabulaire des médias, alors que ce nom était déjà intégré à la terminologie de la finance. » (*Ibid.*)

L'exemple donné par Reutenauer reflète parfaitement la situation de circulation de termes que nous détaillons dans ce chapitre : un terme passe dans la langue générale par le biais des médias, du fait d'une popularité accrue par un évènement particulier, mais, parallèlement à cette intégration, ce terme paraît nouveau pour les locuteurs non spécialistes du domaine financier qui le rencontrent dans les médias, alors qu'il n'est absolument pas nouveau pour les spécialistes de ce domaine.

L'apparition de termes dans la langue générale, et particulièrement dans les médias généralistes, est donc un phénomène qui peut être appréhendé par le champ de la néologie, les termes étant perçus comme tout à fait nouveaux par des locuteurs non spécialistes dans les situations de communication décrites ici. Néanmoins, lorsque le sens des termes change de manière conséquente dans la langue générale, alors c'est dans une perspective de néologie sémantique que ces termes sont abordés.

1.4.2. Changements sémantiques des termes et néologismes sémantiques en langue générale

D'après Meillet, qui qualifie les échanges entre les langues de spécialité et la langue générale d'emprunts, les changements sémantiques qui se produisent lorsque les termes fonctionnent dans la langue générale sont inhérents à ce phénomène et relèvent de fait d'un mécanisme de création néologique : « [s]'ils pénètrent vraiment dans la langue commune et y sont employés couramment, les mots empruntés ne le font qu'en subissant un changement de sens » (Meillet, 1905 : 23).

L'affirmation de Meillet se vérifie toujours aujourd'hui et le champ de la néologie sémantique s'intéresse donc à ces changements sémantiques. Deux facettes du changement sémantique des termes intéressent en particulier les chercheurs : d'une part, lorsque les termes sont utilisés dans la langue générale avec de légères modifications au niveau du sens et, d'autre part, lorsque le sens des termes se modifie dans une plus large mesure et donne lieu à des néologismes sémantiques, où les nouveaux sens s'établissent à partir d'une métaphore, d'une métonymie ou autre. En effet, ces deux facettes représentent les deux grandes tendances observées et décrites dans la littérature. Les travaux qui portent spécifiquement sur ces questions du point de vue de la déterminologisation seront détaillés au Chapitre 2, mais il nous semble nécessaire, à ce stade, de fournir quelques points de repère, afin de développer notre réflexion sur le fonctionnement des termes dans la langue générale, lorsque ces questions sont abordées du point de vue de la néologie sémantique.

Dans le premier cas, il est fait mention de changements sémantiques relativement subtils dans l'emploi des termes dans un contexte non spécialisé, comme l'illustre par exemple la citation

de Béjoint (1988 : 362) *supra*, en particulier avec l'adverbe *slightly*. Cette vision est sans doute celle qui est la plus consensuelle dans la littérature, après peut-être le constat que les termes circulent effectivement en dehors des langues de spécialité et fonctionnent dans la langue générale. Nous la retrouvons ainsi dans les travaux de Gilbert (1973), Varantola (1986), Gaudin (1993, 1997, 2003, 2005), Mortureux (1995), Adelstein (1996), Pearson (1998), Delavigne (2001, 2013, 2020), Ungureanu (2006), L'Homme et Polguère (2007), Moirand (2007) ou encore Botta (2013) par exemple, dont nous avons mentionné les points principaux dans les sections précédentes.

En outre, Guilbert, qui adopte un point de vue sociologique (dans la continuité de Meillet), considère également que le transfert d'un terme d'un domaine vers les locuteurs non spécialistes « le charge d'une signification nouvelle » (Guilbert, 1973b : 22-23). Il situe alors clairement ce phénomène dans le champ de la néologie sémantique, ce qu'il explicite en affirmant qu'il s'agit d'une « forme de néologie sémantique [...] qu'on pourrait qualifier de sociologique » (*Ibid.*). L'aspect sociologique renvoie ici à la fonction essentiellement sociale de la langue.

Dans le second cas, les changements sémantiques sont plus importants et découlent le plus souvent d'usages métaphoriques ou métonymiques qui se sont suffisamment répandus dans la langue générale pour être considérés comme des néologismes sémantiques (Meyer et Mackintosh, 2000b : 130 ; Renouf, 2017 : 26). C'est ce qu'explique par exemple le *Dictionnaire de linguistique* de Dubois *et al.* (2002), à l'entrée *métaphore* :

« [L]a métaphore joue un grand rôle dans la création lexicale ; beaucoup de sens figurés ne sont que des métaphores usées. En particulier, les mutations de type métaphorique sont fréquentes dans l'évolution qui fait passer des termes spécialisés dans le vocabulaire général, c'est-à-dire d'un usage restreint à une communauté limitée à un usage plus large. » (Dubois *et al.*, 2002 : 302)

Ces mécanismes sémantiques sont également bien connus des travaux en néologie, comme en témoigne la typologie des néologismes de Sablayrolles (2011). La métaphore, la métonymie et les autres figures (par exemple l'euphémisme ou le paradoxe) sont ainsi regroupées dans la catégorie du changement de sens. De la même manière, Guilbert (1973b : 22), en donnant les exemples de la métaphore et de la métonymie, explique que ces changements de sens fonctionnent sur la base d'un « changement du groupement des sèmes afférents à un lexème, selon des modalités diverses ». C'est donc de ce point de vue que les termes qui fonctionnent dans la langue générale avec un sens différent de leur sens spécialisé, du fait d'un usage métaphorique ou métonymique qui s'est stabilisé, peuvent être considérés comme des néologismes sémantiques et sont étudiés dans des travaux qui se situent dans le champ de la néologie sémantique.

Parallèlement, Reutenauer (2012b : 87*sqq.*) propose une définition malléable de la néologie sémantique. En effet, partant du constat que « la définition de la néologie sémantique peut recouvrir un périmètre variable, aux frontières floues » (*Ibid.* : 53), l'auteur propose un modèle qui vise à refléter la diversité des définitions de la néologie sémantique présentes dans la littérature (et les avis parfois contradictoires ou peu compatibles). Elle situe ainsi la néologie sémantique dans ses rapports avec les autres types de néologies et explicite ce qui peut entrer dans la définition de la néologie sémantique, selon un certain point de vue (*Ibid.* : 53-54).

Dans sa proposition, différents niveaux de définitions sont représentés. Par exemple, dans sa définition la plus restrictive, la néologie sémantique se caractérise uniquement par des changements de sens issus de figures (métaphore, métonymie, etc.), mais dans sa définition la plus large, la néologie sémantique inclut également les mécanismes d'emprunts internes, dans lesquels Reutenauer regroupe les « migrations depuis certains sociolectes, domaines, langues de spécialité » (*Ibid.* : 54). Ces deux conceptions de la néologie sémantique rendent ainsi compte aussi bien du point de vue de Sablayrolles que de celui de Guilbert. L'inclusion des emprunts internes tels que définis ici renvoie aux propos de Guilbert que nous citons *supra*. En revanche, la définition restreinte de la néologie sémantique renvoie aux propositions de Sablayrolles.

En effet, pour Sablayrolles (2012 : 39), « [l]es évolutions de sens (extension et restriction lors du passage d'un groupe social à un autre) ne sont pas à proprement parler des néologismes ». Au contraire, étant donné que les néologismes sont caractérisés par « la recherche et/ou la production d'un effet : inadéquation assumée de l'emploi d'une unité lexicale » par exemple, et qu'ils doivent être perçus comme tels « par au moins un des participants de l'échange langagier » (Sablayrolles, 2010 : 96), alors l'auteur considère que les évolutions de sens qui se produisent lorsqu'un terme est utilisé dans la langue générale n'entrent pas dans cette définition de *néologisme*. Ces dernières sont plutôt à rapprocher de l'histoire de la langue et des évolutions de sens qui se produisent progressivement au cours du temps. Sablayrolles poursuit et explique que

« [c]e n'est qu'a posteriori que l'on constate ces changements de sens, ces évolutions se faisant de manière insensible et progressive. Aucune volonté consciente ne préside à ces changements, à la différence des néologismes sémantiques qui sont le fruit de créations délibérées le plus souvent, et qui, dans tous les cas, opèrent une rupture brutale. » (Sablayrolles, 2012 : 40)

Cependant, de notre point de vue, la proposition de Reutenauer comporte l'avantage indéniable de fédérer, en quelque sorte, ces différentes visions de la néologie sémantique présentes

dans la littérature²⁷. L'autrice est ainsi amenée à appréhender le fonctionnement de plusieurs termes dans la langue générale par le prisme de la néologie sémantique. La méthode qu'elle développe lui permet de caractériser l'évolution et le fonctionnement de néologismes issus de l'intégration de termes à la langue générale. L'exemple de *mutualiser*, qu'elle reprend de Viprey et Schepens (2010) illustre les phénomènes qu'elle vise à identifier et à caractériser. Elle le résume comme suit :

« [j]usqu'en 90, l'emploi de *mutualiser* est rare et il renvoie presque exclusivement au domaine de l'assurance. Le nouveau sens de partage de moyens et de ressources émerge et se diffuse dans les années 90, puis les emplois connaissent un basculement marqué entre 2001 et 2005 : la deuxième acception, associée à l'idée de mise en commun, s'impose dans les emplois et supplante la première acception. » (Reutenauer, 2012b : 62)

Reutenauer traite ainsi le cas du terme *tsunami*, par exemple. Son « hypothèse est que la diffusion des emplois métaphoriques est déclenchée par la catastrophe naturelle qui a frappé l'Asie du Sud-est le 26 décembre 2004 » (Reutenauer, 2012a : 1931). Cette hypothèse est ensuite vérifiée par la mise en œuvre de différents indices dans un corpus de presse divisé en domaines selon les rubriques des journaux pris en compte²⁸. L'observation de ces indices montre ainsi que « des domaines multiples émergent en corpus, ce qui invite à voir l'apparition d'un nouveau sens général » (*Ibid.* : 1940), qui s'est établi à partir des usages métaphoriques.

Selon nous, cet exemple illustre parfaitement les rapports qu'entretiennent les thématiques de la circulation des termes et de la néologie. En effet, comme nous le disons *supra* (1.2.2), les médias sont responsables dans une large mesure de la diffusion des termes vers le grand public, à la suite d'un évènement particulièrement important ou non, de la même manière qu'ils participent grandement à la diffusion, voire à la création, des néologismes (cf. section 1.4.1). Dans le cas de *tsunami*, l'évènement déclencheur de la diffusion de ce terme dans les médias et auprès du grand public a pu être identifié, ce qui a permis de formuler des hypothèses quant à sa diffusion et aux changements sémantiques inhérents à son passage dans un groupe de locuteurs plus large. Ces hypothèses ont par la suite été vérifiées et l'émergence d'un nouveau sens métaphorique a pu être établie.

²⁷ Nous n'avons dressé ici qu'un portrait partiel de la proposition de Reutenauer, pour les besoins de notre travail. Le Chapitre I.2 de sa thèse présente de manière détaillée cette proposition (Reutenauer, 2012b : 37-58).

²⁸ Sur le corpus exploité, voir Reutenauer (2012b : 117-122).

1.5. Bilan

Dans ce chapitre, nous avons montré que la déterminologisation, entendue comme le passage et le fonctionnement de termes dans la langue générale, peut être envisagée comme prenant part au phénomène global de circulation des termes. Parallèlement, à travers l'abondante réflexion existant autour de ce phénomène, nous avons insisté sur la diversité des angles d'approche du terme lorsqu'il fonctionne dans des textes qui ne sont pas produits par et pour des spécialistes, avec des enjeux qui leur sont propres. En réalité, la déterminologisation est un phénomène transversal, qui peut être questionné à travers ces différents angles d'approche.

Les différents éléments présentés dans ce chapitre permettent de mieux cerner la déterminologisation et d'alimenter la réflexion sur ce phénomène :

- en mettant en exergue la place des discours de transmission des connaissances dans le processus de transfert des termes vers la langue générale, en tant que canaux par lesquels les termes parviennent aux locuteurs non spécialistes ;
- en insistant sur les enjeux de la déterminologisation pour la lexicographie, à travers la question délicate de la description des termes dans les dictionnaires généraux, qui doit tenir compte des particularités du fonctionnement des termes dans la langue générale en comparaison avec leur langue de spécialité « d'origine » ;
- et en mettant l'accent sur la place des termes dans la création néologique en langue générale et en particulier sur les questions soulevées par la déterminologisation dans les mécanismes de néologie sémantique.

Dans le chapitre suivant, nous poursuivons cette réflexion en nous focalisant plus précisément sur la notion de déterminologisation telle que définie par Meyer (2000) à l'origine et en tirant profit des différents travaux qui interrogent cette notion.

Chapitre 2 La déterminologisation : les enjeux d'une définition

Dans ce chapitre, nous nous focalisons plus précisément sur la question de la déterminologisation. Nous cherchons en particulier à retracer l'évolution de cette notion à partir des propositions de Meyer ainsi que des travaux qui approfondissent cette question. Dans le même temps, nous cherchons à préciser notre angle d'approche en mettant en évidence les développements importants de la notion ainsi que les aspects que nous interrogeons dans notre thèse.

Nous organisons notre propos en trois sections. Dans la première section, nous revenons sur les origines de la notion de déterminologisation telle que proposée dans les travaux de Meyer et ses collègues et clarifions les rapports qu'elle entretient avec d'autres notions similaires, développées par d'autres auteurs et auxquelles elle est parfois associée. Dans la deuxième section, nous détaillons les principaux développements de la notion de déterminologisation et précisons en particulier les apports des travaux récents sur trois aspects : la description du fonctionnement des termes dans la langue générale, le rôle de la perspective diachronique et la prise en compte du continuum entre langue de spécialité et langue générale. Enfin, dans la dernière section, nous proposons une définition actualisée de la déterminologisation, qui tient compte de ces différents aspects, et présentons plus en détail la direction de notre travail de thèse dans ce contexte.

2.1. Pluralité des points de vue sur le passage des termes dans la langue générale

Dans notre travail, nous nous focalisons sur la question de la déterminologisation, à travers le point de vue de Meyer essentiellement. Afin de situer notre réflexion parmi les travaux portant sur cette question et afin de préciser les raisons qui nous amènent à conserver ce point de vue en dépit des autres propositions existantes, nous contextualisons d'abord la définition de Meyer puis nous présentons trois notions apparentées, à savoir la *banalisation lexicale*, la *dé-spécialisation* et la *dédomanialisation*.

2.1.1. La proposition de Meyer

Dans cette section, nous nous focalisons sur la proposition de Meyer, qui est le plus souvent reprise dans la littérature et dans la continuité de laquelle nous situons notre thèse. Nous revenons ainsi d'abord sur les travaux de Meyer et ses collègues à l'origine du développement de cette

notion²⁹, puis nous détaillons leur conception de la déterminologisation, à travers la revue de leurs différentes études.

2.1.1.1. Les termes métaphoriques à la base de la réflexion sur la déterminologisation

À la fin des années 1990, Meyer s'est intéressée à l'essor du vocabulaire d'Internet et de l'informatique, et particulièrement à l'émergence de termes métaphoriques et à leur diffusion auprès de locuteurs non experts de ce domaine (Meyer, *et al.* 1997). C'est précisément cet intérêt pour ce type de termes³⁰ qui amène Meyer et ses collègues à développer la notion de déterminologisation.

Dans leur étude, Meyer *et al.* (1997) soulignent l'utilité des métaphores dans la création de termes dans ce domaine, qui facilitent en grande partie la compréhension de ces concepts par des non-experts. Elles expliquent ainsi que

« [m]any laypeople are confused by the large numbers of new computing concepts they have to deal with and are sometimes downright afraid of computers. Metaphorical terms, by allowing computer users to see a potentially complex concept in terms of a well-known and simple one, aid users in understanding and remembering new concepts. At the same time, metaphorical terms allow users to associate unfamiliar concepts with old, "comfortable" ones, thereby helping to palliate technostress. » (*Ibid.* : 3-4)

Par ailleurs, c'est dans un objectif de description que les autrices recensent et analysent plusieurs dizaines de termes métaphoriques d'un point de vue conceptuel et d'un point de vue structurel. D'un point de vue conceptuel, elles montrent la diversité des domaines sources des termes métaphoriques, par exemple les transports (avec des termes tels que *cyberspace*, *anchor*, *to navigate*, *to surf*, *Internaut*) ou l'architecture (qui permet la création de termes tels que *site*, *gateway* ou

²⁹ En réalité, l'origine de la notion de déterminologisation peut être retracée jusque dans les années 1950, dans des travaux de chercheurs tchèques et slovaques, notamment Horecký en 1956 (Nová, 2018 : 388). L'autrice cite également d'autres auteurs qui ont contribué au développement de cette notion dans les décennies qui ont suivi, par exemple Jedlička *et al.* (1970), Poštolková (1977, 1984) ou encore Masár (1991), et souligne les similitudes de ces travaux avec ceux de Meyer et Mackintosh (2000b), principalement dans la description des changements sémantiques qui peuvent se produire au cours du processus de déterminologisation (que nous détaillons *infra*, en 2.1.1.3) (Nová, 2018 : 388-389). Cependant, selon l'autrice (*Ibid.*), et comme nous avons nous-même pu le constater, les travaux de Meyer et Mackintosh restent très largement cités dans la littérature sur la déterminologisation (voir également 2.2.2 et 2.2.3 *infra*). En outre, étant donné qu'à notre connaissance, ces deux autrices ne font jamais mention des travaux d'origine tchèque ou slovaque, il semble qu'elles aient développé leur vision de la déterminologisation parallèlement à ces travaux. Pour ces raisons, et du fait que les travaux que cite Nová ne nous sont pas accessibles (puisque nous ne maîtrisons pas les langues dans lesquelles ils sont publiés), nous ne détaillerons pas davantage ces points de vue.

³⁰ Notons qu'à la fin des années 1990, relativement peu de travaux s'intéressaient à la métaphore en terminologie.

bridge) (*Ibid.* : 10). D'un point de vue structurel, les autrices montrent que les termes métaphoriques peuvent être soit complètement métaphoriques soit partiellement métaphoriques. Dans le premier cas, le terme dans son ensemble, qu'il soit simple ou complexe, est considéré comme métaphorique. À ce propos, les exemples de *Web*, *spam*, *home page* ou encore *mirror site* sont donnés (*Ibid.* : 13-14). Dans le second cas, en revanche, les termes métaphoriques sont composés d'au moins un élément métaphorique et un élément non métaphorique. Il peut s'agir soit de termes dont la tête est métaphorique et l'expansion ne l'est pas (par exemple *digital signature* ou *electronic mail*), soit de termes dont la tête n'est pas métaphorique mais dont l'expansion l'est (par exemple *search engine* ou *data traffic*) (*Ibid.* : 15-16).

Dans cette analyse, les autrices relèvent tout de même plusieurs cas potentiellement problématiques pour les non-experts, notamment lorsque la métaphore peut provoquer une certaine confusion (par exemple *lurking*, dont la connotation péjorative en langue générale rend le terme métaphorique peu parlant pour des non-experts) ou lorsque les métaphores sont mixtes, c'est-à-dire lorsqu'elles s'appuient sur différents domaines conceptuels (par exemple dans l'expression *surfing the information highway*, qui repose sur les domaines des transports et de la navigation maritime) (*Ibid.* : 11-13). D'après les autrices, ces cas restent cependant marginaux et les termes métaphoriques ont plutôt tendance à faciliter leur compréhension par les non-experts qui sont susceptibles de les rencontrer dans leur vie quotidienne.

Ce dernier aspect fait en particulier partie des préoccupations de Meyer *et al.* (*Ibid.* : 20), car le domaine d'Internet et de l'informatique pénètre par nécessité dans le quotidien des locuteurs d'une langue, quel que soit leur degré d'expertise dans ce domaine, et la question de la compréhension de ces termes par des locuteurs non experts est cruciale. Selon les autrices, ce phénomène est d'ailleurs voué à s'accroître, ce qu'elles expliquent par l'émergence et le développement d'une « société du savoir » : « as we evolve into a "knowledge society", more and more laypeople are required to understand and use technical terminology » (*Ibid.*), qu'il s'agisse de termes du domaine d'Internet et de l'informatique ou d'autres domaines.

Dans ce contexte, les autrices interrogent en particulier les effets de cette société du savoir sur l'utilisation et la compréhension de termes par des non-experts. Nous résumons leurs propos dans la section suivante.

2.1.1.2. La société du savoir, un accélérateur de déterminologisation ?

Dans le but d'expliquer l'accroissement de termes qui passent dans la langue générale, Meyer emprunte la notion de société du savoir à Drucker et la définit ainsi : « a society in which

specialized knowledge is replacing manual labour as the axis around which economic development revolves » (Meyer, 2000 : 50). Cette notion se caractérise alors non seulement par le fait que plusieurs domaines de spécialité deviennent centraux dans la société, et donc dans la vie quotidienne des locuteurs, mais également par le fait que « our appetite for this knowledge is voracious » (*Ibid.*). L'autrice suggère ainsi que les locuteurs sont de plus en plus curieux et s'intéressent massivement aux connaissances spécialisées, auxquelles ils attribuent certaines valeurs. Pour Meyer, cela peut notamment s'observer par le nombre croissant d'ouvrages de vulgarisation traitant de thématiques variées et de publications destinées au grand public.

Dans cette perspective, Meyer et Mackintosh expliquent que l'émergence d'une société du savoir n'est pas responsable du passage de termes dans la langue générale, car ce phénomène est en réalité plutôt répandu et connu depuis longtemps (Meyer et Mackintosh, 2000b : 200). Mais avec l'expansion de la société du savoir, ce passage s'accélère et le nombre de termes qui intègrent la langue générale augmente, « which may result in an increased blurring of the boundaries between terminological and general language » (Meyer, 2000 : 39).

Ainsi, si le développement croissant de l'informatique en général et d'Internet en particulier à partir des années 1990 permet aux autrices de repérer une augmentation du nombre de termes de ce domaine dans la langue générale, il semble qu'avec le concept de société du savoir, ce sont également d'autres domaines de spécialité qui prennent de l'importance dans la société³¹, ce qui favorise le transfert de termes vers la langue générale. Comme l'explique Meyer,

« [b]ecoming a knowledge society means that specialized knowledge will percolate into more and more aspects of our everyday lives. Lexically speaking, this implies that increasing numbers of terms migrate into general language. »
(Meyer, 2000 : 50)

La valorisation des connaissances spécialisées dans la société du savoir encourage donc la diffusion de ces connaissances en dehors des cercles de spécialistes et, par conséquent, les termes utilisés pour les exprimer. C'est ce que Meyer *et al.* (1998) proposent de désigner avec le terme *déterminologisation*.

Par ailleurs, les autrices précisent que ce phénomène s'accompagne généralement de certains changements sémantiques (*Ibid.* : 645), qu'elles décrivent largement dans leurs différentes études. Nous en reprenons les points principaux dans la section suivante.

³¹ Par exemple, « les communications, la médecine, l'environnement et l'économie » (Meyer et Mackintosh, 2000a : 201).

2.1.1.3. Déterminologisation et changements sémantiques

La notion de déterminologisation s'est établie essentiellement à travers une analyse du sens des termes dans la langue générale³². Meyer *et al.* (1998) s'intéressent d'abord à certains termes de l'informatique, pour les raisons que nous expliquons *supra*, puis elles affinent la définition de la déterminologisation à travers l'analyse d'autres termes, qui appartiennent à des domaines différents, tels que la protection de l'environnement et la médecine, et qui occupent une place privilégiée dans la vie quotidienne des locuteurs. Ainsi, même si « [i]t is highly likely that the "friendly", uncomplicated nature of many computer words may [...] facilitate their de-terminologization » (Meyer, 2000 : 45), il semble que la déterminologisation est susceptible de concerner tous les termes pour lesquels le grand public trouve un certain intérêt.

Dans le domaine de l'informatique, Meyer *et al.* (1998) traitent par exemple le cas du terme *virtual*, dont la particularité réside dans le fait qu'il a d'abord été terminologisé pour créer notamment les termes *virtual memory* et *virtual reality*, puis qu'il est passé dans la langue générale, où il a acquis un sens différent, plus large, mais qui évoque toujours le domaine de l'informatique (*Ibid.* : 646-647). Pour ce qui est des autres domaines, Meyer et Mackintosh se focalisent sur des termes tels que *to recycle*, du domaine de la protection de l'environnement et de la gestion des déchets, et *to flatline*, du domaine de la médecine en général. Ces termes ont été popularisés auprès du grand public par une prise en compte de plus en plus marquée de la protection de l'environnement, d'une part, et par la culture cinématographique d'autre part (Meyer et Mackintosh, 2000b : 117).

Les autrices expliquent ainsi que, lorsque le terme *to recycle* est utilisé dans un contexte non spécialisé, il renvoie toujours à une idée de transformation, même si le lien avec le domaine d'origine n'est pas toujours explicite (Meyer et Mackintosh, 2000a : 206-207). Les autrices remarquent par ailleurs qu'un changement s'opère dans la nature des objets possibles du verbe : si, dans le domaine de spécialité, les objets syntaxiques de *to recycle* sont des objets ou des déchets, lorsque le terme passe dans la langue générale en revanche, les objets de ce verbe peuvent également être des concepts abstraits ou des personnes (*Ibid.*).

D'une manière similaire, le verbe *to flatline*, qui signifie en médecine « to register no activity on a heart monitor », faisant ainsi référence à la courbe d'un moniteur cardiaque qui reste horizontale lorsqu'aucun rythme cardiaque n'est détecté, se retrouve dans des contextes tout à fait variés (Meyer et Mackintosh, 2000b : 117). Meyer et Mackintosh recensent différents usages du terme *to flatline* et montrent qu'il peut être utilisé comme un synonyme de *se terminer* ou *prendre fin*,

³² Par le biais de textes issus de la presse canadienne (Meyer et Mackintosh, 2000b : 112).

voire comme un synonyme de *s'aplatir*, et même pour désigner quelque chose qui manque d'éclat ou qui est ennuyeux (*Ibid.* : 118).

L'analyse fine du fonctionnement de différents termes dans des textes qui relèvent d'une situation de communication non spécialisée permet aux autrices de dégager deux types de changements sémantiques des termes lorsqu'ils passent dans la langue générale. Le premier type est caractérisé par le « maintien des aspects fondamentaux du sens terminologique » (Meyer et Mackintosh, 2000a : 202). Comme l'expliquent les autrices,

« lorsqu'un terme entre dans la langue générale, l'essence du concept qui lui est sous-jacent est perçue par les non-spécialistes de façon semblable à celle des experts. Autrement dit, la référence à un concept par des non-spécialistes renvoie encore, à ce stade, à son sens terminologique particulier. » (*Ibid.*)

Dans ce cas, seul le « degré de compréhension du concept par le non-spécialiste » (*Ibid.*) diffère. De fait, un non-spécialiste ne possède pas les connaissances suffisantes pour comprendre un terme de manière aussi approfondie ou complète qu'un expert. En outre, selon Meyer et Mackintosh, c'est ce sens plus superficiel qui est habituellement décrit dans les dictionnaires généraux. Les autrices expliquent également qu'avec ce premier type de changements sémantiques, il peut se produire un changement de connotation, ou même l'apparition de nouvelles connotations, lorsque le terme passe dans la langue générale et est utilisé par des non-experts. Elles donnent ainsi l'exemple des termes *atomic* et *peroxide*, qui ont tendance à être associés respectivement à la bombe atomique et Hiroshima et à des produits chimiques servant à colorer les cheveux (Meyer et Mackintosh, 2000b : 115). Néanmoins, il semble que ces connotations ne modifient pas davantage le sens de ces termes dans la langue générale.

Le second type de changements sémantiques est caractérisé par une véritable « dilution du sens terminologique d'origine » (Meyer et Mackintosh, 2000a : 205). Les autrices expliquent ici que lorsqu'un non-expert utilise les termes qui correspondent à ce second type, il « ne les emploie pas en pensant au sens fondamental des termes d'origine », bien que ces termes « conservent une certaine ressemblance avec le terme d'origine » (*Ibid.*). Ainsi, ces termes se retrouvent dans des contextes qui ne font plus du tout référence aux domaines auxquels ils appartiennent au départ, mais leur sens reste lié de manière plus ou moins marquée à ces domaines, comme dans l'exemple de *to recycle*, cité plus haut. Les autrices expliquent à ce propos que ces usages « relèvent d'extensions figurées du sens terminologique d'origine » et que cette « extension sera plus ou moins poussée selon l'unité lexicale en question » (*Ibid.* : 208). Elles illustrent ce second type avec le terme *anorexic* dont l'extension est plus ou moins importante en fonction des contextes dans lesquels il est utilisé :

« we find that the extension of the basic terminological meaning can be seen as a continuum of extensions, some closer to the original meaning, and some further away. *Anorexic*, for example, diluted its original meaning from ‘thinness due to a medical condition’ to simply ‘thinness’ (*anorexic gymnasts*). A further dilution occurs from ‘thinness’ to ‘smallness’ (*anorexic cars*). Yet another dilution occurs from ‘physical smallness’ to ‘non-physical smallness’, which can also be expressed as ‘weakness’ or ‘lack of substance’ (*anorexic dollar, anorexic plot*). » (Meyer et Mackintosh, 2000b : 120)

Dans tous ces exemples, le domaine d’origine de ce terme, la médecine, reste perceptible. Ainsi, avec les différents sens du terme *anorexic* qu’elles observent, les autrices illustrent toute une gamme de changements sémantiques possibles suivant le passage de termes dans la langue générale. Notons également que les changements sémantiques qui résultent du processus d’intégration des termes dans la langue générale peuvent se remarquer par un fonctionnement syntaxique différent. Par exemple, dans le cas de *to flatline*, Meyer et Mackintosh (2000b : 121-122) observent que le verbe s’utilise aussi de manière transitive en langue générale, alors qu’il est exclusivement intransitif en médecine³³.

À travers ces différents exemples, nous constatons que les travaux de Meyer et ses collègues se focalisent majoritairement sur les aspects sémantiques résultant de la détermination. Ce choix s’explique notamment par l’objectif d’application que les autrices poursuivent, qui est fortement ancré dans la pratique terminographique et lexicographique. C’est pourquoi plusieurs enjeux de la détermination sont présentés d’un point de vue lexicographique. La question de la définition est notamment abordée par Meyer :

« [h]ow should lexicographers describe computing words that have moved into general language? Most importantly, they cannot simply “borrow” existing terminographic descriptions intact. The process of de-terminologization can cause words to undergo significant changes in semantics, level of language, and grammar. » (Meyer, 2000 : 46)

Ces observations rejoignent alors les questionnements que nous présentons dans la section 1.3.2 au Chapitre 1 et en soulèvent d’autres. En effet, d’une part, l’équilibre entre précision scientifique et compréhensibilité des définitions des termes dans les dictionnaires généraux est une question récurrente et qui, semble-t-il, ne peut pas être résolue simplement. D’autre part, selon Meyer, la description dans les dictionnaires généraux des termes dont le sens se modifie de manière conséquente au cours du processus de détermination pourrait être améliorée. Meyer remarque par exemple que les changements sémantiques profonds des termes *to download* et *virtual*

³³ Les différents changements sémantiques qu’observent Meyer et Mackintosh rappellent bien évidemment les préoccupations de la néologie sémantique, par exemple dans les travaux de Reutenauer (2012a, 2012b), que nous citons au Chapitre 1 (1.4.2).

– ce qu’elle appelle la « dilution du sens terminologique d’origine » dans Meyer et Mackintosh (2000a : 205) – ne sont pas adéquatement décrits dans les dictionnaires généraux. Meyer suggère alors, pour rectifier ces manquements, que les

« lexicographers need to be aware that any specialized word sense has the potential to “loosen” in this way. Hence, looking for this phenomenon should be a standard part of the job of analyzing terminological senses. Lexicographers also need to be aware that such semantic “loosening” can happen quickly, and that they must therefore consult very recent corpora. » (Meyer, 2000 : 52)

Par ailleurs, selon l’auteurice, les articles des dictionnaires généraux qui rendent compte des termes qui sont passés dans la langue générale devraient également porter une plus grande attention aux différents changements d’ordre syntaxique, voire stylistique (*Ibid.* : 53).

Les travaux en lexicographie et en métalexigraphie que nous présentons au Chapitre 1 évoquent bien évidemment des problématiques similaires. Notamment, le critère de la banalisation lexicale, selon Josselin-Leray (2005), reprend l’idée d’intégration des termes dans la langue générale. De la même manière, le critère de la métaphorisation (*Ibid.*) rappelle le second type de changements sémantiques identifiés par Meyer et Mackintosh. Cependant, il nous semble que Meyer propose de franchir une étape supplémentaire, qui n’est pas nécessairement liée à la sélection des termes à inclure dans les dictionnaires généraux, mais plus spécifiquement à leur description. L’auteurice propose ainsi de tenir compte de la rapidité possible des changements sémantiques qui se produisent lorsque les termes passent dans la langue générale, suggérant des mises à jour plus régulières pour ce qui est de ces termes. Dans le même temps, l’auteurice reconnaît la difficulté de sa proposition, toutes les ressources matérielles et humaines destinées à la mise à jour de dictionnaires ne pouvant bien évidemment pas être allouées à la description des termes uniquement. Pour cette raison, elle propose une approche plus pratique et opérationnelle, qui consiste à expliciter les usages potentiels des termes en langue générale, surtout lorsque ces termes ont tendance à être employés de manière créative et/ou humoristique, et à fournir de nombreux exemples (Meyer, 2000 : 53). Cette approche suppose donc une analyse rigoureuse du fonctionnement des termes qui passent dans la langue générale et une connaissance poussée du processus de déterminologisation.

Les différentes suggestions de Meyer à propos de la lexicographie nous amènent par ailleurs à considérer le fait qu’une analyse fine du processus de déterminologisation, telle que nous proposons de la mettre en œuvre dans notre thèse, pourrait profiter à la lexicographie, par exemple pour améliorer le traitement des termes dans les dictionnaires généraux. Cette proposition s’inspire largement des travaux de Josselin-Leray (2005, 2010), qui suggère entre autres d’exploiter un corpus

de vulgarisation pour « affiner la description des termes dans les dictionnaires généraux » (Josselin-Leray, 2010 : 65).

Nous retenons de ces différents travaux que la déterminologisation est un phénomène qui se produit lorsque des termes présentent un certain intérêt pour des locuteurs non spécialistes. Ces termes sont alors utilisés dans des contextes non spécialisés et par des locuteurs non spécialistes. Nous retenons également que différents changements sémantiques peuvent s'observer dans ce cas, en comparaison avec leur usage dans un contexte spécialisé. Ces aspects sont importants et permettent de contraster avec les notions que nous présentons dans la section suivante.

2.1.2. La banalisation lexicale selon Galisson

La notion de banalisation lexicale est parfois évoquée lorsqu'il est question du passage de termes d'une langue de spécialité à la langue générale, par exemple par Gentilhomme (1994b), Adelstein (1996) ou Josselin-Leray (2005). À l'origine, ce terme a été proposé par Galisson (1978, 1979)³⁴, dans une étude visant à observer « ce qui se passe au niveau du vocabulaire, quand un langage banalisé se juxtapose à un langage technique » (Galisson, 1978 : 11). La proposition de Galisson vise ainsi à distinguer ce qu'il appelle le « langage banalisé » de ce qu'il appelle le « langage technique » et répond, selon lui, à la nécessité de s'intéresser de plus près au langage banalisé.

Le langage banalisé est défini par l'auteur comme une « langue parallèle » (Galisson, 1979 : 76) ou un « langage second, greffé sur un langage "technique", pour assurer une diffusion plus large aux informations relevant du domaine d'expérience couvert par le langage technique en question » (Galisson, 1978 : 9). Dans ce cas, *langage technique* doit être entendu comme le « langage technique "proprement dit" », c'est-à-dire « le langage des spécialistes » (*Ibid.* : 11), ce que l'auteur appelle également parfois *langue de spécialité* (Galisson, 1979 : 76). De la même manière, l'auteur considère qu'une langue de spécialité peut être divisée en différentes strates, selon la spécialisation, par analogie avec les différents niveaux de langue de la langue générale. De ce point de vue, le langage banalisé correspond à une strate d'une langue de spécialité (*Ibid.*). Ainsi, le phénomène de banalisation lexicale, et donc le langage banalisé, relèvent, pour Galisson, des langues de spécialité

³⁴ Précisons toutefois que l'usage du terme *banalisation* dans le sens de « passage de termes dans la langue générale » est antérieur à l'usage de *banalisation lexicale* selon Galisson. Le terme *banalisation* semble en effet avoir été introduit pour la première fois par Dubois (1966), dans un compte-rendu de l'ouvrage de Guilbert (1965), pour désigner les phénomènes de mouvements de termes d'une langue de spécialité à la langue générale qu'observe ce dernier. Dubois explique ainsi que « [d]ans sa thèse complémentaire [...], [Guilbert] ajoute à la problématique des origines du vocabulaire technique dans les milieux de spécialistes, celle de la banalisation, c'est-à-dire de la pénétration dans la langue commune, de cette intégration que les moyens modernes d'information ont accélérée » (Dubois, 1966 : 103).

et non de la langue générale. Comme l'explique l'auteur, le langage banalisé « tend vers le langage courant, mais ne se confond pas avec lui » (Galisson, 1978 : 394)³⁵.

En outre, si le langage banalisé se situe à mi-chemin entre une langue de spécialité et la langue générale et vise à mettre certaines connaissances spécialisées à la portée de locuteurs non spécialistes (mais du moins amateurs) d'un domaine (*Ibid.* : 400), il se distingue toutefois de la vulgarisation. En effet, pour Galisson, le « langage de vulgarisation » (selon l'expression utilisée par l'auteur) constitue un langage parallèle distinct du langage banalisé et vise en particulier une proportion plus large de locuteurs non spécialistes – et non amateurs – d'un domaine. Sur une échelle allant du plus spécialisé au moins spécialisé, l'auteur hiérarchise ainsi le langage technique, le langage banalisé, le langage de vulgarisation et la langue générale (*Ibid.* : 401-402). Les locuteurs destinataires de ces « langues parallèles » jouent également un rôle dans la distinction entre langage banalisé et langage de vulgarisation selon Galisson, car le langage banalisé est essentiellement destiné à une proportion de locuteurs plus large que le langage technique mais plus restreinte que le langage de vulgarisation et la langue générale (*Ibid.*). Autrement dit, comme l'auteur l'exprime lui-même, « la banalisation [...] sert *aux initiés* (plus exactement aux semi-initiés) », tandis que « la vulgarisation [...] sert à *initier* » (Galisson, 1979 : 75).

Ces distinctions désormais établies, soulignons qu'avec le phénomène de banalisation lexicale, Galisson s'intéresse également aux termes. En effet, dans son analyse du langage banalisé du domaine du football, l'auteur recense la part du vocabulaire banalisé qui est composée de termes et celle qui est composée d'unités de la langue générale. Il arrive à la conclusion que les termes et les unités de la langue générale représentent des proportions équivalentes à l'intérieur du vocabulaire banalisé, soit environ 50 % dans les deux cas. Autrement dit, si le vocabulaire banalisé emprunte à la langue générale, il emprunte également, et dans les mêmes proportions, à la langue de spécialité du domaine en question (Galisson, 1979 : 100-101). Dans cette perspective, il est possible de considérer que le phénomène de banalisation lexicale s'apparente à la déterminologisation, car, dans les deux cas, des termes parviennent bien à un ensemble de locuteurs qui ne sont pas à strictement parler des spécialistes du domaine auquel ces termes appartiennent. Néanmoins, nous pensons que plusieurs caractéristiques permettent de différencier la banalisation lexicale de la déterminologisation.

Premièrement, selon Galisson,

« la vocation du langage banalisé est d'élargir l'audience d'une discipline, donc de fonctionner comme courroie de transmission entre les authentiques

³⁵ Le terme *langage courant* est utilisé par Galisson dans le sens de *langue générale* que nous avons précisé au Chapitre 1 (1.1.2).

spécialistes et les amateurs plus ou moins éclairés de la discipline » (Galisson, 1978 : 400)

La banalisation lexicale dénote donc une certaine volonté d'atteindre un public plus large, alors que la déterminologisation telle que définie par Meyer et Mackintosh ne semble pas désigner un processus conscient, mais s'apparente plutôt à l'évolution de la langue de manière générale (nous revenons sur ces aspects dans la dernière section de ce chapitre).

Deuxièmement, le langage banalisé, qui résulte de la banalisation lexicale, est destiné aux amateurs d'un domaine, c'est-à-dire des locuteurs qui s'intéressent explicitement à ce domaine et qui possèdent quelques connaissances, même rudimentaires. En revanche, la déterminologisation concerne le passage et le fonctionnement de termes dans la langue générale, c'est-à-dire dans des situations de communication non spécialisées, où des locuteurs non experts sont susceptibles d'employer ces termes.

Troisièmement, Galisson s'intéresse principalement à la constitution du vocabulaire du langage banalisé, par l'analyse des différentes unités qui le composent, alors que les travaux de Meyer et ses collègues sur la déterminologisation traitent de différents enjeux, à propos notamment des changements sémantiques qui peuvent se produire au cours du processus et des implications lexicographiques que cela représente. Ainsi, si les préoccupations de ces auteurs diffèrent, c'est du fait que les phénomènes linguistiques sur lesquels ils se focalisent sont effectivement différents.

Enfin, soulignons que, pour Galisson, le langage banalisé constitue une strate particulière d'une langue de spécialité donnée, tandis que la déterminologisation désigne un processus de passage de termes vers la langue générale. Les ensembles pris en compte diffèrent donc inévitablement avec, d'un côté, les strates qui constituent les langues de spécialité et, de l'autre, la langue générale. Au vu de ces différents aspects, il semble bien qu'à l'origine, les termes *déterminologisation* et *banalisation lexicale* désignent deux réalités distinctes.

Toutefois, d'autres auteurs reprennent la notion de banalisation lexicale à partir des travaux de Galisson de 1978 et 1979, dans des contextes qui s'éloignent parfois des propositions originales de cet auteur. Par exemple, Jacobi introduit une nuance qui n'est pas présente dans les travaux de Galisson. En effet, cet auteur semble associer la banalisation lexicale à un processus conscient de diffusion auprès d'un groupe de locuteurs plus large, qui, comme la vulgarisation scientifique, contribue à la sociodiffusion des sciences. C'est ce que met en évidence la citation suivante :

« la banalisation lexicale et l'intégration incessante de néologismes, de mots empruntés à des langues étrangères et aux terminologies propres aux langues de

spécialité **témoignent d'une sociodiffusion manifeste et voulue** par les producteurs de concepts. » (Jacobi, 1986 : 57)³⁶

Ainsi, pour cet auteur, « la banalisation marque la réussite d'un terme et la généralisation de son usage hors des limites du petit cercle des spécialistes » (*Ibid.* : 44) et s'inscrit dans une logique de sociodiffusion des connaissances, au même titre que la vulgarisation. Cette conception n'est cependant pas totalement contradictoire avec celle de Galisson qui, bien qu'il n'ancre pas ses propositions dans la problématique de la sociodiffusion des connaissances, attribue au langage banalisé – et donc à la banalisation lexicale – un rôle important de transmission auprès d'un public plus large. Néanmoins, c'est la définition de ce public plus large qui diverge entre ces deux auteurs : si Galisson insiste sur le rôle de la banalisation auprès de locuteurs que l'on pourrait qualifier de semi-experts selon la terminologie de Bowker et Pearson (2002 : 28), Jacobi suggère que la banalisation atteint des locuteurs non spécialistes. Soulignons que cette idée se retrouve également chez Gentilhomme (1994b : 548), Adelstein (1996), Boulanger (1996 : 138-139) et Cabré (1999 : 63).

Enfin, dans des travaux plus récents, la banalisation est explicitement associée à la déterminologisation (Carras, 2002 : 184-185 ; Resche, 2004 : 355 ; Drouin et Langlais, 2006 : 398 ; Díaz Hormigo, 2020 : 74³⁷). Bien qu'il ne soit pas impossible que ces auteurs établissent de légères nuances entre *déterminologisation* et *banalisation*, ces deux termes semblent fonctionner comme des synonymes, comme le montrent par exemple les propos de Sablayrolles (2018 : 35) : « [d]es évolutions collectives se manifestent aussi dans les faits de déterminologisation appelée aussi *banalisation* ». Ces remarques nous permettent donc de supposer que la notion de banalisation lexicale a évolué au fil des travaux qui se sont penchés sur cette question et qui se sont approprié le phénomène d'abord décrit par Galisson. La banalisation lexicale semble désormais s'apparenter à un processus de passage de termes dans la langue générale, de manière similaire à la déterminologisation³⁸. Néanmoins, au vu des quelques divergences soulignées dans cette section, et dans le but d'assurer la clarté de notre propos, nous jugeons qu'il est préférable de ne pas introduire d'ambiguïté avec l'utilisation du terme *banalisation lexicale*.

³⁶ Nous soulignons.

³⁷ Dans ce dernier cas, bien que l'autrice utilise les termes espagnols (*determinologización* et *banalización*), elle les associe tous les deux au passage de termes dans la langue générale (Díaz Hormigo, 2020 : 74).

³⁸ Ce qui se rapproche également de la banalisation selon Dubois (1966).

2.1.3. La dé-spécialisation : une notion aux contours flous

Le terme *dé-spécialisation* (parfois *déspécialisation*) se trouve dans des travaux qui se concentrent de manière générale sur l'évolution d'un terme vers un sens moins spécialisé, avec certaines nuances selon les travaux. Trois points de vue en particulier peuvent être distingués. Le premier s'observe dans une étude effectuée par Mazière (1981), à propos du traitement des termes dans les dictionnaires généraux. L'autrice compare en particulier les marques de domaine pour une sélection de termes dans le Littré et dans le Petit Robert, supposant une évolution entre ces deux dictionnaires élaborés à des époques différentes (Mazière, 1981 : 80-81). Elle observe ainsi que le terme *binage* est accompagné de la marque de domaine *agriculture* dans le Littré, alors qu'aucune marque de domaine n'est précisée dans le Petit Robert. L'autrice attribue cette différence de traitement à la démocratisation de ce concept, voire du domaine de l'agriculture de manière générale. Selon elle, « la multiplication des jardins de banlieue » explique la « dé-spécialisation de *Binage* » (*Ibid.* : 84). Autrement dit, si l'ajout d'une marque de domaine pour *binage* ne paraît pas nécessaire dans le Petit Robert, c'est probablement du fait que cette unité n'est pas perçue comme un terme, étant donné que le binage n'est plus réservé strictement aux agriculteurs, mais qu'il est connu des personnes qui cultivent « des jardins de banlieue ».

De ce point de vue, la dé-spécialisation renverrait au processus par lequel un concept se démocratise auprès d'un public plus large, non spécialiste. C'est d'ailleurs sur cette proposition de Mazière que Meyer et Mackintosh s'appuient pour créer le terme *déterminologisation* (Meyer et Mackintosh, 2000a : 215). Les autrices expliquent cependant qu'elles préfèrent ne pas « le conserver tel quel en français étant donné qu'il pourrait s'appliquer à toute unité lexicale (y compris non terminologique) dont le sens lexical devient plus général » (*Ibid.*).

Le deuxième point de vue est celui de Gouadec (1990), qui évoque la dé-spécialisation dans le but d'illustrer l'« évolution continue » des terminologies (Gouadec, 1990 : 23-26) dans ce qu'il appelle des « cycles terminologiques » (*Ibid.* : 25). Ainsi, selon lui, cette évolution

« intéresse les termes pris individuellement avec, pour chacun d'eux, un cycle dans lequel se succèdent (globalement), selon un rythme aléatoire, l'emprunt ou la création, puis la spécialisation progressive, puis l'état d'équilibre des termes spécialisés, puis la dé-spécialisation ou vulgarisation, et enfin l'émergence de désignations nouvelles. » (*Ibid.* : 23)

La phase du cycle que Gouadec nomme *dé-spécialisation* correspond aux différents changements que peuvent subir les termes lorsqu'ils sont utilisés par des non-experts. L'auteur explique alors que « les termes spécialisés perdent peu à peu leur "pureté" de désignation » lorsqu'ils passent « dans des groupes plus étendus (groupes de non-spécialistes) », ce qui se caractérise par

« une relative modification de leurs acceptions » (*Ibid.* : 24). Il précise également que « les cycles terminologiques sont relativement lents » (*Ibid.* : 25), ce qui suppose une évolution progressive du sens des termes lorsqu'ils passent dans des « groupes de non-spécialistes ».

Enfin, le troisième point de vue est celui de Condamines et Picton et semble plus proche de celui de Mazière. Les autrices expliquent que « la "déspécialisation" de la connaissance [...] renvoie au phénomène de transfert de la connaissance spécialisée vers le grand public » (Condamines et Picton, 2014b : 697). L'accent est porté ici sur les concepts, et, par extension, sur les connaissances spécialisées de manière générale, et moins sur les termes, au contraire de la déterminologisation qui, selon les autrices, « correspond au passage d'une partie de la terminologie d'un domaine spécialisé vers la langue générale » (*Ibid.*). Cependant, étant donné qu'il semble délicat de dissocier complètement les termes des concepts, les phénomènes de déterminologisation et de déspecialisation sont nécessairement apparentés dans la réalité.

Par ailleurs, Condamines et Picton, contrairement à Gouadec, établissent une différence entre déspecialisation et vulgarisation :

« [l]a déspecialisation ne concerne [...] qu'en partie les phénomènes de vulgarisation qui eux, s'inscrivent dans une volonté presque didactique de mettre à portée du grand public une connaissance spécialisée. » (Condamines et Picton, 2014b : 697)

Ces remarques mettent ainsi en lumière, d'une part, le fait qu'il existe plusieurs nuances dans les usages du terme *déspecialisation* selon les auteurs qui traitent de cette question. Ce terme reste par ailleurs très peu utilisé dans la littérature, ce qui pourrait potentiellement expliquer ce manque de consensus. D'autre part, il semble bien que les phénomènes de déspecialisation et de déterminologisation doivent être distingués selon si l'accent est mis sur les connaissances ou sur les termes.

2.1.4. La dédomanialisation : une approche sémique du passage des termes dans la langue générale

Le dernier phénomène dont nous souhaitons traiter est le phénomène de *dédomanialisation* (Rastier et Valette, 2009 : 105-106 ; Rastier, 2011 : 157). Les auteurs, se situant dans une perspective de néologie sémantique et mettant en œuvre une approche sémique du sens, expliquent que « la *dédomanialisation* autonomise un sémème par rapport à son domaine d'origine » (Rastier et Valette, 2009 : 105). Autrement dit, la dédomanialisation concerne les unités dont l'usage se répand en dehors d'un domaine particulier et qui perdent certains sèmes caractéristiques de ce domaine, voire en acquièrent d'autres, qui ne sont pas relatifs à ce domaine.

Plus précisément, pour Rastier et Valette, ce phénomène se produit « par l'inhibition des sèmes inhérents » (*Ibid.* : 114) d'une unité, qui sont spécifiques au domaine. Les auteurs donnent à ce propos l'exemple de *caviar* et montrent que les sèmes relatifs au domaine ne sont pas actualisés dans des contextes où *caviar* est dédomanialisé, mais seuls les sèmes afférents. Ils expliquent que « lorsque le sème /luxe suprême/ – et lui seul – est activé, *caviar* est alors l'expression de la haute qualité » (*Ibid.*), par exemple dans l'expression « c'est du caviar », pour qualifier quelque chose de très bonne qualité (*Ibid.* : 115). Dans ce cas, c'est bien la perte de la référence au domaine d'origine du terme qui permet aux auteurs d'observer le phénomène de dédomanialisation.

À la suite de ces deux auteurs, Reutenauer explique que la dédomanialisation renvoie au nouveau sens d'une unité qui « s'affranchit du ou des domaines spécifiques d'emploi, notamment pour évoluer vers un emploi général » (Reutenauer, 2012a : 1928). Par ailleurs, à travers l'analyse fine d'un corpus diachronique, l'autrice précise cette définition en montrant que l'émergence d'un sens plus général issu de la dédomanialisation d'une unité se caractérise par l'usage de cette unité dans des contextes évoquant plusieurs domaines différents (Reutenauer, 2012b : 200). En effet, selon elle, l'augmentation de la fréquence des unités qu'elle analyse dans plusieurs des domaines couverts par son corpus « peut être le témoin d'une forte diversification des domaines d'emploi et être l'indice d'une dédomanialisation du sens de l'unité lexicale » (*Ibid.*). L'« évolution d'emplois spécialisés [...] vers un usage courant » (*Ibid.* : 197) peut alors également se constater par la propagation des unités en question dans des domaines diversifiés et la multiplication de ces domaines tend à prouver la dédomanialisation de ces unités.

Cependant, la dédomanialisation est un phénomène qui relève d'une perspective linguistique particulière. De fait, les travaux que nous citons s'inscrivent dans la problématique globale de la néosémie, avec une approche sémique du sens, caractéristique de la sémantique interprétative (Rastier, 2009). En outre, bien que le rapprochement entre dédomanialisation et déterminologisation soit indéniable (il s'agit dans les deux cas de termes qui intègrent la langue générale et dont le sens se modifie au cours du processus), ces deux phénomènes renvoient à des approches différentes : la déterminologisation met l'accent sur les termes, avec un point de vue résolument terminologique sur les questions soulevées par ce phénomène, tandis que la dédomanialisation permet d'interroger des problématiques de néosémie en langue générale. Ainsi, et étant donné que nous nous situons dans une perspective de terminologie, il nous semble plus judicieux de nous focaliser exclusivement sur la déterminologisation.

En somme, bien que les notions de dédomanialisation, de dé-spécialisation et de banalisation lexicale partagent certaines caractéristiques avec la déterminologisation, il reste certaines nuances dans les travaux qui traitent de ces questions, que ce soit au niveau des objets

concernés par le phénomène (par exemple les concepts ou les connaissances spécialisées dans le cas de la déspecialisation, ou les termes et les unités de la langue générale dans le cas de la banalisation lexicale, par rapport aux termes uniquement dans le cas de la déterminologisation) ou au niveau des orientations théoriques et méthodologiques à partir desquels les phénomènes sont abordés (par exemple la néologie sémantique et la sémantique interprétative dans le cas de la dédomanialisation, par rapport à la terminologie et aux langues de spécialité dans le cas de la déterminologisation). Par ailleurs, le terme *déterminologisation* est, semble-t-il, celui qui a retenu l'attention d'un plus grand nombre de chercheurs. Pour toutes ces raisons, et afin d'éviter d'introduire une ambiguïté peu souhaitable dans notre travail, nous faisons le choix de conserver uniquement ce terme.

Enfin, soulignons que les travaux de Meyer restent largement cités et repris dans la littérature, ce qui permet d'alimenter la réflexion à propos de la déterminologisation et de contribuer au développement de ces questions. Nous synthétisons la richesse de ces travaux et leur apport pour notre propre travail dans la section suivante.

2.2. Dans la continuité des propositions de Meyer

Dans cette section, nous organisons notre propos en deux temps. Nous présentons d'abord plusieurs travaux qui reprennent la notion de déterminologisation telle qu'elle est définie dans les deux articles de Meyer et Mackintosh (2000a, 2000b)³⁹ et qui s'en servent pour contextualiser différents phénomènes liés à la circulation des termes dans des ouvrages traitant de la terminologie de manière générale ou pour commenter et interpréter certains phénomènes observés dans le contexte d'une étude particulière. Puis nous détaillons trois axes de développement de cette notion, à partir de travaux qui interrogent de manière approfondie la déterminologisation et qui la mettent à l'épreuve des données. Ces travaux constituent une seconde base à nos réflexions, après les études fondatrices de Meyer et ses collègues. Nous reprenons ensuite ces trois axes pour proposer une

³⁹ À ce stade, il nous semble utile de préciser qu'un usage concurrent du terme *déterminologisation* s'observe parfois dans la littérature, en particulier dans des travaux en anglais (avec le terme *determinologization*) et en espagnol (avec le terme *desterminologización*). Dans ces cas, les travaux traitent de problématiques liées à la traduction dans le domaine médical et, de manière plus générale, à la communication entre médecins et patients (Montalt-Resurrecció et Shuttleworth, 2012 ; Montalt-Resurrecció et González Davies, 2014 ; Campos Andrés, 2013 ; Pontrandolfo, 2016 ; Jiménez-Crespo, 2017 ; Jiménez-Crespo et Tercedor Sánchez, 2017 ; Cobos López, 2019). Pour ces différents auteurs, la déterminologisation fait partie des stratégies de reformulation et d'explicitation des termes supposés non connus des destinataires, ce qui suppose un acte volontaire d'adaptation du contenu pour qu'il soit accessible pour un public non expert. Cette conception de la déterminologisation diffère dans une large mesure des propositions de Meyer et Mackintosh et semble évoluer en parallèle, dans une situation particulière. C'est pourquoi elle ne nous semble pas pertinente pour notre travail et nous ne jugeons pas nécessaire de poursuivre davantage la réflexion dans cette direction.

définition actualisée de la déterminologisation, dans la dernière section de ce chapitre, et pour construire notre cadre d'analyse, dans les Chapitres 3 à 5.

2.2.1. Reprise des propositions de Meyer

Dans un premier temps, intéressons-nous aux auteurs qui se réfèrent à la déterminologisation de manière générale. Par exemple, L'Homme, qui considère que les termes sont définis comme des unités lexicales qui ont un sens spécialisé dans un domaine (L'Homme, 2020 : 5), cite les phénomènes de terminologisation et de déterminologisation dans un contexte orienté vers la pratique de la terminologie (*Ibid.* : 92-94). L'autrice aborde en particulier les questions que peuvent soulever ces deux phénomènes dans la gestion des termes dans des ressources terminographiques, lorsqu'un sens spécialisé et un sens général coexistent pour une même unité lexicale (*Ibid.*).

De ce point de vue, l'autrice illustre la déterminologisation avec le terme *to debug*, dont l'usage s'est répandu chez les locuteurs non spécialistes de l'informatique :

« the verb *debug*, although first used by computer scientists, is now quite common outside the realm of computer science. [...] For instance, in general language *to debug* can refer to different situations in which a problem is fixed (and not necessarily a program). » (L'Homme, 2020 : 23)

C'est d'ailleurs un point de vue similaire qu'exprime Cabré (2005, 2016). Dans sa théorie communicative de la terminologie, les termes se caractérisent par le fait qu'ils sont des unités lexicales dont le sens spécialisé est activé en fonction de la situation de communication : « c'est le contexte pragmatique qui est la clé qui active la valeur spécialisée de chaque unité terminologique dans un usage donné » (Cabré, 2016 : 78). L'autrice poursuit :

« [L]e sens spécialisé de chaque terme dans le fonctionnement discursif constitue une valeur associée aux unités lexicales. Cette valeur spécifique, actualisée dans un sens spécifique, n'est inhérente à aucune unité en soi si ce n'est qu'elle est conditionnée par l'usage. Il s'agit donc d'une valeur virtuelle qui peut se matérialiser ou pas dans chacune des unités lexicales d'une langue. Les processus de terminologisation [...] et de déterminologisation [...] peuvent s'expliquer aisément par cette capacité du lexique à associer une valeur spécialisée selon les circonstances communicatives. » (*Ibid.*)

Parallèlement, Humbley rapproche la déterminologisation du glissement de sens, en particulier d'après le point de vue de la théorie continuiste de Nyckees (2006, cité dans Humbley, 2018a : 178). La déterminologisation s'apparenterait alors à une évolution du sens d'un terme, lorsqu'il pénètre dans l'usage d'un groupe plus large de locuteurs, non spécialistes du domaine

auquel appartient le terme à l'origine (Humbley, 2018a : 178). Sablayrolles, quant à lui, évoque brièvement la question de la déterminologisation dans la perspective de la néologie sémantique et suggère que la déterminologisation serait à considérer comme un mode de création néologique (Sablayrolles, 2018 : 35). Citons également Vega Moreno et Llopert Saumell (2017 : 1418) et Díaz Hormigo (2020 : 74), qui expriment un point de vue similaire. Enfin, notons les travaux de Ledouble, dans lesquels le phénomène de déterminologisation est cité dans le contexte plus large de la médiatisation de la science (Ledouble, 2019 : 60-61).

Dans un second temps, nous relevons plusieurs cas où la notion de déterminologisation est invoquée pour contribuer à l'interprétation de certains phénomènes linguistiques constatés dans des données (Resche, 2004 ; L'Homme et Polguère, 2007 ; Bertaccini *et al.*, 2008 ; Botta, 2013 ; Peynaud, 2016 ; Sajous *et al.*, 2018 ; Vogh et Courbon, 2018). Dans ces cas, la déterminologisation ne semble pas être l'objet principal, ou du moins pas le seul objet, de ces travaux. Par exemple, Resche, dans le domaine de l'économie, constate un certain flou autour du terme anglais *recession* et suppose que ce flou pourrait être dû à l'utilisation massive du terme par des locuteurs non experts et/ou dans des contextes non spécialisés (Resche, 2004 : 350*sqq.*). L'autrice qualifie même ce terme de « dénomination fantôme qui n'a plus de notion précise sous-jacente » et rapproche cette « perte de contenu sémantique précis » des conséquences sémantiques du processus de déterminologisation sur les termes (*Ibid.* : 355).

De la même manière, citons l'étude de Botta (2013), où la notion de déterminologisation est exploitée afin de caractériser les différences d'usage des termes dans un corpus de textes de vulgarisation en comparaison avec un corpus de textes que l'autrice qualifie de scientifiques (Botta, 2013 : 283). Dans ce cas, la déterminologisation sert à distinguer les cas où des changements sémantiques sont observables dans le corpus de vulgarisation par rapport au corpus scientifique des cas où les termes semblent relever du même usage dans les deux corpus. Pour l'autrice, seul le premier cas correspond à la déterminologisation (*Ibid.* : 290-292).

Précisons à ce stade que la déterminologisation est mentionnée à plusieurs reprises dans des travaux qui traitent de vulgarisation scientifique ou de médiatisation de la science, tels que ceux de Ledouble et de Botta, que nous citons ici. De la même manière, la vulgarisation est mentionnée dans plusieurs travaux qui traitent de la déterminologisation, comme l'étude de Halskov (2005), par exemple. Par ailleurs, nous montrons au Chapitre 1 que ces deux phénomènes s'intègrent dans la problématique plus large de la circulation des termes et que, dans les deux cas, il est question de l'usage de termes par des non-experts. S'il est clair, dans notre travail, que nous nous focalisons sur la déterminologisation au sens de Meyer et Mackintosh, nous reviendrons néanmoins sur les liens qu'entretiennent ces deux phénomènes, d'abord dans le but de constituer notre objet d'étude, dans

le Chapitre 4 (4.1.1 et 4.2.2.1 en particulier), puis dans le but de réinterroger la nature de ces liens, dans le Chapitre 8.

Dans un contexte différent, signalons également Vogh et Courbon (2018), qui citent le phénomène de déterminologisation précisément dans le but de s'en démarquer. Les auteurs s'interrogent en effet sur la réception et l'appropriation des termes liés au diabète, par des non-experts. S'ils rejettent la perspective de la déterminologisation, c'est en raison de la continuité entre usage spécialisé et usage non spécialisé qu'elle implique, alors que les discours qui les intéressent sont produits par des locuteurs qui, bien qu'ils ne soient pas experts (au sens strict du terme) des questions liées au diabète, sont directement concernés par cette maladie (Vogh et Courbon, 2018 : 86-88). Les auteurs considèrent ainsi que ces locuteurs ne sont pas « passifs » dans la réception des termes liés au diabète, mais participent activement à leur diffusion, ce qui passe par leur appropriation (*Ibid.*). Dans ce contexte, pour eux, la notion de déterminologisation

« ne permet pas véritablement de rendre compte de la complexité des processus en jeu lorsque des termes passent de contextes qui sont spécialisés *dans l'élaboration d'un certain type de savoirs* vers des contextes dans lesquels les besoins cognitifs et interprétatifs sont différents. » (*Ibid.* : 87)

Pour cette raison, les auteurs préfèrent parler d'*appropriation* et de *récupération terminologique*, qui « impliquent la participation active des sujets sans qui les termes n'existeraient pas, ne se diffuseraient pas et ne se transformeraient pas sur le plan sémantique » (*Ibid.* : 88). Notons que cette approche rappelle les travaux en socioterminologie, notamment ceux de Gaudin (1997, 2003 entre autres) et de Delavigne (2001, 2013, 2020), qui traitent de la négociation du sens des termes et de leur validation sociale, et où les locuteurs non experts (ou semi-experts) jouent un rôle actif dans le sens qui est véhiculé dans les discours (cf. Chapitre 1, 1.2.2).

Dans le même ordre d'idée, nous relevons un article de Peynaud (2016), qui s'intéresse au traitement de questions environnementales complexes dans la presse, et en particulier à l'accident nucléaire de Fukushima (Peynaud, 2016 : 209). Partant du principe que l'un des rôles de la presse consiste à transmettre des connaissances spécialisées auprès du grand public (*Ibid.* : 210-211), l'auteur centre son propos particulièrement sur les termes utilisés pour traiter de Fukushima dans la presse. Cette approche l'amène nécessairement à considérer la question de la déterminologisation (*Ibid.* : 213). Par ailleurs, c'est essentiellement à travers la fréquence des termes dans son corpus que l'auteur aborde cette question, les aspects sémantiques (pourtant caractéristiques de l'approche de Meyer, comme nous le disons dans la section 2.1.1.3) n'étant pas interrogés. Nous y reviendrons dans la section 2.3.1.

Ces exemples permettent ainsi de constater que la notion de déterminologisation est exploitée dans différents travaux et dans différents objectifs. Cependant, de manière générale dans ces travaux, la notion en elle-même reste très peu discutée d'un point de vue théorique et les observations effectuées par les auteurs que nous citons n'ont généralement pas pour but de nourrir la réflexion sur la déterminologisation. Nous ne voulons pas dire par là que ces travaux ne sont pas intéressants d'un point de vue théorique, au contraire, seulement que les auteurs cités poursuivent un objectif différent. D'autres travaux permettent en revanche d'enrichir la réflexion sur la déterminologisation. Ils font l'objet de la section suivante.

2.2.2. Trois axes de développement de la notion de déterminologisation

Dans cette section, nous synthétisons plusieurs travaux qui approfondissent la réflexion sur la déterminologisation. Trois aspects ressortent en particulier : l'enrichissement des descriptions des termes dans la langue générale, l'apport de la dimension diachronique pour l'analyse de la déterminologisation et l'intérêt d'analyser d'autres textes que ceux qui relèvent de la langue générale ou d'un degré de spécialisation élevé. Nous reprenons ensuite à notre compte les apports de ces travaux pour élaborer une définition de ce phénomène qui tienne compte de ses différentes facettes et nous mettons en évidence les aspects de la déterminologisation qui peuvent bénéficier d'une étude approfondie.

2.2.2.1. Description affinée du fonctionnement des termes dans un contexte non spécialisé

Selon nous, les descriptions fines du fonctionnement des termes dans la langue générale (à travers la presse généraliste le plus souvent) constituent l'apport principal des travaux qui se focalisent plus en détail sur la question de la déterminologisation. En effet, à partir des deux types de changements sémantiques identifiés par Meyer et Mackintosh (2000a, 2000b), plusieurs auteurs poursuivent la réflexion et enrichissent les descriptions.

Par exemple, Ungureanu, qui vise à démontrer l'interpénétration entre langue générale et langue de spécialité dans le discours d'Internet, décrit les dynamiques sous-tendant les processus de terminologisation, déterminologisation et nomadisation, qui sont responsables de la circulation langagière (Ungureanu, 2006 : 121 ; cf. Chapitre 1, 1.1.1). Elle insiste notamment sur les rapports qu'entretiennent les processus de déterminologisation et de terminologisation, expliquant que « le premier stimule le second, le phénomène entraînant ainsi une polysémie inter-domaines » (*Ibid.*). L'autrice poursuit :

« une unité lexicale déterminologisée peut être utilisée en tant que terme générique qui a perdu sa spécificité dans le domaine. Elle peut acquérir en même temps le statut de mot à la mode. À partir de ce moment-là elle peut être reprise et reterminologisée par une autre langue de spécialité. » (*Ibid.*)

Les contributions de Renouf (2017) se rapprochent de ce point de vue. En effet, à travers la question de la déterminologisation, Renouf s'intéresse surtout au cycle de vie des unités lexicales, ou « life-cycle of a word » (Renouf, 2017 : 27) dans une perspective de néologie. Considérant la déterminologisation comme prenant part aux mécanismes de néologie en langue générale (cf. Chapitre 1, 1.4), l'autrice propose alors de schématiser le cycle de vie des termes en situation de déterminologisation et explique qu'un terme peut se reterminologiser dès lors que son usage est suffisamment répandu en langue générale (*Ibid.*).

À partir de l'observation de plusieurs termes dans un corpus de presse généraliste, Renouf analyse le fonctionnement des termes qui se déterminologisent selon quatre caractéristiques :

- accroissement de la productivité lexicale,
- développement d'un sens métonymique ou métaphorique,
- modifications de la nature grammaticale ou pragmatique,
- accroissement de la créativité (jeux de mots) (*Ibid.*).

Notons que Renouf situe cette étude dans la continuité des propositions originales de Meyer et Mackintosh, à la différence que les données qu'elle possède sont particulièrement volumineuses (1,3 milliard d'occurrences (*Ibid.* : 23)), ce qui lui permet de décrire les fonctionnements de plusieurs termes de manière plus détaillée, à partir d'un plus grand nombre d'occurrences (*Ibid.* : 25-26). À titre d'exemple, citons le cas du terme *irritable bowel syndrome*. Dans un contexte journalistique et dans un but humoristique, ce terme donne en effet lieu à plusieurs jeux de mots, qui consistent à remplacer *bowel* par un autre substantif dans le terme complexe : *irritable mind syndrome*, *irritable desk syndrome*, *irritable bill syndrome*. Ainsi, ce n'est plus le côlon qui est irritable dans ces exemples, mais d'autres objets, d'autres concepts. La nature médicale du terme à l'origine de ces jeux de mots contribue sans doute considérablement à l'effet comique ou ironique. Cet effet transparait par ailleurs également dans des cas où l'analogie va encore plus loin et où le lien avec le terme original est moins évident. Seul le terme *syndrome* est alors maintenu et le jeu de mots se fonde sur la ressemblance entre *irritable* et d'autres adjectifs construits sur le même mode (avec le préfixe *in-* et le suffixe *-able*), par exemple : *interminable meeting syndrome* ou *inexplicable apology syndrome* (*Ibid.* : 40-42).

En outre, bien que les descriptions de Renouf soient particulièrement détaillées, celles-ci se focalisent sur les cas où le fonctionnement des termes diffère dans une large mesure de leur fonctionnement dans leur langue de spécialité « d'origine », c'est-à-dire les cas où les changements sont les plus évidents, répondant ainsi uniquement au second type de changement sémantique identifié par Meyer et Mackintosh. En revanche, la comparaison du comportement des termes entre des données issues de la presse généraliste et des données issues de discours de spécialistes permet d'interroger également le premier type de changement sémantique évoqué par les autrices. C'est ce que propose notamment Dury (2008), dans le domaine de l'écologie.

L'autrice vise ainsi à repérer les termes du domaine de l'écologie qui passent dans la langue générale et à caractériser les changements sémantiques qui se produisent au cours du processus (Dury, 2008 : 231). En outre, et de manière similaire à Renouf ou à Ungureanu, Dury cherche également à saisir et à décrire l'influence de ces changements sémantiques sur la langue de spécialité de l'écologie (*Ibid.*), c'est-à-dire à observer des phénomènes de terminologisation. Par ailleurs, l'autrice aborde ces questions dans un corpus comparable bilingue, en français et en anglais, et cherche à observer des phénomènes parallèles dans les deux langues, pour des paires de termes équivalents. Ainsi, dans un corpus en anglais composé d'un sous-corpus de presse et d'un sous-corpus spécialisé, l'autrice observe que 74 termes de l'écologie extraits du sous-corpus spécialisé sont également attestés dans le sous-corpus de presse (*Ibid.* : 234). L'autrice explique par ailleurs que

« [t]he contexts extracted from the corpus also show that in this case, the original domain sense of the term is unaltered (even though laypersons may understand the concept less deeply than ecologists do [...]). » (*Ibid.* : 234-235)

En outre, Dury approfondit sa réflexion avec l'analyse des termes construits à partir du préfixe *eco-*. L'autrice observe ainsi que ce préfixe, que l'on retrouve dans le terme *ecology*, donne lieu à la création de nouvelles unités dans le corpus de presse en anglais. En effet, dans ce corpus, l'autrice repère 91 unités qui sont construites à partir de ce préfixe. Parmi ces unités, seules 15 sont également attestées dans le corpus spécialisé (*Ibid.* : 234-236). Dans ce cas, étant donné que le préfixe *eco-* permet de créer de nouvelles unités qui ne sont pas nécessairement des termes du domaine de l'écologie⁴⁰, l'autrice considère donc que c'est le préfixe qui se déterminologise⁴¹. Par

⁴⁰ Dans la mesure où 76 unités sur les 91 repérées dans le sous-corpus de presse ne sont pas attestées dans le sous-corpus spécialisé, nous considérons qu'il est très peu probable qu'elles soient utilisées par des experts de ce domaine. Il semblerait donc délicat de les considérer comme des termes de ce domaine, simplement du fait qu'elles sont construites à partir du préfixe *eco-*, bien que celui-ci évoque le domaine de l'écologie (comme le montre Dury (2008 : 236)).

⁴¹ Notons qu'Adelstein (1996) parvient à des conclusions similaires pour l'espagnol, à propos de composants grecs et latins.

ailleurs, l'analyse détaillée des contextes d'apparition de ces nouvelles unités dans le corpus de presse permet à l'auteurice de mettre en évidence deux phénomènes sémantiques : d'une part, le lien avec l'écologie en général et la protection de l'environnement reste visible dans des unités telles que *eco-village*, *eco-tourism* ou *eco-entrepreneur* et, d'autre part, dans des unités telles que *eco-freak*, *eco-shopping* ou *eco-fantasy*, « *eco-* seems to have lost its original ecological sense and to have moved away from its scientific roots » (*Ibid.* : 236). L'auteurice attribue ce changement sémantique plus conséquent au phénomène de déterminologisation.

Comparant ces observations à son corpus en français, Dury constate un phénomène similaire, non seulement avec le préfixe *éco-* (que l'on retrouve dans *écologie*), ce qui dénote donc bien une évolution parallèle en français et en anglais (*Ibid.* : 239-241), mais également avec l'unité *bio*,

« used either as a prefix, like *éco-*, as in *bio-attitude*, *bio-matériaux*, *bio-combustible*, etc., or as an adjective, as in *agriculture bio*, *ligne bio*, *yaourt bio*, etc. It can also be used as a noun, as in *le boum du bio*, *coup de frein sur le bio*, *le bio se vend bien*, etc. » (*Ibid.* : 241)

L'auteurice souligne par ailleurs la productivité de cette unité, avec plus de 100 unités relevées dans son sous-corpus de presse en français (*Ibid.*).

Ainsi, selon nous, l'étude de Dury enrichit les réflexions sur la déterminologisation de deux manières au moins : d'une part, elle étaye les observations de Meyer et Mackintosh quant aux deux types de changements sémantiques susceptibles de se produire au cours du processus de déterminologisation et, d'autre part, elle montre qu'une fois qu'un terme est intégré à la langue générale, il peut servir à la création de nouvelles unités, dénotant ou non le domaine d'origine. Autrement dit, la déterminologisation peut concerner non seulement des termes, mais également des morphèmes.

Citons enfin les travaux de Condamines et Picton (2014a, 2014b), qui s'intéressent particulièrement à la question de la déterminologisation dans le domaine spatial. Dans leur cas, l'approche contrastive entre langue générale et langue de spécialité⁴² permet de mettre au jour un certain nombre de différences de fonctionnement, que les auteurices relient à la déterminologisation. Par exemple, les auteurices s'intéressent à la morphologie des termes complexes et comparent notamment les termes pour lesquels une alternance entre deux modes de formation peut être constatée. En particulier, les termes complexes de type nom + nom et de type nom + préposition

⁴² À des degrés de spécialisation différents, c'est-à-dire à partir de textes produits par des experts pour des semi-experts et des experts du même domaine (Condamines et Picton, 2014b : 699) ou de textes produits par des experts pour des semi-experts ou des non-experts (Condamines et Picton, 2014a : 170-171). Nous y revenons en 2.2.3.3.

+ nom sont considérés, comme *observation satellite* et *observation par satellite* ou *localisation satellite* et *localisation par satellite* (Condamines et Picton, 2014a : 181-182). À partir de la répartition de ces termes dans leurs données, les autrices expliquent alors que

« [l']ellipse du déterminant et de la préposition est souvent considérée comme un indice de terminologisation (Collet, 1997 ; Portelance, 1991). [...] Mais aucune étude n'a, à notre connaissance, étudié le processus inverse, de rétablissement du déterminant et de la préposition, dans les cas de déterminologisation. Or, il semble bien que, de manière globale, lors du passage des termes dans le discours général, les prépositions soient rétablies. » (Ibid. : 182)

Par ailleurs, les autrices interrogent également les aspects sémantiques liés au processus de déterminologisation et se focalisent particulièrement sur le cas du terme *navigation* et des différents termes complexes construits sur cette base (Ibid. : 182-183). En effet, constatant la forte diffusion du terme *navigation satellite* dans leur corpus de presse, les autrices repèrent également le syntagme *navigation terrestre*, qui est, selon elles, probablement une forme réduite de *navigation des véhicules terrestres*. Les autrices supposent alors que l'émergence du syntagme *navigation terrestre* est influencée non seulement par la diffusion du terme *navigation satellite* « à travers le développement et la banalisation des GPS », mais également par le paradigme créé par les autres types de navigation, avec des syntagmes tels que *navigation maritime* et *navigation aérienne* (Ibid.). À l'issue de ces quelques observations, les autrices concluent à une augmentation de « la couverture sémantique de *navigation* dont le sens pourrait être, dans le contexte spatial : guidage par satellite de tout objet se déplaçant dans l'air, l'espace ou sur terre. » (Ibid.)

En bref, il semble que nous pouvons retenir essentiellement deux éléments de ces travaux, pour ce qui est de la description du fonctionnement des termes en situation de déterminologisation. Premièrement, les différentes études que nous citons mettent en évidence le fonctionnement diversifié des termes dans la presse généraliste. Certaines observations vont dans le même sens que les descriptions de Meyer et Mackintosh (2000a, 2000b), par exemple un degré plus faible de compréhension des termes par des non-experts, d'après les observations de Dury, ou les différentes créations lexicales basées sur un terme et recherchant un effet comique d'après Renouf. Mais dans ce dernier cas, l'analyse est soutenue par des données quantitativement plus importantes, ce qui rend possible la description d'un plus grand nombre d'exemples et peut ainsi renforcer les conclusions de l'autrice.

Dans d'autres cas, cependant, ces différents travaux permettent d'ajouter de nouvelles caractéristiques au fonctionnement des termes dans un contexte de déterminologisation et de donner un nouvel éclairage sur ce phénomène. Il nous semble que c'est particulièrement le cas de

l'alternance du mode de construction des termes complexes telle que décrite par Condamines et Picton et de la créativité d'un morphème en particulier une fois qu'un terme est intégré à la langue générale.

Deuxièmement, l'adoption d'une approche contrastive entre langue de spécialité et langue générale, ou du moins entre deux corpus dont les textes relèvent de degrés de spécialisation différents, permet d'identifier et de caractériser des différences de fonctionnement plus subtiles qu'avec une approche impliquant uniquement un corpus non spécialisé. En effet, il semble que, sans point de comparaison, seuls des changements conséquents peuvent être caractérisés, c'est-à-dire lorsque les termes fonctionnent de manière très différente de ce que l'on pourrait attendre en langue de spécialité. Nous faisons ici référence essentiellement aux changements sémantiques profonds et aux usages humoristiques décrits par Renouf (2017). Au contraire, Dury, par exemple, observe bien des fonctionnements qui relèvent du premier type de changement sémantique identifié par Meyer et Mackintosh.

En outre, il semble qu'une approche contrastive offre également la possibilité de faire émerger certains phénomènes des différentes données comparées, qui n'auraient peut-être pas été repérés ou caractérisés de la même manière avec une approche impliquant un seul corpus. Nous pensons en particulier à l'étude de Dury, dans laquelle l'autrice a montré l'apparition de nouvelles unités « pseudo terminologiques », notamment les unités créées à partir du préfixe *eco-*. Il nous semble que, dans ce cas, seule une approche contrastive entre langue générale et langue de spécialité permet de démontrer que ces unités sont caractéristiques de la langue générale (ou du moins de la presse généraliste dans le cas de Dury), précisément du fait qu'elles ne sont pas attestées dans le corpus de langue de spécialité.

Ces travaux enrichissent donc les descriptions des termes en situation de déterminologisation aussi bien grâce à la taille des données analysées (dans le cas de Renouf) que grâce à la complémentarité des différents types de données explorées (dans le cas de Dury et de Condamines et Picton). Mais ces travaux alimentent les descriptions du phénomène de déterminologisation à d'autres égards également, notamment à travers la prise en compte de la dimension diachronique.

2.2.2.2. Prise en compte de la dimension diachronique

Comme nous l'expliquons *infra* (section 2.3.2.1), le processus de déterminologisation s'inscrit nécessairement dans le temps et la prise en compte de la dimension diachronique dans l'analyse permet d'apporter un nouvel éclairage sur ce phénomène. Ainsi, bien que les travaux de

Meyer et ses collègues se situent essentiellement dans une perspective synchronique, d'autres travaux, tels que ceux de Dury (2008), Halskov (2005), Condamines et Picton (2014a) et Renouf (2017), interrogent la perspective diachronique et démontrent son intérêt pour la déterminologisation. Par exemple, Dury montre que la prise en compte de cette dimension permet d'appréhender de manière plus détaillée l'évolution du fonctionnement des termes dans le contexte de la déterminologisation, et plus particulièrement dans un corpus de presse généraliste.

En effet, dans un corpus qui couvre une période de 12 ans et qui est découpé en tranches successives de deux ans (Dury, 2008 : 233-234), l'auteur retrace de manière détaillée l'évolution du préfixe *eco-* en anglais et de l'unité *bio* en français. Elle montre ainsi que les unités construites avec le préfixe *eco-* s'éloignent progressivement du domaine de l'écologie, en particulier les unités qui apparaissent dans le corpus après 2003. L'auteur constate ainsi l'évolution de ce préfixe dans la presse, où il sert à créer d'abord des unités dont le sens reste lié au domaine de l'écologie et de la protection de l'environnement, puis, à partir de 2003, des unités dont le sens s'éloigne dans une plus large mesure de ce domaine (*Ibid.* : 235-236). De la même manière, dans son corpus en français, Dury montre que la déterminologisation de *bio* s'accompagne, dans le temps, d'un accroissement progressif de la fréquence des unités construites sur cette base, passant de 97 occurrences dans la première période de deux ans du corpus (1996-1997) à 404 dans la dernière (2006-2007) (*Ibid.* : 241). Notons que Renouf parvient à des constats similaires en ce qui concerne la productivité des termes lorsqu'ils fonctionnent dans la presse (Renouf, 2017 : 27).

Outre ces observations de nature quantitative, Dury caractérise également l'évolution du sens des termes dans la presse à partir de son corpus diachronique. Par exemple, dans le cas de *bio* en français, l'auteur constate que

« [i]n the first time segments of the corpus ([1996-1997] to [2002-2003]), *bio* is used as the shortened form of *biologique* and applies mainly to food products which are considered healthy and which were cultivated or bred without the help of chemicals or non-natural products [...]. » (Dury, 2008 : 242)

En revanche, les occurrences de *bio* dans les périodes plus récentes du corpus (de 2004 à 2007) dénotent une extension de sens de *bio*, qui modifie non plus uniquement des unités en rapport avec l'alimentation, mais désormais également des unités en rapport avec l'industrie textile ou cosmétique, et même « anything that is nature-oriented and environmentally friendly » (*Ibid.*).

Il semble par ailleurs que ces extensions de sens ne soient pas caractéristiques du domaine étudié par Dury, mais bien des conséquences de la déterminologisation sur le fonctionnement des termes de manière plus générale. En effet, Renouf observe des phénomènes similaires dans un corpus diachronique qui est aussi composé d'articles de presse et qui couvre 32 ans (Renouf,

2017 : 26). Dans ce corpus, l'auteur observe le cas du terme *to downsize*, dont l'usage, originellement réservé au domaine de l'industrie automobile, s'étend à la langue générale (*Ibid.* : 35). Elle explique ainsi que, dans ce domaine, le sens de *to downsize* correspond à « to design or build (a car) of smaller overall dimensions » et à « to meet government and market demands for cars that are lighter and more economical », mais qu'il peut également prendre un autre sens par métaphore, comme synonyme de *renvoyer* ou *licencier* (*Ibid.*). Précisons que l'auteur s'appuie ici sur le dictionnaire Oxford English Dictionary de 2017. Ce nouveau sens par métaphore est donc suffisamment répandu et stable pour être décrit dans un dictionnaire général.

Renouf compare les données de son corpus avec la description de *to downsize* dans le Oxford English Dictionary et détaille son évolution dans le temps. L'auteur repère ainsi plusieurs néologismes formés par dérivation ou par conversion, à partir de ce sens métaphorique de *to downsize* : *downsizer*, pour désigner la personne qui licencie (qui *downsizes*) ainsi que *downsizing* et *downsize* (forme nominale) pour désigner des licenciements massifs au sein d'une entreprise (*Ibid.* : 36).

Parallèlement à cette évolution, Renouf observe un changement dans la distribution de *to downsize*, qui met en évidence une construction intransitive du verbe, alors qu'il est transitif dans les cas que nous venons de décrire (*Ibid.*). Ce changement met alors en évidence l'émergence d'un nouveau sens et le verbe désigne dans ce cas le fait de déménager dans une maison plus petite, d'abord, dans le contexte de personnes à la retraite dès 1999, puis, dès 2013, dans un contexte plus restreint de personnes au bénéfice d'aides sociales qui seraient pénalisées financièrement par le fait d'occuper un logement plus grand que ce qui leur est nécessaire (*Ibid.* : 36-37). Notons le facteur culturel de ce dernier usage, qui pourrait être propre au Royaume-Uni et ne pas se retrouver dans tous les contextes anglophones.

Bien que Renouf parvienne à des descriptions détaillées de l'évolution du fonctionnement de plusieurs termes dans un corpus diachronique très volumineux, Halskov (2005) montre qu'il est possible de saisir certaines évolutions également avec un corpus diachronique couvrant une période de quelques années uniquement. Ainsi, à partir d'une comparaison entre un corpus spécialisé et un corpus de presse, l'auteur souligne deux tendances en diachronie : d'une part, la stabilité de certaines cooccurrences de type nom-nom dans le corpus de presse et à travers toute la période prise en compte, notamment lorsque le terme *server* est modifié par les noms *computer*, *PC*, *e-mail*, *mail*, *Internet*, et, d'autre part, l'apparition ou la disparition de certaines cooccurrences, par exemple *application server* dans le premier cas ou *database server* et *e-commerce server* dans le second cas (*Ibid.* : 57-60).

L'auteur conclut sur le fait que la cooccurrence du terme *server* avec des adjectifs qu'il qualifie de conceptuellement vagues dans le corpus de presse, en comparaison avec le corpus spécialisé, renvoie au flou conceptuel qui peut apparaître lorsque les termes passent dans la langue générale (*Ibid.* : 58). Paradoxalement, à travers la fréquence élevée de *computer server* dans le corpus de presse et de son absence du corpus spécialisé, l'auteur suggère que l'introduction de cette précision répond à un besoin de distinction, dans une situation non spécialisée, entre *server* dans le domaine de l'informatique et *server* dans le domaine de la restauration (*Ibid.*).

Les conclusions de Halskov, de même que celles de Renouf, se fondent sur une analyse en diachronie dans un corpus découpé par année. C'est également ce choix de découpage qu'opèrent Condamines et Picton (2014a), dans un but différent cependant. En effet, c'est essentiellement dans le but de « saisir la manière dont les technologies spatiales se diffusent auprès du grand public » (Condamines et Picton, 2014a : 176) que les autrices observent l'évolution diachronique de la fréquence de plusieurs termes du domaine spatial. Dans ce contexte, les autrices se focalisent sur la diffusion des termes désignant des types de satellites artificiels dans un corpus de presse généraliste représentant une période de 14 ans. Précisons que les types de satellites qui les intéressent sont classés par leur taille, soit les termes *mini-satellite*, *micro-satellite*, *nano-satellite* et *pico-satellite*.

L'analyse de la fréquence et de la répartition de ces termes dans leur corpus permet ainsi aux autrices de dresser un parallèle entre l'actualité du domaine et la diffusion de certains termes dans la presse, notamment *micro-satellite*, pour lequel des pics de fréquence sont repérés en 1998, 2000 et 2004 (*Ibid.* : 178). Condamines et Picton expliquent alors que

« ces pics sont dus à des actualités fortes dans le paysage spatial. Ainsi, depuis fin 1998, le Cnes développe sa première filière de micro-satellites, appelée Myriade, qui s'accompagne notamment d'appels à projets et de rassemblements scientifiques importants jusqu'en 2000. Ces actualités se répercutent alors immédiatement dans les corpus de presse. » (*Ibid.*)

Ce parallèle entre l'actualité d'un domaine et la diffusion de termes liés dans la presse est un aspect important de la déterminologisation. Nous y reviendrons au Chapitre 4 (4.2.2). Cependant, la démarche d'analyse de la déterminologisation proposée par Condamines et Picton nous semble particulièrement intéressante sur un autre point. En effet, contrairement à Dury, par exemple, qui compare le fonctionnement de termes dans un corpus de presse et dans un corpus spécialisé contenant des textes relevant d'un degré élevé de spécialisation (Dury, 2008 : 233-234), les autrices prennent comme point de comparaison un corpus composé de communiqués de presse

(Condamines et Picton, 2014a : 170-171). Nous expliquons l'intérêt de cette approche dans la section suivante.

2.2.2.3. Approche de la continuité de la déterminologisation, entre langue de spécialité et langue générale

Le choix de Condamines et Picton d'intégrer des communiqués de presse à leur étude repose sur l'hypothèse que « dans une perspective de diffusion des connaissances, au moins une autre communauté de locuteurs peut intervenir entre les experts et les journalistes » (Condamines et Picton, 2014a : 171). De ce point de vue, les communiqués de presse joueraient un rôle d'intermédiaire dans la transmission des informations des experts aux journalistes et, par extension, dans la diffusion des termes auprès du grand public.

Le terme *communiqué de presse* est défini dans Termium comme un « [d]ocument destiné aux médias, qui sert généralement à faire des annonces importantes ou à donner de l'information sur les politiques, les programmes ou les activités »⁴³ et dans le Grand dictionnaire terminologique comme un « [a]vis, souvent court, dont le contenu révèle une information précise, qui est destiné à être diffusé dans les médias »⁴⁴. Notons que ces deux définitions mettent en exergue le fait que les communiqués de presse contiennent certaines informations importantes et précises, qui sont destinées aux médias pour une diffusion auprès d'un public plus large, caractéristiques sur lesquelles se basent Condamines et Picton. De ce point de vue, les communiqués de presse sont donc des textes qui relèvent, d'après la classification de Bowker et Pearson (2002 : 28), d'une communication de type expert à semi-expert ou à non-expert, selon que l'on considère les journalistes comme des non-experts ou des semi-experts.

Dans leur étude, Condamines et Picton observent l'évolution des termes non seulement dans la presse généraliste, mais également dans des communiqués de presse, cherchant ainsi à identifier un lien entre les fonctionnements constatés dans la presse et dans les communiqués. Leurs observations confirment partiellement cette hypothèse : si les autrices repèrent effectivement plusieurs similitudes dans l'évolution de la fréquence et de la répartition du terme *micro-satellite*, notamment dans les communiqués de presse et dans la presse (Condamines et Picton, 2014a : 178), tous les phénomènes observés dans leur corpus de presse ne peuvent pas être mis en relation avec des phénomènes parallèles dans leur corpus de communiqués de presse. Les autrices reconnaissent

⁴³ Termium, <https://www.btb.termiumplus.gc.ca> (page consultée le 22 février 2021).

⁴⁴ Grand dictionnaire terminologique, <http://gdt.oqlf.gouv.qc.ca> (page consultée le 22 février 2021).

ainsi que « les informations spatiales, et donc les termes, peuvent être diffusées dans la presse à partir d'autres médias » (*Ibid.*).

Cette remarque renvoie par ailleurs à ce que nous détaillons dans le Chapitre 1 (1.2), lorsque nous montrons que les termes peuvent être diffusés auprès du grand public – et donc potentiellement intégrer la langue générale – par différents biais, notamment les discours de transmission des connaissances. Soulignons enfin qu'à notre connaissance, l'étude de Condamines et Picton est la seule, avec celle de Halskov (2005)⁴⁵, à tenir compte de textes qui ne relèvent pas d'un degré de spécialisation correspondant au type de communication strictement entre experts. Pourtant, comme nous l'expliquons *infra* et dans le Chapitre 4, cette dimension est cruciale dans l'observation du processus de déterminologisation.

Ces différents travaux mettent ainsi en lumière la complexité du phénomène de déterminologisation. Ces trois axes de description permettent alors de saisir cette complexité et de distinguer trois aspects complémentaires de la définition de ce phénomène. Dans la section suivante, nous tirons parti de ces développements pour proposer une définition actualisée de la déterminologisation.

2.3. Pour une double définition de la déterminologisation

À la lueur des éléments discutés dans la section précédente, nous proposons de définir la déterminologisation à la fois comme processus et comme résultat. Au vu de ce qui précède, il semble par ailleurs que ce processus implique deux dimensions, une dimension temporelle et une dimension qui répond au continuum langue de spécialité-langue générale.

2.3.1. La déterminologisation comme un résultat

Le fonctionnement des termes dans la langue générale, qui diffère toujours dans une certaine mesure du fonctionnement en langue de spécialité (d'un point de vue sémantique, syntaxique, etc.), est le point de vue le plus souvent adopté dans la littérature. C'est précisément cet aspect que nous considérons comme le résultat du processus de déterminologisation. Les travaux qui adoptent une telle approche analysent le plus souvent des articles de presse généraliste (par exemple Meyer et Mackintosh, 2000a, 2000b ; Renouf, 2017) et procèdent parfois à des

⁴⁵ Cependant, bien que Halskov considère que, « to gain a better understanding of the linguistic properties of conceptual fuzziness » (ce qui est, selon lui, caractéristique de la déterminologisation), « the dimension of text type should also be examined » (Halskov, 2005 : 54) et bien qu'il affirme explorer un corpus constitué entre autres de textes de vulgarisation scientifique et de discussions de forums techniques (*Ibid.*), l'auteur ne semble pas interroger précisément cette dimension dans l'analyse du terme *server* qu'il détaille (*Ibid.* : 57-60).

comparaisons avec des textes qui relèvent d'une communication de type expert-expert, selon Bowker et Pearson (2002 : 28) (par exemple Halskov, 2005 ; Dury, 2008 ; Condamines et Picton, 2014b ; cf. 2.2.3).

Cette idée de résultat d'un processus est suggérée en particulier dans les travaux de Meyer et ses collègues (Meyer *et al.*, 1998 ; Meyer, 2000 ; Meyer et Mackintosh, 2000a, 2000b), qui utilisent le participe passé du verbe *to de-terminologize* pour qualifier les termes qui ont intégré la langue générale, c'est-à-dire qui se sont déterminologisés. Dans leurs travaux, nous constatons ainsi plusieurs syntagmes pour désigner ces termes et leurs usages dans la langue générale, par exemple *de-terminologized word*, *de-terminologized lexical item*, *de-terminologized term*, *de-terminologized use*, *de-terminologized usage*, *de-terminologized meaning* en anglais et *mot déterminologisé*, *emploi déterminologisé*, *sens déterminologisé* en français. Notons par ailleurs que ces exemples mettent l'accent sur les aspects sémantiques de la déterminologisation (ici avec les unités *use*, *usage*, *meaning*, *emploi* et *sens*), qui est l'orientation principale des travaux de Meyer et ses collègues.

Parallèlement, il nous semble que considérer qu'un terme est déterminologisé, ou qu'il existe un sens déterminologisé pour un terme, suppose différentes étapes dans le processus de déterminologisation, ou du moins qu'il existe un stade à partir duquel il est possible de constater un état déterminologisé par rapport à un état non déterminologisé. En outre, cet état déterminologisé amène inévitablement à interroger ce qui différencie précisément un terme déterminologisé d'un terme qui n'est pas déterminologisé. Par exemple, l'attestation d'un terme dans un texte non spécialisé, c'est-à-dire dans un contexte général, suffit-elle à considérer que ce terme est déterminologisé ?

Si la question de la fréquence n'est pas directement abordée par Meyer et Mackintosh, les autrices suggèrent tout de même que, dès lors qu'un terme fonctionne dans la langue générale, même lorsque « l'essence du sens terminologique demeure » (Meyer et Mackintosh, 2000a : 202), c'est-à-dire lorsque les changements qui se sont produits au cours du processus concernent le degré de compréhension des concepts par les non-experts, ce terme est considéré comme déterminologisé. C'est ce que nous déduisons par exemple de la citation suivante, où le syntagme *de-terminologized word* est mis en relation avec ce type de changement sémantique :

« [d]espite the various changes discussed above, all the de-terminologized words in this category designate the same fundamental concept in general language as their equivalent terms do in specialized discourse » (Meyer et Mackintosh, 2000b : 115)

De plus, les autrices précisent que, si elles considèrent les termes utilisés en langue générale comme déterminologisés, c'est parce que ces termes « undergo minor conceptual shifts as they

cross the border between specialized and general language, and are therefore not fully *terminological* in the strictest sense of this term » (*Ibid.* : 114). De ce point de vue, il semble que ce qui permet de qualifier un terme de déterminologisé relève essentiellement du fait que ce terme, lorsqu'il est employé par des non-experts dans un contexte général, l'est toujours d'une manière qui diffère de celle des experts dans un contexte spécialisé.

En revanche, dans d'autres travaux, la fréquence d'un terme dans des textes non spécialisés semble jouer un rôle déterminant dans le « degré » de déterminologisation de ce terme. C'est du moins ce que suggère Peynaud (2016), qui postule un lien entre la fréquence d'un terme dans la presse et son « degré » de déterminologisation, voire assimile la déterminologisation à la seule fréquence d'un terme dans la presse :

« whether the terms are more or less common in the press in general, **that is to say more or less determinologised**, does not seem to impact the extent to which they are used in the corpus. » (Peynaud, 2016 : 218)⁴⁶

D'après cette citation, il semble bien que l'autrice considère que le « degré » de déterminologisation d'un terme pourrait se mesurer à l'aide de sa fréquence dans la presse de manière générale, ce qui suppose donc que plus un terme est fréquent dans la presse, plus il est déterminologisé. Ce point de vue soulève cependant plusieurs questions. En effet, si le « degré » de déterminologisation d'un terme ne se mesure qu'à travers la fréquence de ce terme dans la presse, différents aspects doivent être précisés, notamment le type ou les types de presse dont il faut tenir compte (presse généraliste uniquement ou non, titres de presse particuliers, etc.), la taille du corpus dans lequel calculer cette fréquence (pour autant qu'une approche en corpus soit adoptée), le seuil de fréquence à partir duquel un terme peut commencer à être considéré comme déterminologisé, un ou des points de comparaison, vu que la fréquence absolue seule n'est probablement pas suffisante (et que toutes les unités d'une langue n'apparaissent pas à la même fréquence dans les textes), etc.

Bien que ces questions restent en suspens, l'importance accordée par Peynaud à la fréquence des termes dans la presse comme facteur déterminant le « degré » de déterminologisation, indépendamment des changements sémantiques éventuels, nous amène à réfléchir sur la possibilité (ou la nécessité) d'une distinction entre le processus d'intégration des termes dans la langue générale en tant que nouvelles unités (ce qui se rapproche des travaux qui abordent la diffusion des termes dans la langue générale du point de vue de la néologie, cf. Chapitre 1, 1.4) et l'évolution du sens des termes lorsqu'ils fonctionnent dans la langue générale. Ce dernier

⁴⁶ Nous soulignons.

point renvoie cependant non plus au résultat de la déterminologisation, mais bien au processus ; nous le développons particulièrement en 2.3.2.1.

2.3.2. La déterminologisation comme un processus

De manière générale, il semble reconnu que la déterminologisation est un processus de passage de termes d'une langue de spécialité vers la langue générale. Les auteurs que nous citons dans ce chapitre associent par ailleurs explicitement la déterminologisation à un processus (par exemple Meyer, 2000 ; Meyer et Mackintosh, 2000a, 2000b ; Halskov, 2005 ; Bertaccini *et al.*, 2008 ; Dury, 2008 ; Cabré, 2016 ; Renouf, 2017 ; L'Homme, 2020). Cependant, en dépit de ce consensus, cette idée de processus n'est le plus souvent pas détaillée davantage et est affirmée comme une évidence. Bien que nous partagions cet avis, il semble que plusieurs aspects interviennent dans ce processus et doivent être précisés, d'abord en relation avec la dimension temporelle de la déterminologisation, puis en relation avec le continuum entre langue de spécialité et langue générale.

2.3.2.1. Première dimension : le rôle du temps

La première dimension est la dimension temporelle. En effet, le processus de déterminologisation s'inscrit nécessairement dans le temps, comme tout autre processus d'ailleurs. Nous retenons notamment cet aspect des définitions de l'unité *processus* dans le Petit Robert en ligne et dans le Trésor de la langue française informatisé (TLFi). Ainsi, dans le Petit Robert, *processus* est défini comme un « [e]nsemble de phénomènes, conçu comme actif et organisé dans le temps »⁴⁷ et, dans le TLFi, comme une « [s]uite continue de faits, de phénomènes présentant une certaine unité ou une certaine régularité dans leur déroulement »⁴⁸. La dimension temporelle apparaît explicitement dans la définition du Petit Robert et elle est sous-entendue dans celle du TLFi. De plus, plusieurs travaux, que nous détaillons *supra*, démontrent largement l'intérêt de la prise en compte de la dimension diachronique dans l'étude de la déterminologisation (en particulier Dury, 2008 ; Condamines et Picton, 2014a ; Renouf, 2017), ce qui confirme son importance dans ce processus.

De même, si, à l'instar de Halskov (2005 : 54), nous considérons que la déterminologisation est un aspect de l'évolution de la langue et du changement sémantique, alors il semble évident que la déterminologisation doit s'envisager dans le temps, de la même manière que le changement

⁴⁷ Le Petit Robert en ligne, <https://petitrobert.lerobert.com/robert.asp> (page consultée le 3 mars 2021).

⁴⁸ Le Trésor de la langue française informatisé, <http://stella.atilf.fr> (page consultée le 3 mars 2021).

sémantique ne peut être appréhendé qu'à travers le temps (Guilbert, 1975 ; Picton, 2009 ; Condamines *et al.*, 2012 ; Reutenauer, 2012b ; Sablayrolles, 2012 ; Boussidan, 2013).

S'il est indéniable que la déterminologisation est un processus qui s'inscrit dans le temps, le rôle du temps dans la diffusion des termes auprès de locuteurs non experts et dans les changements sémantiques qui se produisent au cours du processus reste cependant peu interrogé. En outre, bien que nous n'abordions pas cette dimension dans notre thèse, nous pensons que l'approche de la déterminologisation que nous mettons en œuvre permettra d'ouvrir de nouvelles perspectives d'analyse en diachronie⁴⁹, afin de mieux comprendre ce phénomène dans sa globalité. C'est également dans ce but que nous nous focalisons sur la deuxième dimension du processus de déterminologisation.

2.3.2.2. Deuxième dimension : le continuum entre langue de spécialité et langue générale

La deuxième dimension du processus de déterminologisation est le continuum entre les langues de spécialité et la langue générale. En effet, au Chapitre 1 (1.1.2), nous montrons que les langues de spécialité et la langue générale ne forment pas deux ensembles fermés l'un à l'autre, mais qu'au contraire, il existe une continuité entre ces deux extrêmes. Nous pouvons considérer que cette continuité se retrouve dans le processus de déterminologisation pour plusieurs raisons.

Dans un premier temps, nous avons vu avec la définition de l'unité *processus* dans le TLFi que l'idée de continuité est inhérente à la notion de processus. Ainsi, si la déterminologisation désigne le processus de passage de termes d'une langue de spécialité à la langue générale et que ce processus est nécessairement caractérisé par une certaine continuité, alors il semble évident que le passage de termes s'effectue à travers le continuum langue de spécialité-langue générale.

Dans un deuxième temps, nous avons montré au Chapitre 1⁵⁰ que le transfert des termes vers la langue générale peut s'effectuer par le biais de différents textes qui se situent à l'intérieur du continuum langue de spécialité-langue générale. L'intérêt de tenir compte de ces textes est particulièrement mis en évidence par Halskov (2005) et par Condamines et Picton (2014a), ce qui leur permet d'enrichir l'analyse et de nourrir la réflexion à propos de la déterminologisation. Dans ce contexte, ces travaux participent également à définir la déterminologisation comme un processus de passage de termes dans le continuum entre les langues de spécialité et la langue générale.

⁴⁹ Dont certaines sont initiées en parallèle de notre thèse, par exemple dans Picton *et al.* (2021).

⁵⁰ Essentiellement à travers la notion de discours de transmission des connaissances (cf. Chapitre 1, 1.2 et 1.5).

Soulignons néanmoins que ces travaux sont les seuls, à notre connaissance, à interroger cette dimension dans le processus de déterminologisation ; c'est pourquoi il nous paraît nécessaire de poursuivre dans cette direction et d'approfondir ces réflexions.

2.4. Bilan

Dans ce chapitre, nous avons cherché à définir de manière extensive la notion de déterminologisation. Nous avons d'abord retracé les origines de cette notion à travers les travaux de Meyer et ses collègues et nous avons explicité les rapports qu'elle entretient avec d'autres notions proches. Nous nous sommes ensuite focalisée sur les travaux qui ont repris cette notion et qui ont permis de la faire progresser, affinant ainsi les descriptions existantes de ce phénomène. Enfin, nous avons tiré profit de ces différents éléments pour proposer une définition actualisée de la déterminologisation, qui s'articule autour de trois aspects. De ce point de vue, la déterminologisation désigne le processus d'intégration de termes dans la langue générale, à travers le temps et à travers le continuum entre langue de spécialité et langue générale, ainsi que le résultat de ce processus, c'est-à-dire le fonctionnement de termes dans la langue générale.

En outre, la synthèse des différents travaux abordant ce phénomène a montré que la déterminologisation entendue comme un processus de passage de termes dans la langue générale reste une perspective très peu interrogée dans la littérature. Sa prise en compte semble pourtant cruciale pour mieux définir toutes les facettes de ce phénomène. C'est pourquoi, dans notre travail de thèse, nous faisons le choix de nous concentrer particulièrement sur cet aspect. Nous cherchons ainsi à mieux comprendre comment ce processus se réalise dans les textes à travers une analyse de termes du domaine de la physique des particules dans des textes qui se situent aux extrémités et à l'intérieur du continuum langue de spécialité-langue générale. Dans ce contexte, nous cherchons à saisir les dynamiques qui sous-tendent le passage de termes vers la langue générale et les changements de sens qui se produisent au cours du processus. Dans ce but, nous interrogeons également le rôle que les textes se situant à l'« intermédiaire » entre langue de spécialité et langue générale peuvent jouer dans le processus de déterminologisation, principalement en ce qui concerne les changements sémantiques des termes.

Cette première partie a permis de poser les bases théoriques de notre réflexion à propos de la déterminologisation. L'étude de ce phénomène complexe exige par ailleurs la mise en œuvre d'un cadre méthodologique solide, qui permette non seulement de tenir compte de cette complexité, mais également d'interroger les aspects de la déterminologisation sur lesquels nous nous focalisons

dans notre thèse. La construction de ce cadre méthodologique fait l'objet de la partie suivante : dans le Chapitre 3, nous insistons sur l'intérêt d'adopter une approche outillée en corpus pour notre étude, puis nous définissons un cadre d'analyse adéquat pour notre objectif. Dans le Chapitre 4, nous explicitons les différents choix méthodologiques qui nous permettent de construire notre objet d'étude dans un objectif de systématisme et de reproductibilité, puis, dans le Chapitre 5, nous détaillons la définition et la mise en œuvre des deux approches de la distribution des termes que nous exploitons pour notre analyse.

Deuxième partie

Construire une méthodologie d'analyse
outillée en corpus

Chapitre 3 Outiller l'analyse de la déterminologisation

Dans notre travail de thèse, nous tirons profit des acquis de la linguistique de corpus pour mettre en œuvre une démarche d'analyse outillée, afin d'aborder les termes et leur fonctionnement dans les textes. Ce type de démarche comprend plusieurs avantages pour la réalisation de notre étude, notamment la possibilité d'analyser des données réelles d'une certaine complexité, puisque nous cherchons à tenir compte du continuum langue de spécialité-langue générale dans notre analyse de la déterminologisation, ainsi que la possibilité d'accompagner l'analyse par des outils pertinents et performants, afin non seulement de gérer ces données et de repérer des phénomènes ou des régularités qui ne seraient pas observables « à l'œil nu », mais également de tendre vers une démarche d'analyse reproductible et objective. Nous justifions ce parti pris méthodologique et précisons le contexte de la linguistique outillée dans la section 3.1.

Dans le même temps, puisque nous focalisons notre travail sur l'analyse des termes dans le processus de déterminologisation, les spécificités des approches en terminologie doivent être prises en compte dans la méthode globale d'analyse outillée. Aussi, dans la section 3.2 nous détaillons d'abord les développements théoriques de la terminologie qui prônent un rapprochement avec la linguistique et qui se servent des méthodes de la linguistique outillée pour interroger des problématiques variées sur les questions liées aux usages réels des termes dans le discours, puis nous exposons les propositions de la terminologie textuelle, dans laquelle nous ancrons plus particulièrement notre travail. Enfin, dans la section 3.3, nous explicitons le cadre d'analyse que nous construisons pour aborder la question de la déterminologisation, par le biais du fonctionnement des termes dans des données textuelles.

3.1. Pour une démarche en linguistique outillée

Les approches en linguistique outillée sont caractérisées par l'utilisation d'un corpus, entendu comme « a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language » (Sinclair, 1994 : 2, cité par Pearson, 1998 : 42), et par la mise en œuvre de plusieurs outils, afin de guider l'analyse du corpus. L'intérêt d'adopter ce type d'approches est multiple, qu'il s'agisse de la possibilité d'exploiter des données réelles, et donc d'appuyer l'analyse sur des bases objectives, ou de la possibilité de développer une démarche systématique et reproductible, par le recours à un cadre méthodologique balisé par différents outils. L'utilisation éclairée de plusieurs outils pour accompagner l'analyste dans l'exploration d'un corpus textuel constitué pour un objectif précis

permet en effet de tendre vers une démarche dans laquelle la subjectivité parfois associée à l'analyse linguistique est balisée par les résultats fournis par différents outils et par différentes méthodes statistiques. Mais l'intérêt de mettre en œuvre une démarche de linguistique outillée réside aussi – et surtout – dans la possibilité d'accéder à des phénomènes et des questionnements qui ne seraient peut-être pas accessibles d'une autre manière, et donc d'interroger certaines problématiques précises, dont la question du processus de déterminologisation dans sa dimension continue entre langue de spécialité et langue générale.

Dans ce contexte, afin de montrer l'intérêt des approches linguistiques outillées pour l'analyse du processus de déterminologisation, nous situons d'abord les méthodes de linguistique outillée par rapport à la linguistique de corpus, puis nous dressons un aperçu de la diversité des outils et méthodes disponibles et mettons l'accent sur les avantages qu'ils représentent pour notre étude.

3.1.1. De la linguistique de corpus à la linguistique outillée

Le terme *linguistique de corpus* renvoie globalement à une famille d'approches en linguistique qui se servent de corpus de textes dans l'analyse. Le plus souvent, les auteurs situent la linguistique de corpus par opposition à la linguistique introspective, car l'exploitation d'un corpus vise un travail sur des données attestées (Habert *et al.*, 1997 : 3sqq. ; Tognini-Bonelli, 2001 : 2 ; McEnery et Hardie, 2012 : 1-2 ; Biber et Reppen, 2015a : 1 ; Poudat et Landragin, 2017 : 14). Cette opposition reflète en réalité la nécessité d'un travail fondé sur des textes réels, lorsqu'il est jugé que l'introspection ne suffit pas (ou plus). Jacques cite notamment l'inadéquation des démarches introspectives lorsqu'il est question de décrire des phénomènes de variation et affirme que

« [c]'est précisément là que la linguistique de corpus prend sens, par la réintroduction du locuteur, du locuteur réel, authentique, et de son usage de la langue, dans les études et les descriptions linguistiques. » (Jacques, 2005 : 23-24)

L'autrice poursuit : « [l]e motif essentiel de la linguistique de corpus est à notre avis l'intérêt grandissant de la linguistique pour ces aspects impossibles à traiter par l'introspection et l'intuition » ; le besoin de données réelles fait alors écho à des questionnements linguistiques qu'une démarche introspective ne permet pas nécessairement d'aborder.

De plus, la linguistique de corpus est une approche intéressante de par la possibilité offerte par la technologie de tenir compte d'un plus grand volume de données dans l'analyse que ce qu'il est possible de traiter « à l'œil nu », c'est-à-dire sans recourir à des outils informatiques particuliers (Kennedy, 1998 : 268 ; Tognini-Bonelli, 2001 : 47 ; McEnery et Hardie, 2012 : 1-2 ; Biber et

Reppen, 2015a : 1). Cela permet également de tenir compte d'informations de nature quantitative dans l'analyse, qu'il s'agisse de fréquences brutes ou relatives, ou encore de statistiques descriptives et de mesures d'association (Jacques, 2005 : 25 ; McEnery et Hardie, 2012 : 48*sqq.* ; Poudat et Landragin, 2017 : 17 ; cf. *infra*).

Au-delà de ces aspects plutôt généraux, il existe nombre d'approches et d'objectifs de recherche se situant en linguistique de corpus, qui diffèrent à de nombreux égards (Poudat et Landragin, 2017 : 15 ; Stefanowitsch, 2020 : 21). L'une des distinctions les plus souvent évoquées concerne l'adoption d'une démarche *corpus-based* ou *corpus-driven*. Cette distinction a été proposée à l'origine par Tognini-Bonelli (2001) et est désormais reprise dans la plupart des ouvrages méthodologiques ou didactiques portant sur la linguistique de corpus (McEnery et Hardie, 2012 : 5-6 ; Biber et Reppen, 2015a : 2-3 ; Desagulier, 2017 : 9-10 ; Poudat et Landragin, 2017 : 30-31 parmi d'autres). De manière générale, l'approche *corpus-based* consiste à tester et vérifier des hypothèses sur un corpus, à propos d'un fonctionnement linguistique en particulier (Tognini-Bonelli, 2001 : 65). Le corpus est donc globalement considéré comme un moyen de valider une hypothèse ou une théorie. Comme le dit Tognini-Bonelli,

« corpus evidence is brought in as an extra bonus rather than as a determining factor with respect to the analysis, which is still carried out according to pre-existing categories; although it is used to refine such categories, it is never really in a position to challenge them as there is no claim made that they arise directly from the data. » (Tognini-Bonelli, 2001 : 66)

Au contraire, dans l'approche *corpus-driven*, le corpus est considéré comme l'objet de l'étude, et la démarche de recherche se construit en fonction des usages attestés dans le corpus. Les éléments d'élaboration d'une hypothèse ou d'une théorie proviennent ainsi directement des observations effectuées dans le corpus et interprétées par l'analyste (*Ibid.* : 84) :

« [t]he theory has no independent existence from the evidence and the general methodological path is clear: observation leads to hypothesis leads to generalisation leads to unification in theoretical statement. It is important to understand here that this methodology is not mechanical, but mediated constantly by the linguist, who is still behaving as a linguist and applying his or her knowledge and experience and intelligence at every stage during this process. » (*Ibid.* : 84-85)

D'autres auteurs mettent par ailleurs l'accent sur la diversité des approches en linguistique de corpus en insistant sur leur hétérogénéité (Péry-Woodley, 1995 ; Habert *et al.*, 1997 ; McEnery et Hardie, 2012) ou en évitant simplement le terme *linguistique de corpus*, marquant alors une position plus critique – ou prudente – quant au flou qui entoure parfois ce terme. Ces derniers parlent plutôt d'*approche en corpus* ou d'*approche sur corpus* (Cori et David, 2008 : 111). En anglais également,

différentes expressions sont utilisées pour éviter le terme *corpus linguistics*, telles que *corpus-based* ou *corpus-assisted approach* ou *study* (Taylor, 2008 : 182-183), dénotant une volonté de s'ancrer dans une démarche d'analyse en corpus sans nécessairement employer le terme qui semble parfois problématique.

Parallèlement, il est admis que le travail sur corpus s'effectue par le recours à différents outils informatiques, qu'il s'agisse d'un outillage que l'on peut qualifier de léger ou de plus sophistiqué. Le premier cas peut s'illustrer par la seule exploration d'un corpus à l'aide d'un concordancier, qui constitue la base de l'analyse de corpus. En effet, grâce à la présentation des contextes KWIC typique des concordanciers et grâce aux possibilités de tri à gauche et à droite d'une forme pivot, l'analyse d'un corpus est rendue globalement plus efficace et certaines régularités de fonctionnement sont plus aisément repérables que par une analyse « linéaire » du corpus (par exemple texte par texte) (McEnery et Hardie, 2012 : 1 ; Poudat et Landragin, 2017 : 15*sqq.*).

Le second cas peut faire intervenir des outils plus nombreux, qui accompagnent de manière plus poussée l'observation de certains phénomènes dans le corpus (Tanguy et Fabre, 2014 ; Léon, 2018). Il s'agit alors d'emprunter des méthodes de quantification et d'automatisation de certains traitements au TAL (traitement automatique des langues) et de les exploiter afin d'assister l'analyse linguistique, qui reste en grande partie qualitative (Habert, 2004 : 7 ; Valette, 2016 ; Léon, 2018). Différents étiqueteurs peuvent ainsi être utilisés pour enrichir le texte brut avec des informations de nature morphologique, syntaxique ou sémantique notamment, et afin de permettre aux outils d'effectuer des traitements plus poussés à partir de ces données enrichies (Habert *et al.*, 1997 : 15-22 ; Poudat et Landragin, 2017 : 35*sqq.*). De la même manière, différentes mesures statistiques peuvent être calculées afin d'évaluer la significativité d'indices de fréquence (Brezina, 2018 : 20), par exemple, ou de consolider l'interprétation (Jacques, 2017 : 94-96).

C'est en ce sens que l'on parle parfois de *linguistique outillée*, notamment sous l'influence de Habert (2004, 2006). Notons que ce terme apparaît dans les travaux de plusieurs auteurs, tels que Marchello-Nizia (2004), Tanguy et Fabre (2014), Jacques (2016), Landragin *et al.* (2016) et Poudat et Landragin (2017) mais que d'autres expressions sont parfois employées pour évoquer une approche en linguistique outillée, dans le sens de Habert, par exemple *approche outillée (en corpus* ou *sur corpus)*, *démarche outillée (en corpus* ou *sur corpus)*, *étude outillée en corpus*, *analyse linguistique outillée* (ou simplement *analyse outillée*), *analyse de corpus outillée*, etc. (Picton, 2009, 2018 ; Tanguy *et al.*, 2011 ; Condamines *et al.*, 2012 ; Landragin *et al.*, 2016 ; Poudat et Landragin, 2017 ; Condamines,

2018 ; Léon, 2018 ; Mayaffre *et al.*, 2019)⁵¹. En réalité, selon Habert (2004 : 6), la simple utilisation d'un corpus sur lequel fonder une analyse constitue déjà une manière d'outiller la linguistique, contrastant ainsi avec une démarche tout à fait introspective. De ce point de vue, la linguistique outillée englobe diverses approches de l'analyse linguistique sur corpus, qui s'appuient sur différents outils. Cependant, dans la littérature, il semble que le terme *outil* puisse être utilisé pour désigner tout un éventail de logiciels, programmes, scripts, suite d'outils, etc. (Tutin, 2010 ; Tanguy et Fabre, 2014 ; Landragin *et al.*, 2016 ; Jacques, 2017 ; Poudat et Landragin, 2017 : Mayaffre *et al.*, 2019). Par conséquent, nous pensons qu'il est nécessaire, à ce stade, d'apporter quelques clarifications quant aux différentes « familles » d'outils que l'analyse linguistique est susceptible d'exiger et, dans le même temps, de préciser nos propres choix terminologiques.

3.1.2. Les outils de la linguistique : quelques précisions

Reconnaissant les « confusions conceptuelles qui [peuvent] découler d'un emploi trop large du mot *outil* » (Habert, 2006 : 163), Habert propose une distinction entre les outils dédiés à l'analyse linguistique, qu'il appelle *instruments*, et les outils plus génériques, qui ne sont pas « spécifiquement orientés vers le traitement de données langagières mais qui peuvent être mis à contribution en linguistique » (*Ibid.* : 164), tels que des tableurs et des systèmes de gestion de bases de données. Il réserve ainsi l'emploi du terme *outil* à ces derniers. L'auteur propose également de considérer ce qu'il appelle des *utilitaires* (ou des *scripts*) à part de ces deux catégories. Pour lui, un utilitaire renvoie donc à « un logiciel qui assure une tâche limitée, périphérique à un traitement donné » (*Ibid.*). Habert donne ici l'exemple d'un programme servant à segmenter un texte en phrases.

Néanmoins, bien que ces distinctions rendent compte de la diversité des outils disponibles pour l'analyse linguistique, elles ne sont pas toujours reprises dans les travaux portant sur une approche outillée en corpus et le terme *outil* englobe le plus souvent l'ensemble des outils informatiques qui accompagnent la démarche, comme le soulignent Poudat et Landragin (2017 : 14). Il semble alors que ce terme soit devenu tout à fait générique dans les travaux mettant en œuvre une approche outillée en corpus. Par ailleurs, à la suite de ces auteurs (*Ibid.* : 15), nous pensons que cette généralité peut s'expliquer par la pluralité des approches en linguistique qui requièrent l'utilisation d'outils informatiques et par le nombre et la diversité de ces outils. Habert mentionne par exemple des outils d'annotation sémantique, d'étiquetage morphosyntaxique,

⁵¹ Cette emphase sur l'utilisation de différents outils informatiques dans la dénomination de la discipline semble être caractéristique des travaux de recherche en français. À notre connaissance et de ce que nous avons pu constater chez les auteurs que nous citons, les travaux anglo-saxons semblent préférer les termes cités précédemment, à savoir *corpus linguistics*, *corpus-based* ou *corpus-assisted approach* (ou *study*).

d'analyse syntaxique, de désambiguïsation, de bases de données, etc. (Habert, 2004 : 8). Poudat et Landragin illustrent également très bien cette diversité :

« [a]vec l'apparition des ordinateurs, la linguistique de corpus s'est outillée, c'est-à-dire que la méthodologie de gestion des données langagières s'est informatisée, et s'appuie donc sur des **outils informatiques**. Ces outils sont multiples : outils permettant l'annotation manuelle de textes ou d'enregistrements oraux ; outils pour la visualisation des données annotées ; outils d'interrogation de corpus, permettant l'expression de requêtes d'exploration et l'affichage des résultats obtenus de manière ergonomique ; outils de traitement automatique des langues, permettant d'ajouter automatiquement certaines annotations à un corpus ; outils d'importation et d'exportation de corpus dans des formats divers ; outils d'analyse de données ; outils de calculs statistiques, d'affichage des résultats et de génération d'indicateurs numériques sur le corpus ; outils de traitement de texte ; tableurs ; Systèmes de Gestion de Base de Données (SGBD), etc. » (Poudat et Landragin, 2017 : 14)

Pourtant, il semble qu'en dépit de cette généralité, voire de ce flou qui peut entourer l'emploi du terme *outil*, plusieurs « familles » d'outils peuvent être distinguées, qui renvoient partiellement aux distinctions que propose Habert, entre outils, instruments et utilitaires. C'est du moins ce que nous constatons à partir des différentes expressions qui sont utilisées dans la littérature. Par exemple, Tutin parle d'« outils de traitement de corpus » (Tutin, 2010 : 4) et d'« outils de la linguistique de corpus et du traitement automatique du langage » (*Ibid.* : 145) ; Tanguy et Fabre parlent d'« outils de TAL » (Tanguy et Fabre, 2014 : 17), mais également d'« outils d'investigation et de mesure » (*Ibid.* : 16) ; Jacques utilise aussi le terme *outil de TAL* (Jacques, 2016 : 93) ; Mayaffre *et al.*, enfin, parlent des « outils classiques de la linguistique de corpus » (Mayaffre *et al.*, 2019 : 102).

Il ressort de ces exemples qu'au moins deux familles d'outils spécifiques au travail sur corpus peuvent être distinguées : les outils de la linguistique de corpus, entendus comme les outils d'exploration de corpus (concordanciers, calculs de collocations et de segments répétés, etc.), et les outils de traitement automatique des langues, qu'il faut comprendre comme tout outil permettant d'étiqueter ou d'enrichir automatiquement les corpus (étiquetage morphosyntaxique, lemmatisation, analyse syntaxique, etc.) (Tanguy et Fabre, 2014 ; Jacques, 2016). Par ailleurs, en plus de ces deux familles d'outils clairement identifiables, les auteurs mentionnent également tout l'« outillage statistique » (Tanguy et Fabre, 2014 : 18), ou « appareillage statistique » (Mayaffre *et al.*, 2019 : 107), qui vient compléter la « boîte à outils » des linguistes et qui permet de calculer diverses mesures (corrélations, tests d'indépendance, etc.). Cet « outillage statistique » couvre un large éventail de possibilités de calculs et est rendu disponible selon différentes modalités, par exemple *via* un langage de programmation (R, Python, Perl, etc.) ou non (certains calculs s'effectuent très

bien avec un simple tableur tel qu'Excel) et *via* une interface graphique dédiée (par exemple RStudio (RStudio Team, 2019) ou IRaMuTeQ⁵²) ou non.

Parallèlement, certains des exemples d'outils de traitement de corpus donnés dans ces travaux s'apparentent plutôt à des logiciels qui regroupent plusieurs des outils que nous venons de mentionner. Mayaffre *et al.* (2019 : 102) citent notamment AntConc (Anthony, 2018), Sketch Engine (Kilgarriff *et al.* 2004, 2014), WordSmith Tool (Scott, 2016) et TXM (Heiden *et al.*, 2010). Ainsi, si nous prenons l'exemple de Sketch Engine, nous constatons que ce logiciel permet l'exploration de corpus à travers différents outils (concordancier, extracteur de *keywords*, segments répétés ou collocations, etc.) et applique différents traitements aux corpus, par exemple la segmentation du corpus en paragraphes et en phrases ou l'étiquetage morphosyntaxique⁵³. Ce logiciel regroupe alors plusieurs outils de linguistique de corpus et plusieurs outils de TAL.

La gamme d'outils et de logiciels disponibles pour l'analyse de corpus est donc très large. Le choix d'un outil (ou d'un logiciel) dépend par ailleurs essentiellement de l'objectif de la recherche, des diverses manipulations requises dans le but d'atteindre cet objectif et du type de données exploitées (cf. 3.3.1). Soulignons cependant que d'autres outils informatiques, qui ne sont pas conçus spécifiquement pour l'analyse et le traitement de corpus, entrent également en jeu lorsque les besoins de la recherche l'exigent. Il s'agit le plus souvent d'outils plus génériques de traitement de données, comme des systèmes de gestion de bases de données, ou même des tableurs, tels que MS Excel (Habert, 2006 : 164). L'intérêt de ce type d'outils repose sur la possibilité de gérer, manipuler et trier certaines sorties des outils, en particulier lorsque les outils ne le permettent pas. Par exemple, dans le but de filtrer les listes de candidats-termes extraits du corpus par TermoStat (Drouin, 2003), nous nous servons d'Excel et de Notepad ++, car ces listes ne sont pas directement modifiables dans l'interface de TermoStat (cf. 3.3.2).

Ces réflexions permettent ainsi de montrer non seulement que les outils exploitables dans une démarche de linguistique outillée sont nombreux et variés, mais également que les manières de les nommer et de les classer les uns par rapport aux autres peuvent différer selon les auteurs. De ce point de vue, et à la suite de la majorité des auteurs cités ici, nous faisons le choix d'appeler *outil* tout outil informatique entrant dans le processus d'analyse de corpus, qu'il s'agisse d'un outil dédié à l'analyse linguistique ou non⁵⁴. Nous réservons cependant l'emploi du terme *logiciel* aux systèmes

⁵² Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires, IRaMuTeQ, <http://www.iramuteq.org/> (page consultée le 14 mars 2021).

⁵³ Sketch Engine, <https://www.sketchengine.eu/guide/> (page consultée le 11 mars 2021).

⁵⁴ Il serait bien sûr utile et intéressant de poursuivre la réflexion et d'affiner les distinctions proposées par Habert entre les différentes familles d'outils existants et utilisés en linguistique outillée, à partir des

plus sophistiqués, c'est-à-dire aux logiciels composés de plusieurs outils de traitement et d'analyse de corpus (tels que ceux donnés en exemple *supra* : AntConc, Sketch Engine, TXM, etc.). Enfin, nous parlons de *méthodes statistiques* pour désigner les différents calculs et mesures statistiques que nous évoquons ici et que nous détaillerons dans la suite de notre travail, lesquelles peuvent être intégrées à un logiciel tel que Sketch Engine ou AntConc ou peuvent être mises en œuvre de manière indépendante.

3.1.3. Intérêt d'une approche outillée en corpus

Avant de préciser davantage l'intérêt d'une démarche outillée pour notre travail, une première remarque s'impose. À ce stade, il semble en effet nécessaire d'apporter certaines précisions sur les spécificités des approches en linguistique outillée par rapport aux approches en TAL. Si les deux approches sont apparentées, du fait que le TAL permet à la linguistique outillée de travailler efficacement sur des données attestées et massives et « permet de mettre en place des dispositifs rigoureux d'analyse qui augmentent considérablement les capacités d'observation des linguistes en les dotant d'une boîte à outils très élaborée » (*Ibid.* : 19), au moins une caractéristique importante les différencie. En effet, dans le cas du TAL, « le besoin de données langagières supposées génériques pour le développement de méthodes de traitement automatique » (*Ibid.* : 15) pousse les chercheurs à travailler à partir de corpus toujours plus volumineux. Dans ces cas, la quantité de données importe bien plus que la cohérence du corpus, car elle est absolument nécessaire pour les objectifs poursuivis par les travaux de TAL (*Ibid.*). En revanche, dans la majorité des approches en linguistique outillée, où l'on s'intéresse à des phénomènes précis et où le besoin de caractériser finement ces phénomènes prévaut, la pertinence du corpus reste au centre des préoccupations et doit être privilégiée par rapport à la quantité des données (voir par exemple Condamines (2003)). La masse de données prises en compte est donc bien moindre en comparaison avec les centaines de millions d'occurrences du TAL et se situe plutôt dans un intervalle allant de plusieurs dizaines de milliers d'occurrences pour les corpus les plus restreints à plusieurs millions, selon les textes disponibles (comme dans le cas du corpus Scientext, qui est un corpus composé de textes de spécialité provenant de différents domaines et dont le nombre d'occurrences s'élève à près de 5 millions pour le français (Tutin *et al.*, 2009)).

Dans notre cas, puisque nous nous intéressons à un phénomène précis que nous abordons par le biais de termes du domaine de la physique des particules et pour lequel nous cherchons à interroger le rôle de textes « intermédiaires » dans le passage des termes dans la langue générale, les

nombreux travaux adoptant une démarche en corpus. Néanmoins, une telle analyse requerrait un travail de recherche à part entière, qui ne saurait trouver sa place dans notre thèse.

données à disposition se retrouvent nécessairement restreintes. Pour reprendre une expression de Tanguy et Fabre (*Ibid.* : 19), il est nécessaire de « mieux contrôler la nature des données » dans ce cas (ce que nous détaillons au Chapitre 4). Par ailleurs, ces particularités amènent à considérer l'exploitation d'un outillage adapté, qui permet un équilibre entre quantitatif et qualitatif. C'est dans ce contexte précisément que nous favorisons une approche en linguistique outillée plutôt qu'une approche en TAL⁵⁵.

Dans cette perspective, accompagner la démarche d'analyse de corpus par le recours à différents outils et méthodes statistiques présente un réel intérêt pour notre objectif, de par les possibilités de tendre vers une démarche de recherche systématique. En effet, la mise en œuvre d'outils pour l'analyse linguistique comporte des avantages tels que la « systématisme du recueil d'occurrences, qui repose alors sur des patrons précis, [l']extension du nombre d'occurrences atteintes, [la] prise en compte de la fréquence » (Jacques, 2016 : 97). L'outillage de la démarche rend possible une quantification fiable des occurrences des phénomènes visés dans les données, ce qui facilite également les comparaisons entre différents sous-corpus le cas échéant. Par ailleurs, dès lors qu'il est possible d'obtenir, à partir du corpus, des fréquences d'unités, de cooccurrents, ou d'autres phénomènes qui soient fiables, de nouvelles potentialités apparaissent, comme l'ajout d'une dimension statistique à ces fréquences, somme toute relativement basiques à ce stade. Par exemple, grâce à la mise en œuvre de différentes méthodes statistiques, il devient ainsi possible :

- de calculer des corrélations entre différents phénomènes quantifiés, à l'aide du χ^2 ou d'autres tests statistiques (Levshina, 2015 : 115*sqq.* ; Gries, 2016 : 141*sqq.* ; Desagulier, 2017 : 151*sqq.* ; Stefanowitsch, 2020 : 167*sqq.*) ;
- d'aborder une démarche plutôt exploratoire dans le but de faire émerger de données complexes des phénomènes ou des unités à analyser plus finement, à l'aide d'une analyse factorielle des correspondances ou d'une analyse en composantes principales par exemple (Tutin, 2010 : 146 ; Tanguy *et al.*, 2011 : 146-148 ; Levshina, 2015 : 353-361 ; Desagulier, 2017 : 242-268 ; Poudat et Landragin, 2017 : 25 ; Mayaffre *et al.*, 2019 : 110-112) ;
- de faire émerger les différences lexicales les plus significatives entre deux corpus, à l'aide d'un calcul de *keyness* ou du calcul des spécificités lexicales (Lafon, 1980 ; Scott, 1997 ; Bertels et Speelman, 2013 ; Mayaffre *et al.*, 2019 : 103-105).

⁵⁵ Ces considérations assez générales sont davantage développées dans le Chapitre 5.

Notons également que ces méthodes peuvent se décliner selon le type d'observables sur lequel porte l'analyse (unité lexicale ou terme, contexte syntaxique d'une unité, unité textuelle telle qu'un paragraphe ou un document, etc.) (Poudat et Landragin, 2018 : 15-26 ; Mayaffre *et al.*, 2019 : 102).

Par ailleurs, comme nous le disons *supra*, l'exploitation d'outils provenant du TAL permet d'enrichir les données brutes de manière automatique. Ces traitements, relativement peu coûteux en temps, facilitent alors l'accès à certains phénomènes et réduisent le temps dédié à leur repérage par les analystes et la marge d'erreur humaine potentielle. Par exemple, dans le contexte d'un repérage semi-automatique de candidats-termes, l'étiquetage morphosyntaxique permet de repérer automatiquement les syntagmes qui correspondent à certains patrons définis en amont et qui concordent avec les formes que les termes peuvent prendre dans une langue en particulier. Ainsi, les syntagmes correspondant au patron nom + adjectif en français peuvent être repérés automatiquement grâce aux étiquettes morphosyntaxiques assignées à chaque occurrence (Drouin, 2015). Ce type de traitement permet d'éviter à l'analyste de devoir évaluer chaque occurrence individuellement. De la même manière, toutes les occurrences correspondant aux critères spécifiés peuvent être prises en compte systématiquement et sans introduire de subjectivité à ce stade. Cette part d'automatisation des traitements appliqués au corpus comporte donc un avantage indéniable pour l'analyse, celui de ne pas reposer uniquement sur le jugement de l'analyste, que ce soit pour assigner une étiquette à un élément du corpus (Habert, 2006 : 168) ou pour sélectionner les occurrences à analyser ultérieurement.

Ainsi, le recours à des outils permettant d'enrichir des données brutes et l'exploitation de différentes méthodes statistiques représentent une opportunité pour l'analyse linguistique, notamment pour renforcer les questions méthodologiques (Tanguy et Fabre, 2014 : 21-22 ; Jacques, 2017 : 165). Cette remarque nous amène alors à soulever la question de la reproductibilité de la recherche, qui, de même que la systématité, permet de limiter la subjectivité parfois associée à la recherche en linguistique et contribue de cette manière à inscrire ce domaine dans une perspective de recherche scientifique de manière générale.

De ce point de vue, la démarche d'analyse est balisée par les données fournies par les outils, et bien que l'interprétation humaine reste nécessaire, quelle que soit la part d'automatisation de la démarche (Zweigenbaum et Habert, 2004 : 103 ; Valette, 2016 ; Condamines, 2018), celle-ci s'appuie désormais sur des éléments solides, dont la sélection suit une procédure explicitée. Jacques (2016 : 94) souligne en effet que « le travail sur corpus confronte à la nécessité d'une sélection et d'une mise en ordre des données », ce qui implique inévitablement « une part non négligeable d'interprétation et donc de subjectivité ». Dès lors, il convient de limiter cette part de subjectivité

afin de développer une démarche reproductible, par le recours à divers outils et/ou méthodes statistiques, ainsi que par l'adoption d'« un cadre méthodologique transparent » (Picton, 2018 : 38), par exemple. Ainsi, en linguistique outillée,

« [l]e recours à des mesures statistiques est une suite logique de l'ensemble de la démarche : se baser sur corpus, limiter la subjectivité et l'interprétation par un choix de traits formels, mener des observations sur un grand nombre de données. » (Jacques, 2017 : 94-95)

Dans ce contexte, nous cherchons donc à exploiter le potentiel des outils et des méthodes statistiques existantes pour construire un cadre d'analyse qui permette de mettre en œuvre une démarche systématique et reproductible, ainsi que de limiter notre propre subjectivité. Par ailleurs, comme nous l'expliquons aux Chapitres 4 et 5, l'approche outillée que nous développons doit également permettre d'aborder l'analyse d'un phénomène complexe tel que le processus de déterminologisation, dans des données hétérogènes organisées en sous-corpus. Dans cette perspective, il est nécessaire de sélectionner les outils et méthodes appropriées à notre objet de recherche et à la spécificité de ces données. En effet,

« les phénomènes impliqués sont spécifiques aux objectifs visés, et doivent être abordés par le biais d'observables à définir dans chaque cas. Ces observables [...] ne peuvent pas toujours se réduire aux unités classiques du TAL [...] et ne sont définis que par une observation fine des données. Une fois identifiés, leur traitement et même leur observation au travers de méthodes quantitatives doit souvent faire appel à un outillage *ad hoc*. » (Tanguy et Fabre, 2014 : 20)

Le cadre global de la linguistique outillée est donc pertinent pour notre travail de thèse, en ce qu'il permet non seulement d'aborder l'analyse du fonctionnement des termes dans des données attestées, rassemblées en corpus, mais également de guider l'analyse par le recours à différents outils de traitement de corpus et méthodes statistiques, ce qui favorise la mise en œuvre d'une approche systématique et reproductible. Mais l'analyse linguistique des termes dans un corpus n'est pas une évidence en terminologie, du moins elle ne l'a pas toujours été. De la même manière, la prise en compte de problématiques liées à l'usage des termes dans des situations de communication variées n'est envisageable que si l'on considère que les termes sont effectivement susceptibles de fonctionner dans des contextes diversifiés et d'être sujets à la variation.

C'est pourquoi, avant de détailler notre cadre d'analyse, il est nécessaire de revenir sur les principales évolutions de la terminologie, particulièrement dans la manière d'aborder les termes et leur analyse, qui sont rendues possibles par un rapprochement des méthodes de terminologie avec les méthodes de linguistique de corpus (par exemple Condamines (2005a : 36), voir *infra*). Ainsi, dans la section suivante, nous mettons l'accent sur les principaux changements dans les travaux en

terminologie, qui mettent à profit les méthodes de linguistique outillée afin d'interroger de nouvelles questions, à propos des termes et de leurs usages, et d'apporter un regard nouveau, notamment sur la question de la déterminologisation.

3.2. Spécificités de la terminologie outillée

Dans cette section⁵⁶, nous explicitons d'abord les évolutions théoriques et méthodologiques de la terminologie qui se sont développées en réponse à différentes critiques dirigées contre l'approche plus traditionnelle généralement attribuée à Wüster, puis nous nous arrêtons plus en détail sur les propositions de la terminologie textuelle, qui développent particulièrement la réflexion sur les possibilités d'outiller l'analyse en terminologie et dont nous nous inspirons pour établir notre cadre d'analyse.

3.2.1. Évolution des théories de la terminologie

Parallèlement à l'évolution de la linguistique rendue possible notamment par l'analyse outillée de corpus, les études en terminologie se modifient dans une large mesure. Différents chercheurs (et parfois praticiens) estiment en effet que la Théorie Générale de la Terminologie (TGT) ne convient plus à la réalité des données impliquées dans le travail terminologique, et doit par conséquent évoluer, voire être remplacé.

Parmi les critiques les plus souvent observées dans la littérature, nous constatons une remise en question presque systématique du principe de biunivocité ainsi que de l'approche onomasiologique et en synchronie (Gambier, 1991a ; Gaudin, 1993 ; Boulanger, 1995 ; Cabré, 1999 ; Béjoint et Thoiron, 2000a ; Slodzian, 2000 ; Temmerman, 2000 ; L'Homme, 2004 ; Van Campenhoudt, 2006 parmi d'autres). En effet, le constat généralisé que les termes varient en discours remet sérieusement en cause le principe de biunivocité de l'approche « traditionnelle » de la terminologie. De même, des modifications importantes dans les manières d'approcher le terme et son analyse, notamment par la prise en compte de textes réels, tendent vers l'adoption d'une démarche sémasiologique, à travers laquelle il est désormais possible d'aborder l'évolution des termes.

Le rôle prescriptif des travaux en terminologie fait également l'objet de nombreuses critiques, et un rôle au contraire descriptif est revendiqué⁵⁷, non seulement dans une optique

⁵⁶ Dont le titre reprend une formulation proposée par Condamines et Picton (2014a, 2015).

⁵⁷ Pour autant, le rôle prescriptif des travaux en terminologie ne doit pas être rejeté, car lorsque l'objectif d'application l'exige, ces travaux peuvent mener à une normalisation de la terminologie d'un domaine (Bourigault et Slodzian, 1999 ; Condamines, 2005a).

socioterminologique (Gambier, 1991a : 51*sqq.* ; Gaudin, 1993 : 65*sqq.*, 2005 : 85-89), mais également dans d'autres optiques. Ainsi, d'un point de vue sociocognitivist, Temmerman (2000 : 37-28) argumente en faveur d'un élargissement des objectifs de la terminologie, dans le but de traiter des problématiques plus vastes que la seule normalisation de la terminologie, notamment au profit d'une meilleure description des termes et de leur fonctionnement en discours. De la même manière, Cabré (2005) propose une théorie communicative de la terminologie, où la description des données en contexte permet d'accéder entre autres à la variation terminologique, reconnaissant pleinement son rôle dans la communication spécialisée. Enfin, Slodzian insiste sur une perspective textuelle de la terminologie, qui privilégie une « approche descriptive des textes et des unités lexicales au détriment de l'approche normative » (Slodzian, 2000 : 74).

En réponse à ces critiques à l'égard de l'approche traditionnelle de la terminologie, d'intensité et de nature variables⁵⁸, de nouvelles propositions théoriques se sont développées, notamment la terminologie sociocognitive (Temmerman, 2000), la socioterminologie (Gambier, 1991a, 1991b ; Gaudin, 1993, 2003 ; Boulanger, 1995), la théorie communicative de la terminologie (Cabré, 1999, 2005), la terminologie textuelle (Bourigault et Slodzian, 1999 ; Condamines, 2003), l'optique lexico-sémantique (L'Homme, 2004, 2020). La majorité de ces propositions, sinon toutes, prônent un certain rapprochement de la terminologie avec la linguistique, au niveau théorique comme au niveau méthodologique. D'un point de vue théorique, un changement majeur concerne la définition du terme. Considéré d'abord, dans la TGT, comme l'« étiquette » du concept (Gambier, 1991a : 51 ; Slodzian, 2000 : 66 ; Gaudin, 2005 : 86-87 ; Cabré, 2016 : 71), le terme est aujourd'hui le plus souvent reconnu comme une unité fonctionnant en discours comme toute autre unité lexicale (Cabré, 1999, 2005 ; L'Homme, 2004 ; Gaudin, 2005 : 82). Pour L'Homme,

« [l]a particularité du terme, par rapport aux autres unités lexicales d'une langue, est d'avoir un *sens spécialisé*, c'est-à-dire un *sens* qui peut être mis en rapport avec un domaine de spécialité. La définition du "terme" [...] est donc relative. Elle dépend de la délimitation qu'on a faite d'un domaine spécialisé. » (L'Homme, 2004 : Chapitre 1, §38)

De manière similaire, Cabré (2016 : 78) envisage le terme comme une unité lexicale dont la valeur terminologique peut être activée ou non selon la situation de communication. Dès lors, toute unité lexicale est susceptible d'acquérir un « sens spécialisé » en discours et toute « unité terminologique » peut perdre cette valeur, selon le « contexte pragmatique ».

⁵⁸ Une discussion plus poussée de ces points de critique dépasserait le cadre de notre travail. Pour une revue détaillée, voir Campo (2012).

D'un point de vue méthodologique, et tenant compte de cette conception du terme, les auteurs de ces nouvelles propositions argumentent en faveur d'une évolution des pratiques terminologiques, qui doivent désormais se fonder sur les usages réels des termes, tels qu'ils peuvent être observés dans des textes produits par des experts (Bourigault et Slodzian, 1999 : 30*sqq.* ; Temmerman, 2000 : 42*sqq.* ; Cabré, 2003 : 178 ; L'Homme, 2004 ; Condamines, 2005a : 42*sqq.* ; Gaudin, 2005 : 89*sqq.* parmi d'autres). Comme l'affirment Bourigault et Slodzian,

« [c]'est dans les textes produits ou utilisés par une communauté d'experts, que sont exprimées, et donc accessibles, une bonne partie des connaissances partagées de cette communauté, c'est donc par là qu'il faut commencer l'analyse. » (Bourigault et Slodzian, 1999 : 30)

Par ailleurs, pour L'Homme, les textes ont presque toujours joué un rôle central en terminologie, il s'agit donc surtout de reconnaître ce rôle :

« [m]ême s'il est vrai que la délimitation précise d'un concept le guide lorsqu'il sélectionne des termes et qu'il les définit, le terminographe procède généralement à un repérage des termes dans des textes. Une fois qu'il les a identifiés, il en appréhende le sens. Il adopte donc la démarche inverse de celle préconisée par la terminologie classique. [...] Par ailleurs, on pourrait croire que les outils informatiques accentuent cette pratique, mais dans les faits, il en a presque toujours été ainsi. » (L'Homme, 2004 : Chapitre 1, §27)

Ainsi, dans une pratique de la terminologie centrée sur la description du fonctionnement des termes et sur la constitution de ressources, la prise en compte de textes réels constitue un enrichissement à la fois méthodologique et théorique. Les textes occupent alors une place fondamentale dans le développement de nouvelles propositions théoriques :

« objet empirique d'une linguistique textuelle, le texte est le point de départ de la description lexicale à construire. On va du texte vers le terme. Les bases théoriques de la terminologie doivent être ancrées dans une linguistique textuelle. » (Bourigault et Slodzian, 1999 : 31)

De ce point de vue, le potentiel qu'offre la linguistique de corpus pour traiter de grands volumes de textes⁵⁹ à l'aide d'outils dédiés est reconnu et mis à profit en terminologie (Cabré, 1999 : 6 ; Béjoint et Thoirion, 2000a : 14 ; Slodzian, 2000 : 73*sqq.* ; Condamines, 2005a). Cela passe notamment par la mise au point d'outils plus spécifiques à la terminologie, qui visent le repérage semi-automatique de termes ou de variantes terminologiques. Ces outils sont alors pensés pour accompagner autant que possible le travail de recensement et de description des termes (Slodzian,

⁵⁹ De l'ordre de plusieurs centaines de milliers d'occurrences, voire de plusieurs millions selon les textes disponibles dans les domaines et les langues étudiées (cf. section 3.1).

2000 : 61-62 ; Cabré *et al.*, 2001 ; Bowker et Pearson, 2002 : 165*sqq.* ; Drouin, 2003 ; Bourigault *et al.*, 2005 ; Daille, 2005, 2017).

Parallèlement à ces développements, le travail sur des données réelles a également contribué au développement des connaissances sur le terme et sur son fonctionnement dans les textes. Grâce au travail descriptif revendiqué par les auteurs des propositions théoriques mentionnées *supra*, il est de plus en plus reconnu que le terme est une unité lexicale, certes, mais une unité qui varie, selon différents facteurs (Gambier, 1991a : 51 ; Boulanger, 1995 : 195-196 ; Cabré, 2003 : 178 ; Gaudin, 2003 : 149-150 entre autres). L'approche textuelle désormais adoptée permet alors de rendre compte de ces variations, créant ainsi de nouvelles possibilités de recherche en terminologie (Béjoint et Thoiron, 2000a : 16-17 ; Temmerman, 2000 : 37 ; Gaudin, 2003 : 46*sqq.* ; Cabré, 2016 : 74). C'est dans ce contexte qu'il est possible d'aborder les questions liées à la circulation des termes en dehors de la sphère spécialisée, notamment à travers le prisme de la socioterminologie (Gambier, 1991a : 51 ; Gaudin, 1993, 2003, 2005), comme nous l'expliquons au Chapitre 1 (1.1). De même, des problématiques plus variées sont interrogées en lien avec cette question, par exemple la vulgarisation scientifique (Delavigne, 2001 ; Gaudin, 2003 : 105*sqq.* ; Chapitre 1, 1.2) le traitement des termes dans les dictionnaires (Béjoint, 1988 ; Thoiron, 1998 ; Josselin-Leray, 2005 ; cf. Chapitre 1, 1.3) ou encore la déterminologisation (cf. Chapitre 2), sur laquelle nous nous focalisons plus précisément.

Le cadre d'analyse que nous développons tire ainsi profit de ces différents développements de la terminologie, en particulier du travail sur des données textuelles réelles et à partir de différents outils informatiques. Nous cherchons cependant à analyser le processus de déterminologisation dans sa dimension continue à l'intérieur du continuum entre langues de spécialité et langue générale, ce qui impose de réfléchir au type de données à recueillir et à la manière de les organiser en sous-corpus, afin qu'elles permettent d'aborder cette dimension continue de la déterminologisation et qu'elles soient exploitables avec les outils existants. Nous développons cette réflexion dans le Chapitre 4.

En outre, la définition d'indices pertinents rendant possible l'exploration de ces données complexes et nécessairement hétérogènes constitue également une étape cruciale de notre travail. L'articulation entre ces indices et les différents outils permettant de les interroger fait l'objet de la section 3.3, où nous présentons notre cadre d'analyse. La construction de ce cadre d'analyse se fonde principalement sur les propositions de la terminologie textuelle, qui tire profit des acquis de la linguistique de corpus et du développement des outils dédiés à l'analyse linguistique pour aborder l'étude des termes, de leur fonctionnement et de leur variation. C'est pourquoi nous exposons dans

la section suivante les spécificités de cette approche de la terminologie et insistons en particulier sur la place qu'y occupe l'outillage.

3.2.2. Les propositions de la terminologie textuelle

3.2.2.1. De la constitution de ressources à la prise en compte de différentes situations de variation terminologique

Nous l'avons dit, la terminologie textuelle est une proposition théorique de la terminologie qui a émergé dans un contexte où l'approche traditionnelle était fortement remise en question. Cette proposition revendique notamment une meilleure prise en compte du texte spécialisé, qui doit constituer le point de départ des descriptions des termes permettant de construire des ressources terminologiques (Bourigault et Slodzian, 1999 : 30 ; Condamines et Picton, à paraître). La diversité des textes exploités dans ce but ainsi que la nécessité d'une systématisation des analyses et des descriptions favorisent le développement de cette approche textuelle de la terminologie, qui tire profit des outils de la linguistique de corpus et du TAL (Condamines, 2005a : 42). De ce point de vue, selon Bourigault et Slodzian (1999 : 30), « l'activité de construction d'une terminologie est désormais essentiellement une tâche d'analyse de corpus textuels ».

Dans le même temps, la diversité des objectifs auxquels les ressources terminologiques sont susceptibles de répondre montre qu'il est difficilement envisageable de constituer une seule terminologie qui corresponde aux besoins de chacun de ces objectifs (Bourigault et Slodzian, 1999 : 29-30 ; Condamines, 2005a : 42-43). Les choix des termes à retenir dans une ressource ainsi que la manière de les décrire peuvent en effet différer selon l'objectif, par exemple si la ressource est utilisée dans un contexte de traduction, de gestion des connaissances en entreprise ou encore si elle vise à alimenter les outils de TAL afin d'améliorer leurs performances (Bourigault et Slodzian, 1999 : 29 ; Aussenac-Gilles *et al.*, 2002 ; Aussenac-Gilles et Condamines, 2007 : 135-137). L'objectif de description joue donc un rôle majeur, dans la mesure où toute la démarche d'analyse s'articule autour de cet objectif, depuis la constitution du corpus sur lequel les analyses se fondent jusqu'à l'interprétation des résultats et leur validation par les experts de domaine (Condamines, 2005a : 43). Ainsi, par la prise en compte de l'objectif dans la constitution du corpus, on s'assure de la pertinence du corpus pour cet objectif.

Parallèlement, afin d'assister autant que possible le travail de repérage et de description, les outils d'exploration de corpus et de TAL sont mis à profit. Le repérage des termes est ainsi assisté par une extraction semi-automatique (Bourigault *et al.*, 2005) et le repérage de relations conceptuelles, en vue de construire un réseau terminologique, est systématisé par la recherche de

marqueurs de relation en corpus, par exemple des marqueurs d'hyponymie/hyperonymie (Condamines, 2005a : 42*sqq.*). Cependant, le repérage de ces marqueurs ne suffit pas à constituer une terminologie, une interprétation des données, en fonction de l'objectif d'analyse, reste toujours nécessaire (Bourigault et Aussenac-Gilles, 2003 ; Aussenac-Gilles et Condamines, 2007 ; Condamines, 2018), que celle-ci se fonde sur des marqueurs de relation ou sur une approche distributionnelle plus « classique ». En réalité, ces deux approches sont le plus souvent combinées : « l'interprétation se fait toujours par un ajustement entre connaissances langagières a priori et données du corpus et ce, jusqu'à ce que l'interprétation se stabilise » (Aussenac-Gilles et Condamines, 2007 : 140).

En outre, si la constitution de ressources tenant mieux compte des usages des termes dans les textes spécialisés est la préoccupation principale des recherches en terminologie textuelle à l'origine, le développement de la méthode et l'enrichissement des réflexions à propos du fonctionnement des termes dans les textes conduisent à la prise en compte de problématiques plus larges (Condamines, 2017 : 343*sqq.*). Ainsi, des questions liées à la variation des termes selon différents facteurs sont appréhendées du point de vue de la terminologie textuelle. Par exemple, Condamines *et al.* (2004) s'intéressent aux changements sémantiques des termes dans le temps. L'origine de cette étude provient d'une demande du Cnes⁶⁰, où la question de la variation dans le temps est associée au suivi de « projets spatiaux de longue durée, c'est-à-dire des projets dont la durée dépasse dix ans » (*Ibid.* : 547). L'objectif consiste alors à repérer et caractériser des changements sémantiques à partir d'une étude en corpus, afin d'aider les experts du Cnes à gérer les évolutions des connaissances dans ce type de projets de longue durée. Dans le même cadre théorique, Picton (2009) développe une méthode outillée en corpus de repérage de ces évolutions en diachronie courte, c'est-à-dire « sur de courts intervalles temporels » (*Ibid.* : 66). L'autrice identifie et caractérise ainsi 17 phénomènes d'évolution des connaissances à partir d'observations dans deux corpus du domaine spatial (*Ibid.* : 254*sqq.*).

D'autres questions de variation sont également abordées en terminologie textuelle, dans une perspective synchronique. Il s'agit de variation « en situation d'innovation disciplinaire » (Condamines et Dehaut, 2011 : 266) dans le domaine de l'exobiologie, où la recherche fait appel aux concepts d'autres domaines : chimie, biologie, physique et astronomie. L'objectif consiste alors à « mesurer l'impact de l'interdisciplinarité sur la constitution de la terminologie [de ce] domaine » (Condamines *et al.*, 2012 : 168). Différents phénomènes, de synonymie et de polysémie notamment, sont ainsi repérés dans un corpus composé de sous-corpus spécifiques à chaque discipline.

⁶⁰ Centre national d'études spatiales (<https://cnes.fr/fr>, page consultée le 5 mai 2020).

L'analyse de ces phénomènes permet alors de mettre en lumière des fonctionnements potentiellement conflictuels, pouvant avoir un impact sur la communication entre les experts de chacune de ces disciplines (Condamines et Dehaut, 2011 : 278*sqq.*).

Enfin, la question de la circulation des termes en dehors de la sphère spécialisée est également abordée par Condamines et Picton (2014a, 2014b) dans la perspective de la terminologie textuelle, en vue de répondre à une demande du Cnes. L'organisme souhaitait en effet « savoir comment le spatial, en tant qu'ensemble de connaissances, réelles ou imaginaires, se diffuse dans le grand public, de façon plus ou moins consciente » (Condamines et Picton, 2014a : 167). Les autrices ont interprété cette question d'un point de vue terminologique, ce qui les a amenées à s'intéresser à la déterminologisation, comme nous l'expliquons au Chapitre 2 (2.2.2).

C'est directement dans la continuité de cette étude que notre travail se situe. C'est pourquoi nous poursuivons notre réflexion sur le cadre d'analyse à adopter pour appréhender le processus de déterminologisation d'un point de vue outillé en corpus avec une présentation des principes d'analyse de la terminologie textuelle, mis au point à partir des différentes études citées dans cette section.

3.2.2.2. Principes d'analyse en terminologie textuelle

La démarche d'analyse en terminologie textuelle évolue et s'affine grâce aux différentes études que nous citons dans la section précédente, à l'élargissement des problématiques prises en compte et au perfectionnement des outils d'exploration de corpus et des méthodes statistiques disponibles (Condamines, 2018 ; Picton, 2018). Cette démarche repose généralement sur quatre caractéristiques (Condamines et Picton, à paraître).

Premièrement, le rapprochement de la terminologie avec la linguistique amène à travailler à partir de textes réels, dont la sélection et l'organisation en corpus sont pensées en fonction de l'objectif poursuivi. Par ailleurs, dans la perspective de la terminologie textuelle, les textes sont le plus souvent rassemblés en corpus comparables. Le principe sous-tendant l'exploitation de ce type de corpus repose sur l'idée que c'est de la comparaison de deux sous-corpus au moins que des différences pertinentes peuvent émerger, lesquelles seront ensuite interprétées en fonction de l'objectif. L'approche est ainsi généralement contrastive.

Deuxièmement, les termes sont le plus souvent considérés comme le point de départ de l'analyse. De manière générale, il s'agit alors de repérer des régularités de fonctionnement des termes, à partir de différents indices observables en corpus. Dans ce cas, la prise en compte des contextes distributionnels des termes représente une importante part de la démarche.

Troisièmement, l'exploration du corpus s'appuie sur différents outils selon l'objectif poursuivi et s'inscrit dans une démarche de recherche qui tend vers la systématique et la reproductibilité. Dans cette perspective, les outils doivent être sélectionnés en fonction de leur pertinence pour le type de données sélectionnées et leur organisation en sous-corpus et pour permettre de répondre à cet objectif. En outre, les sous-corpus sont explorés par le biais de différents indices, qui sont définis pour chaque étude selon les types d'observables à prendre en compte pour interroger le phénomène visé. Les différents travaux que nous citons dans la section précédente ont permis d'affiner ces indices. Ils se déclinent en quatre types principaux :

1. **indices quantitatifs** : ce premier type d'indices concerne principalement des variations de fréquence des termes à travers différents sous-corpus. Il peut s'agir de fréquences plutôt inattendues dans un sous-corpus par rapport à un autre, dont la significativité se mesure à l'aide de méthodes statistiques (spécificité, χ^2 , etc.), ou d'absence/présence de termes dans les sous-corpus (Picton 2009 : 105*sqq.*, 2018 : 31 ; Condamines et Dehaut, 2011 : 271-272) ;
2. **indices centrés sur des variations formelles des termes** : la présence de variantes terminologiques, c'est-à-dire différentes dénominations renvoyant généralement au même concept, peut indiquer l'existence de différents points de vue sur ce concept, voire des variations conceptuelles (Picton, 2009 : 189*sqq.*, 2018 : 31 ; Condamines et Dehaut, 2011 : 272) ;
3. **indices distributionnels** : ce type d'indices se focalise sur l'analyse des contextes d'apparition des termes en corpus, c'est-à-dire sur leur distribution. Il se décline en deux approches :
 - a. **avec interprétation *a priori*** : dans ce cas, la recherche se concentre sur des structures spécifiques, qui permettent de cibler avec plus ou moins de précision l'élément analysé ou l'information pertinente et dont on sait qu'elles peuvent être systématiquement (ou du moins le plus souvent) associées à une interprétation « stable ». En d'autres termes, il s'agit de focaliser la recherche sur des marqueurs de contextes riches en connaissances (Meyer, 2001 ; Condamines, 2002, 2005a). En outre, comme le dit Condamines (2018), « la notion de "connaissance" peut varier selon l'objectif d'étude des corpus spécialisés », les marqueurs doivent donc être définis pour chaque étude, afin qu'ils correspondent aux éléments précis sur lesquels porte l'analyse. Par exemple, Picton (2009 : 147*sqq.*) exploite cet indice dans le but de repérer des

contextes riches en connaissances évolutives, la définition de marqueurs pertinents tient donc compte de cet objectif ;

- b. **sans interprétation *a priori***: dans ce dernier cas, l'analyse se focalise sur l'environnement distributionnel des termes et cherche à mettre au jour des similarités ou divergences de fonctionnement selon l'hypothèse que le sens d'une unité peut être appréhendé par sa distribution dans un texte, ou un corpus. En particulier, c'est de la comparaison de la distribution des termes dans plusieurs sous-corpus que des différences de fonctionnement peuvent émerger (Condamines, 2018 ; Picton, 2018 : 31).

Ces indices peuvent être interrogés par le biais de différents outils d'exploration de corpus et sont retravaillés pour correspondre précisément aux besoins de chaque étude et des phénomènes analysés. Par ailleurs, ces indices mettent en évidence différents éléments, qui doivent être interprétés par l'analyste en relation avec l'objectif de l'étude. Cette interprétation se construit généralement avec le concours d'experts de domaine.

Ainsi, quatrième, la démarche globale d'analyse s'effectue en collaboration avec des experts de domaine. De manière générale, le travail avec les experts est envisagé comme une collaboration, les experts possédant la connaissance du domaine et les analystes entretenant la distance nécessaire avec les données langagières pour permettre des descriptions objectives (Bourigault et Slodzian, 1999 : 30 ; Bourigault et Aussenac-Gilles, 2003). Dans cette perspective, des experts sont sollicités tout au long du processus d'analyse, c'est-à-dire dès l'étape de constitution du corpus (Condamines, 1994 : 48-49, Bourigault et Slodzian, 1999 : 30). Leur participation peut cependant varier, en fonction des besoins de l'étude et de la nature des analyses effectuées. En effet, les experts peuvent être interrogés dans une étape de validation des résultats dans un objectif de construction de ressources terminologiques, mais ils peuvent également intervenir à d'autres moments de l'analyse, dans diverses étapes d'interprétation des données fournies par les outils (comme nous le montrerons aux Chapitres 6 et 7). Par ailleurs, le rôle des experts a notamment été mis en évidence par Picton (2009 : 300*sqq.*), qui développe un principe de « co-construction de l'interprétation ». Ce principe implique ainsi que les experts participent pleinement au processus d'interprétation des résultats. En effet, une

« analyse ne peut être conçue comme la mise au jour de phénomènes *donnés* par les indices, mais bien comme le résultat de la *construction* d'une interprétation de phénomènes langagiers. » (Picton, 2009 : 300)

La construction d'une interprétation repose alors sur les données fournies par les outils et selon les indices mis en œuvre en corpus, sur la compétence linguistique de l'analyste et sur une collaboration avec des experts de domaines, et elle reste toujours guidée par l'objectif poursuivi ou l'application visée (*Ibid.* : 302).

Dans ce contexte, travailler sur la déterminologisation en tant que processus continu crée une opportunité de poursuivre la réflexion sur les possibilités méthodologiques d'analyse de ce processus en corpus, que ce soit dans la manière de constituer les données ou dans la manière de les explorer. Ces deux aspects sont détaillés dans les Chapitres 4 et 5, respectivement. En outre, dans une perspective de terminologie outillée, notre approche de la déterminologisation nous pousse à continuer de mettre à profit les développements de la terminologie textuelle, du TAL et de la linguistique outillée, afin de construire un cadre d'analyse rigoureux, qui nous permette de mettre en œuvre une démarche d'analyse systématique et reproductible et, dans le même temps, de limiter notre propre subjectivité.

3.3. Cadre d'analyse

3.3.1. Construction d'un cadre d'analyse adéquat

Dans notre travail de thèse, nous abordons le processus de déterminologisation à travers sa dimension continue entre langue de spécialité et langue générale et nous nous intéressons particulièrement au fonctionnement des termes dans ce processus. Afin d'aborder ces questions, nous mettons en œuvre une démarche d'analyse outillée en corpus. Plus précisément, nous inscrivons notre travail de manière générale dans une démarche de terminologie textuelle, c'est-à-dire que notre analyse est outillée et s'appuie sur un corpus comparable dont la constitution fait partie intégrante de notre réflexion, que nous nous focalisons essentiellement sur les termes et que nous collaborons avec des experts de domaine. Dans cette perspective, les termes sont considérés comme les points d'entrée dans les données et l'analyse s'articule autour des termes et de leurs fonctionnements dans le corpus. De même, deux physiciennes des particules, de l'Université de Genève et du CERN⁶¹, sont interrogées à plusieurs reprises dans notre travail.

Cependant, la constitution d'un corpus dans ce but introduit nécessairement certaines limites. Au Chapitre 2, nous définissons la déterminologisation comme un processus continu dans le continuum langue de spécialité-langue générale, mais le recours à un corpus pour analyser ce

⁶¹ Organisation européenne pour la recherche nucléaire (<https://home.cern/fr>, page consultée le 24 mai 2021).

processus impose de figer les dynamiques dans une représentation statique des phénomènes linguistiques. Bien que cette stabilisation soit absolument nécessaire si l'on souhaite décrire les fonctionnements des termes en situation de déterminologisation, nous cherchons à créer la possibilité d'interroger ces dynamiques, par l'organisation des données en différents sous-corpus (cf. Chapitre 4) et par la définition d'une méthode de comparaison adéquate (cf. Chapitre 5).

Par ailleurs, nous nous intéressons plus particulièrement aux changements sémantiques susceptibles de se produire au cours du processus de déterminologisation. Il est donc nécessaire de définir une méthode de repérage et de caractérisation de ces changements. Dans notre cas, nous nous servons de deux indices de type distributionnel sans interprétation *a priori*, dont le choix est déterminé pour leur complémentarité et pour leur capacité à permettre des descriptions fines des fonctionnements des termes dans le corpus. Ces indices passent alors par l'observation de deux types de contextes distributionnels : un contexte composé de l'ensemble des unités qui apparaissent dans une fenêtre déterminée autour d'un terme et un contexte composé uniquement des unités qui entretiennent une relation de dépendance syntaxique avec un terme. La réflexion qui sous-tend le choix de ces indices, leur fonctionnement et leur mise en œuvre est détaillée au Chapitre 5. L'exploration de ces deux indices en corpus fait l'objet des Chapitres 6 et 7.

Afin d'accompagner l'exploration de ces deux indices, nous mettons en œuvre des indices de nature quantitative, essentiellement à travers la prise en compte d'informations de fréquence. En particulier, la fréquence de certains phénomènes intervient à différentes étapes de l'analyse, par exemple pour sélectionner les phénomènes à observer plus finement, pour évaluer la significativité de phénomènes ou encore pour mettre en œuvre certaines méthodes statistiques. Nous nous servons par ailleurs de méthodes statistiques dans différents buts, notamment pour repérer automatiquement le lexique significativement plus présent dans un sous-corpus par rapport à un autre (cf. Chapitre 5), ou pour limiter les biais dus à la prise en compte de notre jugement seul, par le recours au calcul d'un accord inter-annotateurs (cf. Chapitre 6). Ces méthodes statistiques sont donc mises en œuvre dans le but de construire une interprétation qui limite notre propre subjectivité et qui repose sur une démarche d'analyse systématique et reproductible.

En outre, l'interprétation des données fournies par les outils et les méthodes statistiques se construit d'une part en collaboration avec les expertes lorsque cela est adéquat et d'autre part par le retour systématique au contexte (Valette, 2016 ; Condamines, 2018 ; Egbert *et al.*, 2020 : 46*sqq.*), afin de ne pas surinterpréter les données fournies par les outils, qui sont parfois décontextualisées. Dans ce cas, plusieurs allers-retours entre les données fournies par les outils et les contextes du corpus peuvent être nécessaires.

Dans ce même objectif de systématique et de reproductibilité, et afin de gérer la complexité et l'hétérogénéité des données textuelles sélectionnées et rassemblées en sous-corpus (cf. Chapitre 4), nous tirons également profit des possibilités de visualisation offertes, notamment, par le langage de programmation R. Comme nous l'expliquons notamment au Chapitre 6 (6.3.1), la visualisation de certaines données extraites du corpus peut permettre de faciliter le repérage de phénomènes intéressants à analyser plus en détail et de rendre les comparaisons entre plusieurs sous-corpus plus aisées.

En conclusion, bien que notre travail se situe de manière générale dans la continuité des travaux s'inscrivant dans la perspective de la terminologie textuelle, nous affinons la démarche par le recours à différents outils et méthodes statistiques, que nous mettons en œuvre dans le but d'analyser le processus de déterminologisation dans un corpus comparable complexe. Par ailleurs, la spécificité du phénomène étudié, de même que l'hétérogénéité des données que nous analysons (qui est d'ailleurs nécessaire pour répondre à l'objectif que nous poursuivons), requiert un outillage adapté, voire *ad hoc*, comme le suggèrent Tanguy et Fabre (2014 : 20) lorsque le contexte de travail est aussi spécifique. C'est pourquoi nous détaillons nos choix en matière d'outillage dans la section suivante.

3.3.2. Choix des outils

Nous détaillons ici les différents choix d'outils que nous exploitons dans notre travail, nous décrivons le fonctionnement de chacun de ces outils et exposons leurs avantages pour les besoins de notre objectif de recherche. Ainsi, en plus du logiciel AntConc, nous exploitons un extracteur de termes : TermoStat. Par ailleurs, afin de préparer les données, nous utilisons TreeTagger pour lemmatiser et étiqueter le corpus et, dans une perspective distributionnelle où seuls les contextes syntaxiques sont pris en compte, nous exploitons également Talismane, un analyseur syntaxique en dépendances. Enfin, nous utilisons le langage de programmation R, à travers l'interface graphique RStudio, afin d'effectuer différentes manipulations et de mettre en œuvre plusieurs méthodes statistiques.

AntConc

AntConc (Anthony, 2018) est un logiciel libre, développé par Laurence Anthony (Anthony, 2005) et régulièrement mis à jour⁶². Outre le fait que ce logiciel soit gratuit, d'autres avantages à son utilisation expliquent notre choix. L'ergonomie de son interface rend son utilisation plutôt aisée,

⁶² Les versions d'AntConc que nous utilisons sont la 3.4.2 et la 3.5.6.

même avec de nombreux sous-corpus. Il est en effet très facile de visualiser la répartition d'unités selon leur fréquence dans différents sous-corpus avec AntConc, chaque fichier contenant un sous-corpus étant affiché individuellement dans l'interface. Cette possibilité rend son utilisation tout à fait intéressante pour une analyse essentiellement fondée sur la comparaison de sous-corpus. Par ailleurs, AntConc gère très bien les corpus de taille modeste, tel que celui que nous exploitons dans notre travail de thèse (cf. Chapitre 4).

Cette flexibilité représente également un atout de ce logiciel, de même que les nombreuses possibilités de paramétrage. Avec AntConc, il est en effet possible de véritablement personnaliser les paramètres de recherche, ce qui permet de tenir compte des besoins particuliers de chaque recherche et de cibler précisément les éléments sur lesquels se focaliser. Par ailleurs, l'outil est totalement transparent, rien n'est laissé au hasard, que ce soit au niveau des résultats (aucun tri *a priori* de la part de l'outil) ou au niveau des paramètres (tous les choix requis pour l'utilisation d'une fonctionnalité peuvent être contrôlés par l'utilisateur). Ainsi, le processus de recherche est entièrement maîtrisé par l'analyste.

Enfin, soulignons que les outils regroupés dans ce logiciel et ses différents modes de recherche constituent d'autres avantages à son utilisation. Bien sûr, l'outil principal d'AntConc reste le concordancier, mais la possibilité d'affiner les recherches avec des expressions régulières puissantes le rend d'autant plus intéressant pour des objectifs précis. Par exemple, c'est de cette manière que nous repérons et isolons des contextes composés d'un nombre fixe d'occurrences avant et après un terme pivot (cf. Chapitre 5, 5.3.1.1). Un autre outil pertinent pour notre travail est la *Keyword List*, qui permet d'identifier les *keywords* d'un corpus par rapport à un autre corpus et de les classer selon un score de *keyness*. En outre, du fait qu'AntConc est régulièrement mis à jour, cette fonctionnalité tient compte d'un grand nombre de calculs de *keyness* actuellement utilisés par différents chercheurs et qui ne sont pas nécessairement regroupés dans d'autres logiciels⁶³. Cela permet encore une fois une vraie personnalisation des paramètres, afin que ceux-ci correspondent autant que possible aux besoins de la recherche et de ce que l'on vise à identifier.

TermoStat

TermoStat Web 3.0 (Drouin, 2003) est un système d'extraction automatique de candidats-termes hybride, développé à l'Université de Montréal et utilisable librement à des fins de recherche. L'hybridité de ce système repose sur un fonctionnement à la fois à base de connaissances

⁶³ En comparaison notamment avec Sketch Engine (<https://www.sketchengine.eu/guide/keywords-and-term-extraction/#toggle-id-4>, page consultée le 6 avril 2020) et WordSmith Tools (https://lexically.net/downloads/version7/HTML/keywords_calculate_info.html, page consultée le 6 avril 2020).

linguistiques et à base de connaissances statistiques. Plus précisément, le système travaille à partir d'un corpus lemmatisé et étiqueté par TreeTagger (le corpus d'analyse), d'où il extrait des candidats-termes (CT) qui correspondent à certains patrons syntaxiques. Ces patrons sont définis pour chaque langue supportée, selon la forme que peuvent prendre les termes dans cette langue. Il peut s'agir de termes simples (nom, adjectif, verbe, par exemple) ou de termes complexes (nom + adjectif, nom + préposition + nom, par exemple). La fréquence de ces CT est ensuite comparée à leur fréquence dans un corpus de référence. Pour le français, le corpus de référence est un corpus composé d'articles du journal *Le Monde* publiés en 2002. Sur la base de ces comparaisons, un score est assigné à chaque candidat-terme, à partir d'un calcul statistique. Quatre calculs peuvent être choisis selon les besoins et les préférences de l'analyste : les spécificités, le χ^2 , le *log-likelihood* et le *log-odds ratio*. Ainsi, de manière générale, plus le score d'un CT est élevé, plus il est probable que ce CT soit un terme du domaine à l'étude⁶⁴.

Les résultats de TermoStat sont présentés sous la forme d'une liste de candidats-termes, triés selon leur score et qui sont accompagnés de différentes informations : la fréquence des candidats-termes, leur score, les différentes variantes orthographiques (ou flexionnelles) repérées dans le corpus d'analyse et la matrice, c'est-à-dire le patron auquel ils correspondent (nom, nom + adjectif, etc.). Le tableau 3.1 illustre le format de sortie de l'extraction à partir d'un extrait des candidats-termes obtenus avec le corpus et pour lesquels le calcul des spécificités est utilisé.

⁶⁴ TermoStat 3.0, http://termostat.ling.umontreal.ca/doc_termostat/doc_termostat.html (page consultée le 8 avril 2020).

Candidat de regroupement	Fréquence	Spécificité	Variante orthographiques	Matrice
jet	4847	318.09	jet ; jets	Nom
masse	4538	278.98	masse ; masses	Nom
boson	3376	272.24	boson ; bosons	Nom
transverse	1622	188.91	transverse ; transverses	Adjectif
détecteur	1657	186.1	détecteur ; détecteurs	Nom
désintégration	1515	179.12	désintégration ; désintégrations	Nom
bruit de fond	1274	166.7	bruit de fond ; bruits de fond	Nom Préposition Nom
lepton	1206	162.87	lepton ; leptons	Nom
section efficace	930	143.01	section efficace ; sections efficaces ; sections efficaces	Nom Adjectif
impulsion transverse	724	126.16	impulsion transverse ; impulsion transverses ; impulsions transverses	Nom Adjectif
standard	787	124.77	standard	Adjectif
spin	400	93.09	spin ; spins	Nom
électrofaible	394	93.01	électrofaible	Adjectif
hadron	394	92.88	hadron ; hadrons	Nom
énergie transverse manquante	274	77.52	énergie transverse manquante	Nom Adjectif Adjectif
désintégrer	213	65.66	désintégrer ; désintégrera ; désintègre ; désintègrent ; désintégreraient	Verbe

Tableau 3.1 : Extrait des candidats-termes extrait du sous-corpus *Spécialisé* avec TermoStat

TermoStat est un logiciel aujourd'hui largement utilisé dans des tâches de repérage de termes, notamment pour son fonctionnement intuitif dû à une interface ergonomique et simple d'utilisation ainsi que pour ses performances tout à fait satisfaisantes (Alecú *et al.*, 2012 ; Bertels et Speelman, 2013 ; Inkpen *et al.*, 2016 ; Marín Pérez, 2016 ; Daille, 2017). De manière générale, en effet, lorsque ce logiciel est comparé à d'autres systèmes d'extraction automatique de termes, le rappel et la précision qui sont rapportés sont soit comparables aux autres systèmes (Alecú *et al.*, 2012 : 516) soit supérieurs (Marín Pérez, 2014 : 83 ; Inkpen *et al.*, 2016 : 160). Bien évidemment, les résultats ne sont pas parfaits, c'est pour cette raison que l'on parle généralement de *candidats-termes* et non de *termes*, pour désigner les sorties du logiciel. Par ailleurs, ce bruit est un écueil bien

connu en terminologie (L'Homme, 2004 : Chapitre 6, §4-6). La validation du statut terminologique des CT est une tâche qui revient aux terminologues et/ou aux experts, selon leur disponibilité et l'approche de l'analyse terminologique qui est adoptée (*Ibid.*). Nous reviendrons plus en détail sur ce point dans le Chapitre 4.

Dans notre cas en particulier, l'intérêt d'utiliser un extracteur automatique de termes performant tel que TermoStat réside dans la possibilité de tenir compte exclusivement des termes qui sont réellement attestés dans le corpus et de ne pas fonder notre analyse sur une liste constituée *a priori* et de manière peut-être subjective, même si nous faisons appel aux experts de domaine. Enfin, soulignons deux autres avantages à l'utilisation de TermoStat. D'abord, ce logiciel permet de télécharger les résultats de l'extraction, ce qui rend leur traitement plus aisé avec des programmes tiers, notamment afin de nettoyer la liste de CT, qui est nécessairement bruitée. De la même manière, bien que l'interface Web de TermoStat soit relativement simple à utiliser, elle reste peu pratique dès lors que différents sous-corpus sont comparés. C'est donc de ce point de vue que la possibilité de télécharger les résultats pour les traiter avec d'autres programmes est intéressante. Un dernier avantage non négligeable de ce système consiste en la possibilité d'extraire des candidats-termes non exclusivement nominaux, ce que ne permettent pas nécessairement d'autres systèmes existants (Alecú *et al.*, 2012 : 514). Or, dans le domaine que nous étudions, la physique des particules, nous pensons que la prise en compte des termes verbaux et adjectivaux doit être envisagée (et donc rendue possible par l'extracteur exploité)⁶⁵.

TreeTagger

TreeTagger est un étiqueteur morphosyntaxique à fonctionnement probabiliste (Schmid, 1994), librement disponible à des fins de recherche et installable facilement sur les principaux systèmes d'exploitation actuels. Il compte parmi les étiqueteurs les plus exploités, ce qui s'explique notamment par ses performances satisfaisantes et le grand nombre de langues qu'il est possible de traiter⁶⁶. En effet, pour le français, la précision rapportée sur des corpus divers est de plus de 92 % (Allauzen et Bonneau-Maynard, 2008 ; Denis et Sagot, 2010 ; Balvet *et al.*, 2014). En outre, selon Denis et Sagot (2010) et Falk *et al.* (2014), la précision de TreeTagger sur des mots nouveaux (c'est-à-dire qui n'apparaissent *a priori* pas dans le corpus sur lequel l'étiqueteur a été entraîné) varie entre 75 % et 82 % environ, ce qui reste intéressant en terminologie, si l'on part du principe que certains termes parmi les plus spécifiques du domaine à l'étude peuvent être considérés par l'outil comme

⁶⁵ Sans rentrer dans le détail à ce stade, mentionnons tout de même que nous nous attendons à ce que certains verbes soient relevés par TermoStat, notamment *désintégrer* et *interagir*.

⁶⁶ TreeTagger, <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/> (page consultée le 6 avril 2020).

des mots nouveaux. Par ailleurs, cet étiqueteur est intégré à d'autres outils de traitement de corpus, tels que Sketch Engine⁶⁷, TXM⁶⁸ et Termostat⁶⁹, par exemple, ce qui atteste de son utilisation tout à fait répandue. Notre choix s'est donc naturellement porté sur cet étiqueteur.

Pour chaque langue, TreeTagger est entraîné sur un corpus d'apprentissage annoté manuellement avec un jeu d'étiquettes morphosyntaxiques. Pour le français, 33 étiquettes sont utilisées⁷⁰. Enfin, TreeTagger utilise un arbre de décision pour résoudre les cas d'ambiguïté (par exemple lorsqu'une forme fléchie peut être rattachée à plusieurs lemmes ou lorsqu'un lemme peut relever de plusieurs catégories grammaticales) (Schmid, 1994).

L'utilisation de cet outil est très simple : il suffit de soumettre un corpus en texte brut et de spécifier la langue pour que TreeTagger en fournisse une version lemmatisée et étiquetée. L'exemple ci-dessous (tableau 3.2) illustre le résultat obtenu, à partir de la phrase suivante tirée du corpus : « Ces collisions produisent des particules instables. ». La première colonne du tableau indique l'occurrence telle qu'elle apparaît dans le corpus, la seconde indique l'étiquette morphosyntaxique correspondant à l'occurrence et la dernière indique le lemme de l'occurrence.

Occurrence	Étiquette	Lemme
Ces	PRO:DEM	ce
collisions	NOM	collision
produisent	VER:pres	produire
des	PRP:det	du
particules	NOM	particule
instables	ADJ	instable
.	SENT	.

Tableau 3.2 : Exemple d'une phrase lemmatisée et étiquetée avec TreeTagger

Talismane

Talismane est un analyseur syntaxique en dépendances développé au sein du laboratoire Cognition, Langues, Langage, Ergonomie (CLLE) (Urieli, 2013), à l'Université Toulouse – Jean Jaurès, et mis à disposition librement pour des fins de recherche⁷¹. La documentation très riche qui

⁶⁷ Sketch Engine, <https://www.sketchengine.eu/french-treetagger-part-of-speech-tagset/> (page consultée le 6 avril 2020).

⁶⁸ TXM, <http://textometrie.ens-lyon.fr/spip.php?rubrique96> (page consultée le 6 avril 2020).

⁶⁹ Termostat Web 3.0, http://termostat.ling.umontreal.ca/doc_termostat/doc_termostat.html (page consultée le 6 avril 2020).

⁷⁰ TreeTagger, <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/> (page consultée le 6 avril 2020).

⁷¹ Ressources développées à CLLÉ (REDAC), Talismane (<http://redac.univ-tlse2.fr/applications/talismane.html>, page consultée le 7 avril 2020).

l'accompagne rend ce programme prêt à l'emploi, facilitant ainsi son installation et son utilisation. L'annotation en dépendances de Talismane se fonde majoritairement sur une adaptation du French Treebank, qui est la ressource de référence pour le français et dont les informations syntaxiques ont été manuellement annotées (Abeillé *et al.*, 2003 ; Candito *et al.*, 2010 ; Urieli, 2013).

La chaîne de traitement de Talismane comprend quatre étapes principales, à partir d'un corpus en version brute (au format .txt). L'outil procède d'abord à la délimitation des phrases, puis à la tokenization et à l'étiquetage morphosyntaxique et enfin à l'analyse syntaxique en dépendances (Urieli, 2013 : Chapitre 3). Le tableau 3.3 illustre le résultat d'une telle analyse effectuée par Talismane, à partir d'un extrait du corpus.

1	Ces	ces	DET'	n=p	2	det
2	collisions	collision	NC	n=p g=f	3	suj
3	produisent	produire	V	n=p t=PS p=3	0	root
4	des	des	DET'	n=p	5	det
5	particules	particule	NC	n=p g=f	3	obj
6	instables	instable	ADJ	n=p	5	mod
7	.	.	PONCT		6	ponct

Tableau 3.3 : Exemple d'une phrase analysée syntaxiquement en dépendances par Talismane

Dans cet extrait, chaque ligne contient les informations relatives à une occurrence. Les informations, qui sont séparées par des tabulations dans le fichier de sortie de Talismane, sont de la nature suivante (*Ibid.*, Chapitre 1) :

- colonne 1 : identifiant unique de chaque occurrence d'une phrase ;
- colonne 2 : occurrence, telle qu'elle apparaît dans le corpus ;
- colonne 3 : lemme de cette occurrence ;
- colonne 4 : catégorie grammaticale de l'unité ;
- colonne 5 : informations morphosyntaxiques (genre et nombre, mais aussi temps verbal et personne lorsque l'unité est un verbe) ;
- colonne 6 : identifiant du gouverneur de l'occurrence. Le zéro indique la racine de l'étiquetage ;
- colonne 7 : relation de dépendance partagée.

Par exemple, la seconde ligne concerne la deuxième occurrence de la phrase (colonne 1), qui est *collisions* (colonne 2) et dont le lemme est *collision* (colonne 3). Il s'agit d'un nom commun

(colonne 4) au féminin pluriel dans la phrase (colonne 5). Son gouverneur est l'unité *produire*, dont l'identifiant est 3 (colonne 6) ; la relation partagée entre les deux unités est *subj* (colonne 7), c'est-à-dire que *collision* est le sujet de *produire* dans cette phrase.

Enfin, outre l'utilisation relativement facilitée de Talismane grâce à son côté « prêt à l'emploi », les différentes annotations effectuées par Talismane représentent un atout. Ces annotations permettent en effet de cibler la recherche aussi bien sur une certaine relation de dépendance entretenue entre plusieurs unités que sur des unités relevant de certaines catégories grammaticales, les deux options pouvant également être aisément combinées. De ce point de vue, l'étiquetage effectué par Talismane ainsi que le format des fichiers de sortie de l'outil offrent une certaine flexibilité selon ce que l'on souhaite observer. Par ailleurs, ce format rend les résultats de l'analyseur aisément exploitables avec une base de données relationnelle, moyennant une conversion de ces fichiers au format .csv.

R et RStudio

R est un langage de programmation et un logiciel libre destiné à l'analyse statistique (R Core Team, 2018), qui fait partie des langages de programmation utilisés en linguistique (Levshina, 2015 ; Gries, 2016 ; Desagulier, 2017 ; Brezina, 2018 ; Egbert *et al.*, 2020). RStudio est un environnement de développement qui permet de gérer plus aisément et plus ergonomiquement les données et les scripts R (RStudio Team, 2019). Notre choix d'utiliser RStudio se fonde essentiellement sur l'ergonomie de cette interface, qui nous convient davantage que R seul.

R permet la mise en œuvre de méthodes statistiques puissantes et offre de nombreux packages destinés à l'analyse linguistique, où les chaînes de traitement des données sont « prêtes à l'emploi », ce qui représente à nos yeux un avantage non négligeable et facilite la mise en œuvre de ce langage. Par ailleurs, plusieurs ouvrages didactiques dédiés à l'analyse linguistique avec R permettent de prendre en main rapidement ce langage, notamment les ouvrages de Levshina (2015) et de Desagulier (2017), dont nous nous servons à plusieurs reprises et que nous citons dans plusieurs chapitres.

Plus précisément, nous nous servons de R et de RStudio afin de rendre certaines données issues d'extractions automatiques plus aisément comparables entre les sous-corpus, par exemple en transformant des listes brutes d'unités en tableaux de fréquences. De la même manière, les différents packages disponibles permettent de visualiser plusieurs données, permettant ainsi d'identifier certains phénomènes intéressants plus rapidement et plus aisément qu'à partir de listes de fréquences. Nous détaillons l'utilité de ces manipulations dans le Chapitre 6.

Enfin, nous exploitons R et RStudio pour mettre en œuvre différentes méthodes statistiques, notamment un calcul de l'accord inter-annotateurs, et pour extraire des fréquences d'unités automatiquement des sous-corpus. Dans ce dernier cas également, c'est la possibilité d'accélérer et de systématiser certains traitements qui rend l'utilisation de R et de RStudio attractive selon nous.

Ainsi, nous considérons R et RStudio comme des outils complémentaires aux autres outils précédemment cités, qui font désormais partie de la « boîte à outils » de l'analyste et qui permettent d'accompagner toujours plus la démarche d'analyse linguistique et, par extension, de tendre toujours plus vers une démarche objective et reproductible.

Outils génériques

Pour terminer, précisons que d'autres outils, plus génériques, sont utilisés tout au long de notre travail, aussi bien pour la constitution du corpus que pour son analyse. Plus particulièrement, nous utilisons :

- Notepad ++ pour traiter les données brutes et, dans une moindre mesure, les données étiquetées ;
- MS Excel pour gérer les données issues d'extractions automatiques, qui, même si ce n'est pas son utilité première, permet de filtrer aisément les données selon plusieurs critères ;
- le système de gestion de bases de données relationnelles MySQL, afin de gérer les corpus étiquetés par Talismane et d'extraire automatiquement les contextes syntaxiques des termes à l'étude à l'aide du langage de requête SQL (cf. Chapitre 5, 5.3.2).

Dans ce chapitre, nous nous sommes attachée à situer notre thèse dans une démarche outillée en corpus et nous avons particulièrement montré l'intérêt d'un travail sur des données attestées, qui est guidé par le recours à différents outils informatiques et méthodes statistiques. L'importance d'une telle démarche est cruciale pour aborder la question du fonctionnement des termes dans le processus de déterminologisation, dans le continuum entre langue de spécialité et langue générale : par la construction d'un cadre méthodologique solide et transparent, qui permet de contrôler un maximum de variables entrant *a priori* dans le processus de déterminologisation, nous sommes à même d'accéder au fonctionnement des termes dans différents textes prenant part à ce processus et de caractériser de manière fine et systématique les phénomènes repérés grâce à

différents indices. La construction de ce cadre contribue ainsi à inscrire notre travail de thèse dans une démarche scientifique systématique et reproductible.

Dans la suite de notre travail, nous nous focalisons sur la constitution de notre objet d'étude – le corpus et les termes à analyser – et sur la définition d'indices pertinents dans ce but.

Chapitre 4 Constitution de l'objet d'étude

Dans ce chapitre, nous abordons les questions méthodologiques spécifiques à la constitution de notre objet d'étude, selon l'angle d'analyse défini au Chapitre 3. Plus précisément, nous présentons notre réflexion sur le type de données à exploiter et la manière de les sélectionner, afin de permettre l'analyse du processus de déterminologisation en corpus comparable. Dans le même temps, nous insistons sur le besoin de transparence et d'objectivité de notre démarche, afin de mettre en œuvre une méthodologie d'analyse reproductible.

Différents choix sont nécessaires dans ce but, notamment au niveau du domaine, du corpus et de l'échantillon de termes à analyser. C'est pourquoi nous expliquons d'abord, dans la section 4.1, les raisons qui rendent le domaine de la physique des particules adéquat pour une analyse de la déterminologisation, puis, dans la section 4.2, nous exposons l'ensemble des critères mis en place afin de construire un corpus qui soit pertinent pour le domaine et pour observer le dynamisme du processus de déterminologisation dans le continuum entre langue de spécialité et langue générale. Enfin, nous détaillons dans la section 4.3 la méthode de sélection de l'échantillon de termes soumis à l'analyse, mise en place à la suite d'une réflexion visant l'objectivité et la reproductibilité de cette sélection.

4.1. Choix du domaine de la physique des particules

Dans le but d'analyser le fonctionnement de termes en situation de déterminologisation dans un domaine précis, il est nécessaire de délimiter un domaine dont les termes sont susceptibles de circuler en dehors de la sphère spécialisée et d'intégrer la langue générale. Dans cette section, nous argumentons notre choix du domaine de la physique des particules à l'aide de deux critères, développés à partir des travaux synthétisés dans les Chapitres 1 et 2 : la médiatisation du domaine et sa proximité avec le quotidien des locuteurs. Enfin, afin de situer notre travail et de délimiter le domaine pour lequel le corpus doit être pertinent, nous précisons ce que recouvre la physique des particules.

4.1.1. Pertinence de la physique des particules

4.1.1.1. Médiatisation du domaine

Le premier critère concerne le rôle des médias dans le processus de passage de termes dans la langue générale. En effet, nous avons vu au Chapitre 1 que les termes peuvent circuler en dehors

de la sphère spécialisée par différents types de médias, notamment des médias de vulgarisation (1.2.1) et des médias ordinaires (1.2.2), et donc intégrer la langue générale par le biais de différents canaux. Aussi, le domaine sur lequel nous nous focalisons doit être un domaine dont les termes sont susceptibles de figurer dans les médias (de vulgarisation et ordinaires). Dans ce contexte, la physique des particules est un domaine intéressant, car, au même titre que d'autres domaines de recherche scientifique, il est régulièrement traité dans ces types de médias.

Dans un premier temps, il est reconnu que la science, au sens le plus large, occupe une place privilégiée dans les médias, qu'il s'agisse de presse généraliste (Moirand, 2007 ; Costes et Laval, 2019 ; Gonon *et al.*, 2019), de quotidiens gratuits (Girardi, 2006) ou encore du journal télévisé (Deleu, 2006) par exemple. En effet, selon Moirand (*Ibid.* : 64), « [l]a presse ordinaire constitue [...] un lieu de rencontre des discours sur la science et sur ces événements ».

Tous les domaines de la recherche scientifique ne sont cependant pas couverts avec le même enthousiasme. Deleu (2006 : 79) oppose en effet « tout ce qui a trait à l'espace, l'astrophysique, l'astronomie, le biomédical (biologie avec application médicale, génétique, clonage), les nouvelles technologies, l'environnement » à d'autres domaines comme « la chimie, la biologie non médicale, les nanotechnologies, les mathématiques ». Les premiers auraient tendance à être privilégiés, tandis que les seconds seraient plus rarement traités. De la même manière, Beacco (1999 : 9) oppose les mathématiques à des domaines tels que la médecine, la sociologie, l'économie, l'histoire et les sciences de l'univers, ces derniers bénéficiant selon lui d'une « "forte" diffusion extérieure ». Dans ce contexte, la physique des particules étant un domaine proche de l'astrophysique (et donc des sciences de l'univers), il semble légitime d'affirmer que ce domaine fait également partie des domaines privilégiés par les médias.

Parallèlement, la notion de sensationnalisme peut aussi entrer en jeu dans le choix non seulement des sujets à traiter dans les médias, mais également de la manière dont ils sont traités (Labasse, 2012 ; Incelli 2018). Bien qu'il n'existe pas de définition claire et que cette notion « demeure étonnamment incertaine, tant dans ses origines que dans ce qu'elle dénote » (Labasse, 2012 : 115), il semble qu'elle renvoie à un ensemble de pratiques journalistiques, qui tendent à exploiter les aspects spectaculaires ou effroyables de certaines nouvelles, dans le but de capter l'attention du lectorat. Par exemple, selon Villedieu, le sensationnalisme consiste à « "gonfler" quelque peu la nouvelle scientifique », à lui « conférer plus de poids et de portée qu'elle n'en a réellement » (Villedieu, 1996 : 68). Pour cet auteur, cette pratique renvoie à la fascination que les journalistes éprouvent pour certains sujets scientifiques (*Ibid.* : 69), qui peut donc dépendre de la sensibilité de chaque journaliste. Dans le même ordre d'idée, Bernier, qui synthétise les propositions de plusieurs auteurs, explique que « le sensationnalisme consiste à insister trop fortement sur

l'aspect émotif et spectaculaire d'un événement » (Bernier, 2004 : 293), avec le risque de se laisser tenter par l'exagération et de ne pas rapporter fidèlement un fait (*Ibid.* : 295-297)⁷². Enfin, Labasse, qui discute abondamment la notion de sensationnalisme, propose de la définir selon l'impact émotionnel provoqué par un contenu sensationnaliste : « la composante émotionnelle, quelle que soit sa nature, sa source ou le procédé utilisé pour la susciter, semble bien constituer le dénominateur commun à toutes les acceptions du sensationnalisme » (*Ibid.* : 135). Ce type de contenu se situerait « aux deux extrémités du traitement de l'actualité que sont, d'un côté, les sujets réputés les plus malséants (les faits-divers) et, de l'autre, les plus relevés (les informations scientifiques) » (*Ibid.* : 114).

Cette notion de sensationnalisme est donc étroitement liée aux aspects spectaculaires de certaines découvertes scientifiques, que ces aspects soient réellement spectaculaires ou seulement perçus et relatés comme tels. Ainsi, du fait que « [l]es normes journalistiques [...] privilégient la nouveauté et le spectaculaire » (Gonon *et al.*, 2019 : 51), les découvertes scientifiques figurent parmi les sujets les plus traités dans les médias ordinaires (Beacco, 1999 : 129*sqq.*). De ce point de vue, la physique des particules représente un sujet de choix pour les journalistes, puisque ce domaine a vécu des développements majeurs ces dernières années, sur le plan théorique comme sur le plan expérimental, avec notamment la découverte du boson de Higgs en 2012. L'existence de cette particule a été postulée dans les années 1960⁷³ et il a fallu mettre au point « l'accélérateur de particules le plus grand et le plus puissant du monde »⁷⁴, le LHC (Large Hadron Collider, Grand collisionneur de hadrons), pour en démontrer l'existence. L'observation du boson de Higgs au LHC au CERN (Organisation européenne pour la recherche nucléaire) a donc été rendue possible environ 50 ans après sa prédiction par la théorie.

Dans ce contexte, il semble évident que tant la construction et la mise en marche du LHC que la recherche et la découverte du boson de Higgs entrent dans cette idée de nouveau et de spectaculaire d'un point de vue journalistique. Ces sujets sont donc susceptibles d'être traités dans une perspective sensationnaliste dans les médias, ce que suggèrent notamment Labasse (2012 : 114) et Incelli (2018 : 193*sqq.*), et, plus largement, d'y apparaître régulièrement.

⁷² Pour cet auteur, toutefois, le sensationnalisme ne se rapporte pas uniquement aux faits de science, mais peut concerner d'autres types de faits. À ce propos, il donne l'exemple du traitement par les médias de la mort de la Princesse Diana, où la rareté des informations disponibles immédiatement et leur manque de fiabilité auraient poussé les journalistes à exagérer certains faits (Bernier, 2004 : 296).

⁷³ Le boson de Higgs, CERN, <https://home.cern/fr/science/physics/higgs-boson> (page consultée le 29 mars 2021).

⁷⁴ Grand collisionneur de hadrons, CERN, <https://home.cern/fr/science/accelerators/large-hadron-collider> (page consultée le 29 mars 2021).

Dans un second temps, la physique des particules est un domaine pour lequel il existe de nombreuses publications de vulgarisation. Comme nous le disons au Chapitre 1 (section 1.2.1), la vulgarisation scientifique a pour objectif de transmettre des connaissances spécialisées à un public non spécialiste. Ainsi, tout domaine scientifique est susceptible d'être couvert par des médias de vulgarisation. Cependant, nous pensons que la physique des particules jouit d'un statut privilégié, qui est largement dû à l'importance mondiale du CERN. Cette organisation semble en effet redoubler d'efforts pour rendre accessibles les résultats des recherches qui y sont menées, soit directement soit par le biais des nombreux laboratoires partenaires.

Par exemple, le CERN organise des journées portes ouvertes, tous les cinq ans environ⁷⁵, ainsi que des événements TED (Technology, Entertainment and Design)⁷⁶, qui ont lieu presque chaque année depuis 2013⁷⁷. Le site de cette organisation possède également une version grand public, qui consacre plusieurs rubriques aux concepts fondamentaux du domaine et aux différentes expériences qui y sont menées⁷⁸. Dans certains cas, la vulgarisation prend même une forme un peu plus ludique, comme dans *La BD du LHC*, disponible sur le site LHC-France, lequel détaille la contribution française dans la recherche au LHC⁷⁹.

En outre, nombre de revues de vulgarisation qui ne sont pas spécialisées dans un domaine en particulier, comme *Pour la Science*, *La Recherche*, *Science & Vie*, etc., publient des articles sur la physique des particules, selon l'actualité du domaine. D'autres canaux, spécifiques à la physique des particules, participent également à la transmission des connaissances. Le projet *Élémentaire* en constitue un bon exemple : son but premier est de « faire connaître au plus grand nombre la physique des particules »⁸⁰, par le biais d'une revue, qui explique tous les aspects de la recherche en physique des particules, ainsi que de jeux éducatifs, classiques ou sur ordinateur. Enfin, certains chercheurs œuvrent eux-mêmes pour la vulgarisation du domaine. Citons à ce propos Etienne Klein, auteur notamment du livre *Les secrets de la matière*⁸¹, et Pauline Gagnon, autrice de l'ouvrage *Qu'est-ce que le boson de Higgs mange en hiver et autres détails essentiels*⁸². Ces deux scientifiques ne sont d'ailleurs pas avares de conférences destinées au grand public.

⁷⁵ CERN Open Days, <https://opendays.cern/fr> (page consultée le 29 mars 2021).

⁷⁶ TED, <https://www.ted.com/> (page consultée le 29 mars 2021).

⁷⁷ TEDxCERN, <https://tedxcern.web.cern.ch/fr> (page consultée le 29 mars 2021).

⁷⁸ CERN, <https://home.cern/fr> (page consultée le 29 mars 2021).

⁷⁹ LHC-France, *La BD du LHC*, <http://www.lhc-france.fr/spip.php?rubrique65> (page consultée le 29 mars 2021).

⁸⁰ Élémentaire, <https://elementaire.lal.in2p3.fr/> (page consultée le 29 mars 2021).

⁸¹ Etienne Klein, <https://etienneklein.fr/> (page consultée le 29 mars 2021).

⁸² Pauline Gagnon, <https://paulinegagnon3.wixsite.com/boson-en-hiver> (page consultée le 29 mars 2021).

Au vu de ces éléments, il semble clair que ce domaine bénéficie d'une importante couverture médiatique, dans des médias ordinaires comme dans des supports de vulgarisation, ce qui le rend particulièrement pertinent pour une analyse de la déterminologisation.

4.1.1.2. Proximité avec les locuteurs

Le second aspect dont il faut tenir compte concerne l'intérêt supposé du grand public pour certains domaines scientifiques et, par extension, pour certains des termes qui appartiennent à ces domaines. C'est ce que Meyer et Mackintosh (2000a : 199) expliquent lorsqu'elles affirment que le phénomène de déterminologisation « se produit lorsqu'un terme attire l'attention du public ».

Cet aspect est bien évidemment étroitement lié au premier aspect discuté, puisque l'on peut supposer qu'un domaine qui figure régulièrement dans les médias ordinaires est un domaine qui intéresse le grand public. Mais un domaine peut également attirer l'attention par l'importance que certains des concepts qui lui sont rattachés prennent dans la vie quotidienne des locuteurs. L'idée sous-jacente, partagée notamment par Guilbert (1975) et Meyer (2000), est que les locuteurs qui sont régulièrement confrontés à certains termes auraient plus de chances de se les approprier et de les reprendre ensuite à leur guise, favorisant ainsi l'intégration de ces termes dans la langue générale. Abordant cette question du point de vue de la néologie, Guilbert affirme que « le processus de métaphorisation des termes techniques » (Guilbert, 1975 : 84), qui renvoie à la possibilité pour un terme d'être considéré comme un néologisme sémantique en langue générale, si le « nouveau sens » qui y est constaté se fonde sur une métaphore (cf. Chapitre 1, section 1.4),

« est intimement lié à la vulgarisation des sciences et des techniques **qui sont mêlées étroitement à la vie quotidienne de chaque locuteur**, comme l'étaient autrefois les animaux à la vie des gens dans une société à dominante paysanne. » (Guilbert, 1975 : 84)⁸³

Meyer exprime un point de vue complémentaire :

« some computer words de-terminologize [...] [p]rimarily for the obvious reason that computers are central to our everyday lives. In the words of Ayto [1999:iv], the lexicon is "a mirror of our times". We cannot help adopting words such as *surfing*, *downloading*, *desktop*, and *email* because we are faced with these concepts daily. » (Meyer, 2000 : 46)

Le domaine choisi doit donc démontrer une certaine proximité avec le quotidien des locuteurs non nécessairement spécialistes de ce domaine. Nous pensons que la physique des particules satisfait à ces conditions, pour plusieurs raisons.

⁸³ Nous soulignons.

Tout d'abord, précisons que, s'il est difficile d'imaginer que la physique des particules fasse partie de notre vie quotidienne, le CERN, en revanche, y occupe une place particulière. En effet, les infrastructures du CERN étant situées à la frontière franco-suisse, nous pouvons raisonnablement penser qu'elles pourraient susciter l'intérêt des personnes habitant dans cette même région⁸⁴. De la simple curiosité pour la recherche, pour « ce qui se fait » dans ces bâtiments, mais peut-être également quelques inquiétudes provoquées par les expériences conduites dans ces locaux, du moins par la méconnaissance de ces expériences. Cette hypothèse reprend en réalité celle de Delavigne à propos de la représentation du nucléaire dans la presse régionale. Selon elle, en effet, les journaux régionaux « peuvent être le reflet de préoccupations d'une région qui possède deux centrales nucléaires » (Delavigne, 2001 : 267-268). La présence de centrales nucléaires dans une région aurait ainsi une influence dans le traitement du nucléaire par la presse locale, les articles sur ce sujet seraient par exemple plus abondants. Pour ce qui est de la physique des particules, la proximité du CERN avec la population alentour pourrait donc renforcer les efforts de communication de la part de l'organisation, en particulier avec les médias d'information locaux. De la même manière, la prise en compte des inquiétudes potentielles de la population pourrait également passer par ces médias.

Ensuite, si le CERN est probablement l'organe de recherche le plus important à l'échelle mondiale, ou du moins européenne, d'autres laboratoires contribuent à la recherche en physique des particules. L'IN2P3 (Institut national de physique nucléaire et de physique des particules) en recense par exemple 16 en France (qui sont pour la plupart partenaires du CERN)⁸⁵. Nous pensons alors que les canaux par lesquels les termes pourraient se diffuser auprès des locuteurs non spécialistes sont multipliés et que la prise en compte des préoccupations du grand public de la part des organes de recherche encourage la communication sur le domaine.

La physique des particules constitue donc bien un domaine pertinent pour notre étude de la déterminologisation. Il nous semble à présent nécessaire de définir et circonscrire précisément ce domaine, notamment dans une optique de constitution de corpus, dans le but de développer des critères de sélection des textes pertinents pour ce domaine.

⁸⁴ Par ailleurs, la présence du CERN dans deux pays francophones représente un atout considérable pour une étude en corpus en français, car la majorité, sinon l'ensemble des documents émanant de cette organisation sont rédigés directement en français ou traduits dans cette langue, constituant ainsi une source de données textuelles en français non négligeable.

⁸⁵ IN2P3, <https://in2p3.cnrs.fr/fr/recherche> (page consultée le 29 mars 2021).

4.1.2. Précisions sur la physique des particules

De manière générale, la physique des particules est définie comme un domaine de la physique qui étudie les particules et leurs interactions (Taillet *et al.*, 2018 : 564). Afin de décrire plus clairement ce domaine, il est nécessaire de préciser les relations qu'il entretient avec d'autres domaines de la physique, notamment la physique atomique et la physique nucléaire. Les informations que nous synthétisons dans ce but proviennent en grande partie des différents entretiens que nous avons menés auprès des expertes tout au long de notre travail de thèse. Plusieurs aspects ont été vérifiés dans les textes du corpus et dans le *Dictionnaire de physique* (Taillet *et al.*, 2018)⁸⁶ *a posteriori*.

La physique des particules entretient des rapports étroits avec la physique nucléaire et la physique atomique, notamment en raison de la proximité de leurs objets d'étude respectifs⁸⁷. Par ailleurs, bien que ceux-ci soient clairement définis dans les textes et dans les ressources, ils restent liés les uns aux autres, non seulement dans la réalité, mais également dans la recherche⁸⁸.

Observons d'abord les définitions de ces trois champs de recherche, avec une emphase sur les objets d'étude. Le *Dictionnaire de physique* définit ces trois termes comme suit (Taillet *et al.*, 2018 : 564)⁸⁹ :

- *physique atomique* : « [d]omaine de la physique s'intéressant à la **structure des atomes**, en particulier à leurs niveaux d'énergie électroniques et à leur interaction avec la lumière » ;
- *physique nucléaire* : « [d]omaine de la physique qui s'intéresse aux **propriétés des noyaux atomiques** et à leurs transformations (radioactivité, fusion, fission, spectroscopie), dans les situations où l'on considère qu'ils sont constitués de nucléons (la prise en compte de la structure en quarks est l'objet de la physique des particules) » ;

⁸⁶ Il s'agit du dictionnaire de physique le plus récent et le plus régulièrement mis à jour que nous ayons trouvé. La dernière édition date de 2018.

⁸⁷ Pour un historique détaillé de la terminologie de la physique (nucléaire, atomique et des particules), voir Candel (1995), Delavigne (2001), Ahmad et Musacchio (2003).

⁸⁸ Notons également que les termes *atomique* et *nucléaire* sont parfois confondus et utilisés l'un pour l'autre, notamment par les non-experts, comme le soulignent Calberg-Challot *et al.* (2006 : 232). Cette confusion s'explique principalement par des raisons historiques, liées non seulement à la militarisation et à l'industrialisation de l'énergie nucléaire, mais aussi au développement des connaissances sur l'atome et sur ses propriétés (Calberg-Challot *et al.*, 2006 ; Delavigne, 2006). En conséquence, en plus du fait que leurs objets d'étude sont très proches, les domaines de la physique atomique et de la physique nucléaire sont parfois confondus. Cette confusion ressort également de nos entretiens avec les expertes. En effet, si pour l'une des expertes il s'agit bien de deux branches de la physique, pour la seconde, ces deux termes sont en réalité synonymes et la distinction entre *physique atomique* et *physique nucléaire* serait plutôt de nature historique.

⁸⁹ Nous soulignons.

- *physique des particules* : « [d]omaine de la physique qui **décrit les particules élémentaires et leurs interactions**, d'un point de vue théorique et expérimental. [...] Une grande partie de la recherche du domaine s'effectue dans le cadre du modèle standard de la physique des particules [...] ».

Les différences entre ces trois domaines de la physique sont en apparence tout à fait claires du fait que l'objet de chaque domaine est spécifié (atomes, noyaux, particules élémentaires), ce que nous retrouvons par ailleurs dans notre corpus, comme dans cet exemple⁹⁰.

1. À partir du 18e siècle (et jusqu'à nos jours), cette approche réductionniste de la matière a progressivement donné naissance à l'étude des molécules (la chimie), **des atomes (la physique atomique), des noyaux (la physique nucléaire) et de leurs constituants (la physique des particules)**. (SC-V, Élémentaire, 2016)

Or, s'il semble clair que ces trois domaines s'intéressent à des objets tout à fait distincts, dès lors que l'on prend en compte la composition d'un atome, on se rend compte que, dans la réalité, ces objets sont en fait liés. La figure 4.1 illustre parfaitement les rapports entre ces trois domaines (image Pixabay, libre de droits)⁹¹. L'atome de lithium qui y est représenté est composé d'un noyau (au centre) et d'électrons qui orbitent autour du noyau. Selon les définitions citées ci-dessus, l'atome dans son ensemble constitue l'objet d'étude de la physique atomique, et le noyau pris isolément celui de la physique nucléaire. En outre, les électrons étant des particules élémentaires, ils sont étudiés dans le champ de la physique des particules.

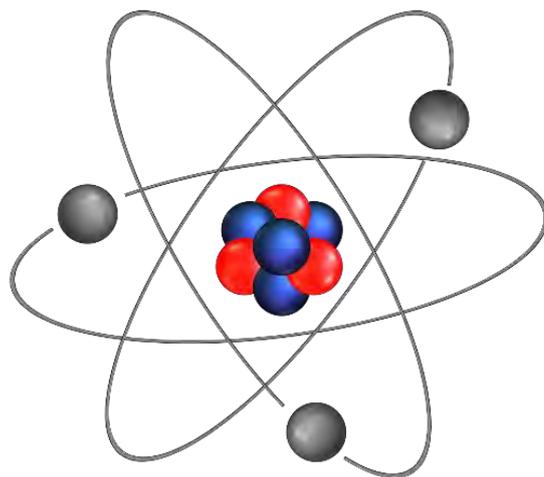


Figure 4.1 : Illustration d'un atome

⁹⁰ Nous soulignons

⁹¹ Cette image est libre de droits et a été téléchargée sur Pixabay (<https://pixabay.com/fr/>) le 5 septembre 2019.

Ces rapports entre les différentes composantes de l'atome et l'atome lui-même expliquent en partie la confusion qui existe parfois entre ces trois domaines de la physique. Mais les différences sont encore plus subtiles. En effet, le noyau est composé de protons et de neutrons, qui sont eux-mêmes constitués de quarks. Les quarks ne sont pas composés de plus petits éléments, ce sont donc des particules élémentaires. Par ailleurs, ils sont liés entre eux par des particules appelées *gluons*, qui sont également des particules élémentaires. Les quarks et les gluons font donc l'objet de recherches en physique des particules. En revanche, étant donné que les protons et les neutrons forment le noyau de l'atome, ils sont principalement étudiés dans le champ de la physique nucléaire.

En outre, pour permettre l'étude de certaines particules élémentaires, notamment que le boson de Higgs, celles-ci doivent être issues de collisions entre d'autres particules, comme des protons, qui sont des particules composites (c'est-à-dire des particules qui sont composées de particules élémentaires). Par conséquent, les particules composites jouent un rôle important en physique des particules. Le rôle de ces particules participe au flou qui peut entourer la délimitation précise de la physique des particules, ce qui a également été mis en avant par les expertes lors d'entretiens. En effet, pour l'une des expertes, ce domaine s'intéresse aux particules élémentaires uniquement, tandis que, pour l'autre experte, les particules composites font partie intégrante de ce champ d'études. S'il n'est pas sûr que cette position domine, elle offre l'avantage de reconnaître clairement la place qu'occupe ce type de particules dans la recherche en physique des particules.

La physique des particules est donc un domaine de la physique qui étudie les particules élémentaires et leurs interactions, et dont les activités de recherche font intervenir les particules composites. La délimitation de ce domaine représente une étape dans la constitution de notre objet d'étude, qui est nécessaire pour construire un corpus qui soit pertinent pour observer le processus de déterminologisation pour les termes qui appartiennent à ce domaine. Mais le domaine n'est pas le seul critère dont nous devons tenir compte pour constituer le corpus. C'est pourquoi nous poursuivons notre réflexion à ce propos dans la section suivante.

4.2. Constituer un corpus adéquat pour l'analyse de la déterminologisation

Nous l'avons dit au Chapitre 3, le cadre d'analyse que nous développons se base sur une analyse de corpus, lequel se trouve au cœur de la démarche. Sa constitution revêt donc une importance considérable. Les critères de constitution choisis doivent en effet être le résultat d'une réflexion centrée sur l'objet d'étude et sont fonction du but poursuivi. Nombre de chercheurs travaillant à partir de corpus reconnaissent l'importance de la cohérence du corpus pour l'objectif

de recherche, par exemple Engwall (1994 : 54), Biber *et al.* (1998 : 246), Habert *et al.* (1998 : 35), Kennedy (1998 : 70), Hunston (2008 : 154), Nelson (2010 : 58), Poudat et Landragin (2017 : 12*sqq.*), et en particulier dans un contexte de langues de spécialité et de terminologie. Par exemple, Bowker et Pearson affirment que les corpus sont des « collections that have been put together according to specific criteria. These criteria are determined by your needs and by the goal of your project » (Bowker et Pearson, 2002 : 45) et Condamines explique que « [l]a nature du travail que l'on souhaite mener joue un rôle déterminant sur la nature du corpus que l'on va constituer » (Condamines, 2003 : 43).

Dans notre cas, nous poursuivons le but d'analyser le phénomène de déterminologisation en tant que processus, dans un domaine précis et dans le continuum entre langue de spécialité et langue générale. Le corpus doit donc permettre d'interroger cet aspect continu du processus dans l'analyse et les textes inclus doivent être pertinents pour le domaine étudié. Dans ce but, nous discutons d'abord deux critères fondamentaux entrant dans la constitution d'un corpus – la représentativité et l'équilibre du corpus – puis nous présentons la procédure de constitution du corpus mise en œuvre, qui vise l'objectivité et la reproductibilité de cette étape⁹².

4.2.1. Représentativité et équilibre : deux principes à nuancer

La représentativité et l'équilibre d'un corpus sont deux caractéristiques généralement présentées comme essentielles par des auteurs traitant de la constitution de corpus, qu'il s'agisse d'un corpus de référence, où l'on cherchera la représentativité de la langue dans son ensemble, ou d'un corpus de langue de spécialité, où l'on cherchera la représentativité de cette langue de spécialité en particulier (Sinclair, 1991 ; Atkins *et al.*, 1992 ; Biber, 1993 ; Meyer et Mackintosh, 1996 ; Kennedy, 1998 ; Habert, 2000 ; Tognini-Bonelli, 2001 ; Leech, 2007 ; Nelson, 2010 parmi d'autres). Cependant, définir ces concepts reste une tâche ardue, car les auteurs ne sont pas toujours d'accord entre eux, que ce soit à propos de la définition de la représentativité d'un corpus ou à propos de la manière de parvenir à un corpus qui soit véritablement représentatif et équilibré. Le débat perdure depuis plusieurs dizaines d'années et nous n'avons pas la prétention de le clore ici, mais certains points méritent d'être soulignés.

En premier lieu, la représentativité d'un corpus constitue une condition nécessaire pour rendre possible la généralisation des observations effectuées sur ce corpus. Pour Leech (1991 : 11), par exemple, « a corpus is 'representative' to the extent that findings based on its contents can be

⁹² La procédure de constitution du corpus détaillée dans cette section a fait l'objet d'une publication dans (Humbert-Droz *et al.*, 2019).

generalized to a larger hypothetical corpus ». Cette définition fait en réalité référence à un principe d'échantillonnage selon lequel, dans la mesure où il est impossible d'analyser tout un ensemble de phénomènes, il est nécessaire de sélectionner certains échantillons de cet ensemble, échantillons qui doivent être représentatifs. Cette condition est présentée comme essentielle pour que les résultats tirés de l'échantillon soient applicables à l'ensemble. Selon Biber,

« a thorough definition of the target population and decisions concerning the method of sampling are prior considerations. Representativeness refers to the extent to which a sample includes the full range of variability in a population. »
(Biber, 1993 : 243)

Ainsi, d'un point de vue théorique, un corpus représentatif, par exemple d'une certaine langue, est donc un corpus composé d'échantillons de cette langue qui ont été sélectionnés avec soin et à partir duquel des généralisations à propos de cette langue dans son ensemble sont possibles. Leech précise par ailleurs que, si tel n'est pas le cas, « whatever is found to be true of a corpus, is simply true of that corpus – and cannot be extended to anything else » (Leech, 2007 : 136). Dans cette perspective, la définition de la population cible (*target population*, pour reprendre les termes de Biber), c'est-à-dire de la langue ou de la variété de langue choisie, est donc une étape cruciale dans la constitution d'un corpus. Et c'est précisément sur ce point que le concept de représentativité pose problème.

En effet, ce point de vue suppose qu'il est possible de définir clairement la variété de la langue que l'on souhaite étudier, qu'il s'agisse de la langue générale dans le cas d'un corpus de référence ou d'une langue de spécialité d'un domaine en particulier dans le cas d'un corpus spécialisé. Or, du fait que la langue est en constante évolution et que le découpage strict en domaines ne reflète pas nécessairement la réalité⁹³, la définition de l'ensemble étudié, pour en sélectionner un échantillon représentatif, semble peu réaliste⁹⁴. Atkins et Rundell (2008 : 64) considèrent d'ailleurs que cette méthode d'échantillonnage ne convient pas à la langue, du fait que « it is difficult to define what the total population is » et que « the population is continually growing ». Plus loin, les auteurs ajoutent que « [a] truly representative corpus is an impossible goal because we are sampling from a population whose nature is unknowable and whose extent is unlimited » (*Ibid.* : 74-75).

⁹³ Rondeau (1981 : 26) affirme par exemple que les « frontières [entre domaines] sont perméables, notamment dans des domaines connexes comme par exemple ceux des sciences de la santé, ou encore ceux de la chimie-physique ». Dans la même idée, Meyer et Mackintosh (1996 : 268) ainsi que Bowker et Pearson (2002 : 50) mentionnent la difficulté à circonscrire un domaine, en particulier lorsque ce domaine est interdisciplinaire ou multidisciplinaire.

⁹⁴ En outre, comme nous l'avons discuté au Chapitre 1, le concept de langue générale reste difficile à définir et peu opérationnel pour la constitution d'un corpus de référence.

Condamines (2005a : 41) évoque également ce problème, en insistant sur le fait qu'il est impossible de s'assurer qu'un corpus représente réellement l'ensemble qu'il est censé représenter :

« s'il s'agit de décrire le fonctionnement tel qu'il se manifeste dans les corpus étudiés, il faut donc que ces corpus, puisqu'ils ne peuvent rendre compte de tous les usages, soient au moins représentatifs de tous ces usages. Or, non seulement, nul n'a les moyens de vérifier cette représentativité mais, et c'est plus ennuyeux, on voit nécessairement réapparaître l'introspection, que l'on pensait évacuer par le recours aux corpus. »

En second lieu, différents auteurs considèrent que la représentativité d'un corpus est à conceptualiser en fonction de la qualité du matériel inclus dans le corpus.

« Le potentiel de représentativité d'un corpus est alors étroitement lié à sa diversité et à la qualité de sa documentation. » (Péry-Woodley, 1995 : 18)

« Issues associated with how to make a corpus representative or balanced or able to be used for comparative purposes are essentially issues concerning the quality of a corpus. » (Kennedy, 1998 : 66)

Dans ce contexte, l'idée de qualité ne renvoie pas nécessairement à la qualité d'un texte en tant que tel (supposé bien écrit, bien structuré, etc.), mais aussi – et surtout – à l'adéquation d'un texte en particulier par rapport à l'ensemble visé, au phénomène linguistique étudié, au domaine sélectionné et à l'objectif de recherche pour lequel le corpus est constitué.

Dans certains cas, la taille d'un corpus a été mise en avant pour pallier le problème de la représentativité (et, par là, de la qualité des textes composant le corpus). Habert (2000) explique que cette position repose sur

« [l]a conviction [...] que l'élargissement mécanique des données mémorisables [...] produit inévitablement un échantillon de plus en plus représentatif de la langue traitée. Si l'on n'arrive pas à cerner précisément les caractéristiques de l'ensemble des productions langagières, il ne reste qu'à englober le maximum d'énoncés possibles. »

De ce point de vue (aussi mentionné par Leech (2007 : 6) ou Pearson (1998 : 51)), plus les données recueillies dans un corpus sont volumineuses, plus le corpus est censé s'approcher de l'ensemble linguistique visé. Cette idée est également défendue par les chercheurs prônant l'utilisation du Web comme corpus : vu la quantité des données textuelles disponibles sur le Web et leur diversité, leur prise en compte constitue nécessairement un avantage pour tout type de recherches en linguistique. C'est du moins à cette conclusion que parviennent notamment Kilgarriff et Grefenstette (2003).

Enfin, si l'utilisation de corpus toujours plus vastes est parfois revendiquée comme un moyen de s'approcher de la représentativité, cette position n'est pas toujours partagée. Kennedy remet clairement cette idée en question, en insistant, lui aussi, sur l'importance de la qualité des textes :

« [a] huge corpus does not necessarily 'represent' a language or a variety of a language any better than a smaller corpus. At this stage, we simply do not know how big a corpus needs to be for general or particular purposes. Rather than focusing so strongly on the quantity of data in a corpus, compilers and analysts need also to bear in mind that the quality of the data they work with is at least as important. » (Kennedy, 1998 : 68)

Ainsi, si la taille d'un corpus reste un critère essentiel dont il faut tenir compte pour constituer le corpus et pour évaluer sa représentativité, ce critère ne devrait pas supplanter la qualité des textes sélectionnés.

En troisième lieu, la question de l'équilibre d'un corpus se trouve parfois mise en avant, soit pour sa complémentarité avec le critère de représentativité, soit pour contourner ce critère, du fait que celui-ci reste, malgré tout, peu opérationnel : « [t]he word *representative* has tended to fall out of discussions, to be replaced by the meeker *balanced* » (Kilgarriff et Grefenstette, 2003 : 342). D'après Leech,

« a corpus is "balanced" when the size of its subcorpora (representing particular genres or registers) is proportional to the relative frequency of occurrence of those genres in the language's textual universe as a whole. » (Leech, 2007 : 137)

Cette définition de l'équilibre d'un corpus n'est pas sans poser problème. Elle suppose en effet que ce que Leech désigne comme étant « the language's textual universe as a whole » peut être défini. Or, comme nous l'avons vu à propos du concept de représentativité, il semble impossible de définir précisément de quoi est composée la langue. Par ailleurs, même si les propos de Leech renvoient à la langue générale et aux corpus de référence, nous pouvons raisonnablement supposer que les mêmes conclusions peuvent être tirées pour un corpus spécialisé. Il semble en effet ardu d'estimer l'ensemble des productions textuelles d'une langue de spécialité, d'autant plus si l'on tient compte des différents degrés de spécialisation existant au sein d'une même langue de spécialité.

Cette notion d'équilibre se rapproche donc de la notion de représentativité, tant par l'idéal auquel elle renvoie que par l'impossibilité de parvenir à cet idéal. Son manque d'opérationnalité est particulièrement bien mis en évidence par Péry-Woodley (1995 : 219) :

« la recherche de corpus équilibrés semble bien constituer une impasse : la notion d'équilibre s'apparente à celle de "langue générale", et elle paraît tout aussi insaisissable. Elle suppose également une recherche irréaliste d'exhaustivité : le

corpus équilibré est sans doute celui qui a "de tout un peu", mais encore faudrait-il savoir ce qu'est "tout", c'est-à-dire quelles sont les classes à représenter. »

Enfin, citons Atkins *et al.* (1992 : 6), qui remettent sérieusement en doute la possibilité – et même la nécessité – de parvenir à un corpus parfaitement équilibré :

« [a]t present, corpus 'balance' relies heavily on intuition, although work on text typology is highly relevant. It is not our purpose to lay down a methodology for 'balancing' a corpus, for we believe that this is not something that can be done in advance of the corpus-building process, if it can be done at all. [...] Knowing that your corpus is unbalanced is what counts. It would be short-sighted indeed to wait until one can scientifically balance a corpus before starting to use one, and hasty to dismiss the results of corpus analysis as 'unreliable' or 'irrelevant' simply because the corpus used cannot be proved to be 'balanced'. »

À la lueur de ces éléments, nous considérons que la notion d'équilibre pose les mêmes problèmes pratiques que la représentativité elle-même.

En dernier lieu, pour toutes les raisons présentées dans cette section, il semble qu'aucune stratégie efficace n'ait encore été développée pour constituer un corpus qui soit véritablement représentatif de l'usage visé. La représentativité soulève encore en effet de nombreuses questions, et l'idée d'équilibre qui y est associée reste tout aussi peu opérationnelle. Dans ce contexte, nous considérons, à l'instar de Prévost (2005 : 156), Leech (2007 : 144-145), Hunston (2008 : 161) Nelson (2010 : 60) ou encore McEnery et Hardie (2012 : 10), que la représentativité est un objectif vers lequel tendre. Nous soulignons ainsi la nécessité pour un corpus d'être effectivement constitué en accord avec l'objectif d'analyse visé, afin que les observations effectuées puissent raisonnablement conduire à des conclusions qui soient tout de même valides. Par ailleurs, afin de nous approcher de cet « idéal » de représentativité, nous jugeons qu'il est préférable de recourir à une méthode de sélection des textes qui soit systématique et éclairée, comme le recommande par exemple Hunston (2008 : 162). En d'autres termes, nous estimons que la constitution du corpus doit suivre un certain nombre de principes, développés en fonction du but de l'analyse. C'est pourquoi, dans la section suivante, nous détaillons les trois principes qui guident la sélection des textes constitutifs de notre corpus, afin que celui-ci soit pertinent par rapport au processus de déterminologisation et adéquat pour le type d'analyses que nous visons.

4.2.2. Trois principes de constitution pour garantir la pertinence du corpus

La constitution d'un corpus adéquat pour observer le processus de déterminologisation dans le domaine de la physique des particules passe par différentes étapes de réflexion sur les possibilités existantes de récolte et d'organisation des données ainsi que de sélection des textes.

Dans ce but, nous développons trois principes de constitution, qui visent également à limiter autant que possible l'introduction de biais dans le choix des textes.

4.2.2.1. Représenter un continuum en corpus par l'inclusion de genres textuels diversifiés

Le premier principe vise d'une part à identifier les textes pertinents pour observer la déterminologisation en corpus, à l'aide de différentes caractéristiques, notamment les genres textuels dont relèvent les textes et le degré de spécialisation de la communication auquel ils correspondent, et, d'autre part, à organiser les données de manière à ce que la continuité du processus de déterminologisation puisse être interrogée.

4.2.2.1.1. Identifier des genres textuels pertinents

L'observation en corpus de la déterminologisation en tant que processus continu entre langue de spécialité et langue générale implique de constituer un corpus non seulement à partir de textes d'un degré de spécialisation élevé – c'est-à-dire des textes produits par les experts et pour les experts (Bowker et Pearson, 2002 : 28) – et de textes de langue générale (ou non spécialisés, comme nous l'expliquons *infra*), mais également à partir de textes dont le degré de spécialisation se situe entre ces deux pôles. Afin d'identifier les textes à sélectionner et de caractériser leur pertinence pour l'observation de la déterminologisation dans le domaine de la physique des particules, nous nous servons de la notion de genre textuel.

De manière générale, cette notion renvoie à l'idée que les textes produits dans certaines situations de communication partagent certaines caractéristiques, qui permettent de les regrouper et qui les distinguent d'autres textes. Cependant, deux points de vue au moins existent. D'une part, Biber introduit une distinction entre ce qu'il appelle *genre* et *text type*, insistant sur le type de caractéristiques permettant d'effectuer des regroupements, selon si elles sont d'ordre linguistique ou extralinguistique :

« I use the term 'genre' to refer to categorizations assigned on the basis of external criteria. I use the term 'text type', on the other hand, to refer to groupings of texts that are similar with respect to their linguistic form, irrespective of genre categories. » (Biber, 1988 : 70)

D'un autre côté, un genre peut être défini en fonction de caractéristiques à la fois linguistiques et extralinguistiques. C'est notamment le point de vue de Bhatia, pour qui la notion de genre

« essentially refers to language use in a conventionalized communicative setting in order to give expression to a specific set of communicative goals of a disciplinary or social institution, which give rise to stable structural forms by imposing constraints on the use of lexico-grammatical as well as discoursal resources. » (Bhatia, 2004 : 23)

Dans cette perspective, les textes peuvent être regroupés en genres non seulement en fonction de règles déterminées par la situation de communication, mais également car ils partagent certaines régularités linguistiques. Par ailleurs, la prise en compte des aspects extralinguistiques a l'avantage de rendre la catégorisation en genres plus opérationnelle que la seule prise en compte des régularités linguistiques des textes. En effet, cette seconde option devrait déjà être le fruit d'une analyse, à partir de textes présélectionnés. Or, tout l'enjeu réside dans cette présélection : sur quels critères ces textes devraient-ils être choisis ? À partir de régularités linguistiques supposées ? Ou qui proviennent également d'une analyse préliminaire ? Le caractère cyclique évident de cette démarche la rend peu opérationnelle pour les objectifs de notre travail.

Parallèlement, soulignons la difficulté à établir une typologie des genres qui soit parfaitement exhaustive (Pearson, 1998 : 53 ; Condamines, 2003 : 36-37). En effet, tout comme la langue, les genres sont en constante évolution (Bowker, 1996 : 45 ; Condamines, 2003 : 35 ; Leech, 2007 : 143) et, en langues de spécialité, ils peuvent également varier en fonction du domaine et du degré de spécialisation en jeu (Meyer et Mackintosh, 1996 : 270). En outre, cette complexité à catégoriser les genres s'accroît dès lors que les genres du Web sont également pris en compte, en plus des genres « traditionnels » (presse, roman, article scientifique, etc.). Les genres du Web ne font d'ailleurs pas non plus consensus, comme Daille (2017 : 84) le résume : « [t]here are several typologies of web genres, from the most concise (8 genres) to the most exhaustive (2,000 genres) ».

Face à ces difficultés d'ordre théorique et pratique sur la question du genre textuel, nous reprenons à notre compte le constat de Pearson (1998 : 53) : « each corpus project tends to have its own method of classifying genre ». Dans ce contexte, nous cherchons à exploiter la notion de genre textuel pour identifier des textes adéquats pour observer le processus de déterminologisation en corpus à partir de caractéristiques extralinguistiques pertinentes dans ce but. Pour caractériser les genres, nous nous focalisons ainsi sur le statut des locuteurs, notamment leur degré d'expertise, et le potentiel des textes de diffuser des connaissances (et des termes liés), et donc de contribuer au processus d'intégration des termes dans la langue générale. De plus, même si nous nous servons principalement de caractéristiques extralinguistiques, nous nous attendons à observer des fonctionnements linguistiques différents pour les termes, selon le genre pris en compte. Comme nous l'expliquons au Chapitre 2 (2.4), cet aspect revêt une importance particulière, puisque nous cherchons à interroger le rôle des textes intermédiaires dans le processus de déterminologisation.

Notre réflexion s'articule en deux temps : dans un premier temps, afin d'inclure des textes qui permettent de représenter les deux extrémités du continuum, il est nécessaire d'identifier des textes pertinents qui relèvent d'un degré de spécialisation élevé et des textes pertinents qui relèvent de la langue générale. Dans le premier cas, nous tenons compte de textes relevant d'un degré de spécialisation expert-expert, selon la classification de Bowker et Pearson (2002 : 28). Pour les autrices (*Ibid.*) et pour Loffler-Laurian (1983 : 10*sqq.*) notamment, un genre textuel particulièrement pertinent à considérer lorsque l'on s'intéresse à une situation de communication de type expert-expert est l'article de revue scientifique. C'est le choix que font Halskov (2005) et Dury (2008) par exemple, dans leurs études de la déterminologisation en corpus (cf. Chapitre 2, section 2.2.2).

Cependant, les articles scientifiques sont désormais le plus souvent rédigés en anglais, particulièrement dans des domaines de sciences dites « dures », comme le soulignent également Dury (2008 : 234) et Nicolae et Delavigne (2013 : 220). Le choix de ne sélectionner que des articles de recherche comporterait ainsi un risque de manque de données pour cette partie du corpus. Pour cette raison, nous intégrons aussi des thèses de doctorat, à l'instar par exemple des auteurs du corpus Scientext (Tutin *et al.*, 2009)⁹⁵. Ce choix se justifie également par le fait que, comme l'explique notamment Loffler-Laurian (1983 : 12), c'est précisément la rédaction de la thèse qui confère aux étudiants en doctorat le statut de spécialistes. De ce point de vue, malgré une dimension pédagogique inhérente à ce genre, les thèses relèvent bien d'un degré de spécialisation d'expert à expert et, de fait, « participent aussi des discours scientifiques spécialisés » (*Ibid.*).

Dans le second cas, puisque le concept de langue générale est complexe et difficile à définir de manière opérationnelle pour une étude en corpus⁹⁶, nous faisons ici le choix de qualifier les textes à sélectionner de textes non spécialisés et non de textes de langue générale. Par ailleurs, les corpus de presse généraliste sont le plus souvent considérés comme des alternatives opérationnelles et acceptables dans ce cas⁹⁷. C'est pourquoi, à la suite de la majorité des auteurs qui se sont penchés sur la question de la déterminologisation avec une approche en corpus (Meyer et Mackintosh, 2000a, 2000b ; Halskov, 2005 ; Dury, 2008 ; Condamines et Picton, 2014a ; Renouf, 2017), nous faisons le choix d'intégrer des articles de presse généraliste pour la partie non spécialisée du corpus. Ce choix se justifie également par le rôle joué par les médias dans le processus de

⁹⁵ Corpus Scientext, <https://scientext.hypotheses.org/corpus-des-ecrits-scientifiques-francais> (page consultée le 30 mars 2021).

⁹⁶ Si l'on en croit le grand nombre de critères largement traités dans le cadre de la constitution d'un corpus de référence (par exemple Sinclair, 1991 ; Kennedy, 1998 ; Atkins et Rundell, 2008 ; Siepmann *et al.*, 2017) et les questions fondamentales que pose le concept de représentativité (cf. 4.2.1).

⁹⁷ À noter que l'accès aux articles de presse facilité par les moyens actuels ainsi que le peu de prétraitement à fournir pour rendre ces articles exploitables avec des outils d'exploration de corpus sont des atouts non négligeables de la presse.

déterminologisation (cf. 4.1.1.1). De ce point de vue, la presse représente non pas la langue générale dans son ensemble, mais bien une étape dans le processus d'intégration de termes dans la langue générale, et c'est précisément cette caractéristique qui la rend si pertinente pour notre étude.

Enfin, soulignons que nous privilégions la presse écrite non seulement pour des raisons pratiques de collecte de données, mais également en raison de son caractère pérenne⁹⁸ et de son influence potentielle sur ses lecteurs. Pour reprendre un concept de Guespin (1985), développé par Delavigne (2001 : 264-265), nous prenons en compte le « poids glottopolitique » de la presse, et en particulier des journaux les plus lus et les plus diffusés. À l'instar de Delavigne,

« [n]ous prendrons ici glottopolitique comme un critère de pertinence : à quoi bon rassembler des documents très peu lus, donc n'ayant aucun poids sur les pratiques langagières, qui sont d'un certain point de vue "morts", hors circulation ? » (Delavigne, 2001 : 264-265)

Dans un second temps, afin de sélectionner des textes qui se situent entre ces deux extrémités sur le continuum langue de spécialité-langue générale, nous devons identifier des textes pertinents qui relèvent d'un degré de spécialisation « intermédiaire ». Ce que nous appelons ici « intermédiaire » correspond en réalité aux situations de communication spécialisées qui se caractérisent par différents degrés d'expertise chez les locuteurs impliqués (Bowker et Pearson, 2002 : 27-28). Les autrices distinguent en effet trois degrés de spécialisation dans la communication spécialisée, à savoir d'expert à expert, d'expert à semi-expert et d'expert à non-expert (*Ibid.* : 28). Le premier peut être illustré avec des articles de recherche, comme nous le précisons *supra*, et seuls les deux derniers renvoient à un degré de spécialisation que nous qualifions d'intermédiaire. Bowker et Pearson expliquent que certains genres de textes qui sont produits par des experts sont destinés à des semi-experts, qui sont censés avoir certaines connaissances sur un domaine mais moins poussées que celles d'un expert (par exemple des étudiants assez avancés ou des experts de domaines connexes), et d'autres sont destinés à des non-experts, qui sont censés n'avoir aucune connaissance d'un domaine, ou presque aucune (*Ibid.*).

La différence entre ces deux degrés repose ainsi principalement sur les connaissances que chaque individu peut avoir d'un sujet. D'un point de vue théorique, cette différence est tout à fait claire et pertinente, la distinction entre le niveau de connaissances de locuteurs semi-experts et de locuteurs non experts étant parfaitement concevable. Néanmoins, en pratique, la limite entre le niveau de connaissances requis pour considérer un locuteur comme un semi-expert ou comme un non-expert reste difficile à établir. Cela est d'autant plus vrai si l'on tient compte du fait que le

⁹⁸ Cet argument est également mis en avant par Kennedy (1998 : 63).

lectorat d'un texte peut être extrêmement diversifié. Il semble en effet difficile, voire impossible, d'identifier avec précision le niveau de connaissances de tous les individus susceptibles de lire un article d'une revue de vulgarisation, par exemple, surtout si l'on tient compte de l'accès facilité à de nombreuses publications *via* Internet. Dans ce contexte, nous faisons le choix de considérer l'ensemble des textes produits par des experts pour des non-experts ou des semi-experts comme des textes qui relèvent d'un degré de spécialisation *intermédiaire*, c'est-à-dire qui ne sont ni des textes destinés à des experts ni des textes non spécialisés. Ce choix nous permet d'opérationnaliser la sélection des genres pertinents en nous focalisant sur les différences d'expertise entre les émetteurs des textes et les locuteurs ciblés, plutôt que sur la détermination précise du degré d'expertise de ces derniers.

Plusieurs genres textuels, qui correspondent à différentes situations de communication, peuvent relever de ce degré de spécialisation intermédiaire, par exemple des communiqués de presse, tels que ceux du Cnes, dans l'étude de Condamines et Picton (2014a : 171), ou des articles de revues de vulgarisation et des messages de forums, comme dans l'étude de Halskov (2005 : 54*sqq.*) (cf. Chapitre 2, 2.2.2.3). Afin de déterminer les genres de textes à inclure dans notre corpus, nous nous servons d'une caractéristique supplémentaire et prenons en compte les genres qui entrent dans ce que l'on appelle des « discours de transmission des connaissances », selon la définition de Beacco et Moirand (1995) présentée au Chapitre 1 (1.2.2). En effet, nous avons montré que les discours de transmission des connaissances contribuent à la circulation des termes en dehors de la sphère spécialisée et, par extension, à leur intégration dans la langue générale. Aussi, nous cherchons à identifier des textes pour lesquels nous pouvons faire l'hypothèse qu'ils contribuent à la déterminologisation, du fait que le genre dont ils relèvent peut être qualifié de discours de transmission des connaissances.

Les textes à inclure dans la partie intermédiaire du corpus doivent donc appartenir à des genres qui se caractérisent par un degré de spécialisation intermédiaire et par le fait qu'ils participent à transmettre des connaissances. De manière complémentaire à cette réflexion sur les caractéristiques générales des genres à considérer, nous reprenons à notre compte les hypothèses des chercheurs ayant travaillé sur la déterminologisation avec une approche en corpus et ayant interrogé cette dimension continue de la déterminologisation. Ainsi, nous sélectionnons :

- des communiqués de presse,
- des rapports de laboratoires orientés grand public,
- des articles de revues de vulgarisation et des sites Web de vulgarisation.

Les communiqués de presse qui émanent d'institutions ou d'organismes actifs dans la recherche en physique des particules représentent un genre particulièrement intéressant, car ils sont « destinés par définition aux professionnels de la presse » (Nicolae et Delavigne, 2013 : 219) et qu'ils entrent dans le concept des discours de transmission des connaissances. Ainsi, à la suite de Condamines et Picton (2014a : 171-172), nous faisons l'hypothèse que les termes présents dans les communiqués de presse sont susceptibles d'être repris par les journalistes qui exploitent ces communiqués. Les communiqués permettent donc de représenter une « étape » dans le processus d'intégration des termes dans la langue générale.

En ce qui concerne le choix des rapports, nous sélectionnons un type de rapport d'activités bien précis : le rapport annuel (ou bisannuel) disponible pour le grand public. Ce type de rapport se caractérise par son libre accès sur les sites Web des laboratoires et se distingue des rapports internes par le fait qu'il vise plutôt à informer le public de manière générale sur les activités de recherche menées au sein du laboratoire. La communication avec le grand public constitue en effet l'un des enjeux actuels de la recherche. Il existe par exemple des formations à l'intention des chercheurs pour les aider à communiquer efficacement avec les médias⁹⁹. Mais l'importance de ces enjeux de communication avec le grand public est surtout illustrée par la volonté et la nécessité pour les organismes de recherche d'informer le grand public des avancées scientifiques et de rendre des comptes à la société, puisque la recherche est le plus souvent financée par des fonds publics¹⁰⁰. Dans cette perspective, nous faisons l'hypothèse que des rapports orientés grand public peuvent participer à la transmission des connaissances spécialisées vers le grand public, directement ou par le biais des médias, et donc représenter une autre « étape » dans le processus de déterminologisation.

Enfin, nous sélectionnons deux genres relevant de la vulgarisation scientifique : des articles de revue et des sites Web. Nous l'expliquons au Chapitre 1 (1.2.1), la vulgarisation poursuit le but de transmettre des connaissances spécialisées auprès d'un public plus large, non expert. De ce point de vue, les textes de vulgarisation contribuent à la circulation des termes en dehors de la sphère spécialisée et, plus largement, au processus de déterminologisation. Rappelons que c'est également pour cette caractéristique que Halskov (2005) inclut des textes de vulgarisation dans son étude de

⁹⁹ Par exemple, le CNRS organise régulièrement des formations de ce type (<https://cnrs-hebdo.dr14.cnrs.fr/Actualites/23736/Suite.aspx>, <https://www.insis.cnrs.fr/fr/cnrsinfo/formation-communiquer-la-science-et-travailler-avec-les-medias>, pages consultées le 30 septembre 2019).

¹⁰⁰ Que ce soit par le biais des universités, qui sont le plus souvent publiques en France et en Suisse, et donc financées par l'État, ou par le biais d'organismes de promotion de la recherche, financés entièrement ou partiellement par l'État, par exemple le FNS en Suisse (Fonds national suisse de la recherche scientifique, <http://www.snf.ch/fr/Pages/default.aspx>, page consultée le 6 avril 2021) ou le CNRS en France (Centre national de la recherche scientifique, <https://www.cnrs.fr/fr/page-daccueil>, page consultée le 6 avril 2021).

la déterminologisation. En outre, le choix d'inclure deux genres relevant de la vulgarisation scientifique est motivé en particulier par leur complémentarité. En effet, les articles de revue ont tendance à couvrir des sujets d'actualités, des événements ou des découvertes, par exemple, tandis que les sites Web ont plutôt tendance à expliquer le domaine dans son ensemble et dans toute sa complexité. La sélection de ces deux genres représente donc mieux la diversité des moyens de la vulgarisation.

Le tableau ci-dessous résume nos différents choix de genres textuels, en fonction des degrés de spécialisation dont ils relèvent.

Degré de spécialisation	Genre textuel
Élevé	Article de revue spécialisée Thèse de doctorat
Intermédiaire	Communiqué de presse Rapport de laboratoire « grand public » Article de revue et site Web de vulgarisation
Non spécialisé	Article de presse généraliste

Tableau 4.1 : Composition du corpus : genres et degrés de spécialisation

4.2.2.1.2. Approcher le continuum par l'organisation des données en plusieurs sous-corpus

Le choix de textes qui relèvent de genres et de degrés de spécialisation différents constitue en réalité une première phase de la représentation en corpus du continuum langue de spécialité-langue générale. Afin de rendre possible l'observation du processus de déterminologisation dans ce corpus, les données doivent être organisées pour permettre l'interrogation de la continuité de ce processus dans le corpus. Il est donc nécessaire à ce stade d'organiser les données de manière à se rapprocher de ce continuum.

Parallèlement, nous cherchons à mettre en œuvre une démarche d'analyse comparée de corpus, pour faire émerger des fonctionnements différents de la comparaison de sous-corpus, que nous interprétons ensuite en relation avec le processus de déterminologisation (cf. Chapitre 3, 3.3.1). L'organisation des données dans le corpus doit donc répondre à cette exigence de comparabilité des données et plusieurs sous-corpus doivent être créés dans ce but.

Représenter un continuum en corpus en vue d'interroger la continuité d'un processus reste toutefois un défi de taille, étant donné le caractère nécessairement figé des corpus. Bien que ce constat puisse finalement rendre les corpus peu adéquats pour les besoins de notre travail, il s'agit d'un problème récurrent dans toute étude sur corpus et

« [l]a difficulté vient donc de concilier à la fois le dynamisme de la langue en lien avec des situations extralinguistiques toujours mouvantes et la nécessaire stabilisation que requiert une description "scientifique". » (Condamines, 2003 : 72)

Dans notre cas, cela signifie que la représentation du continuum langue de spécialité-langue générale est nécessairement biaisée d'une façon ou d'une autre, mais, comme le souligne Sinclair (2008 : 30), « since there is no known alternative [...], we use corpora in full awareness of their possible shortcomings ». En outre, du fait que, « même constitué avec le plus grand soin, un corpus ne se suffit pas à lui-même pour son interprétation » (Condamines, 2003 : 72), les lacunes inévitables du corpus sont prises en compte dans l'interprétation des résultats, comme le recommandent également Bowker et Pearson (2002 : 58).

Dans ce contexte, le défi principal de ce corpus consiste alors à trouver un équilibre dans le nombre de sous-corpus à créer et à assurer leur pertinence pour notre objectif. En effet, le nombre de sous-corpus doit permettre d'approcher le continuum entre la langue de spécialité de la physique des particules et la langue générale (par le biais de textes non spécialisés dans ce cas), tout en maintenant les comparaisons entre sous-corpus interprétables. Autrement dit, la multiplication des points de comparaison entre spécialisé et non spécialisé est nécessaire pour tendre vers la continuité, mais un nombre raisonnable de ces points de comparaison est tout aussi nécessaire pour assurer la comparabilité des données. En effet, plus ce nombre est élevé, plus les comparaisons risquent d'être difficiles à interpréter, d'un point de vue tant technique que cognitif. À ce stade, nous jugeons que trois sous-corpus au moins sont nécessaires, afin de représenter les deux extrémités du continuum telles que l'on peut les observer en corpus et ce qui se passe entre ces deux extrémités.

Parallèlement, dans notre travail nous poursuivons le but d'interroger également le rôle des différents genres qui relèvent d'un degré de spécialisation intermédiaire dans le processus d'intégration des termes dans la langue générale. Cet aspect est encore très peu abordé dans les travaux sur la déterminologisation, mais il est pourtant crucial pour parvenir à une meilleure compréhension de ce phénomène. Cet objectif impose donc de créer la possibilité, dans l'organisation des données en sous-corpus, de considérer individuellement chacun des genres pertinents qui se situent à l'intermédiaire entre les textes très spécialisés et les textes non spécialisés. C'est pourquoi nous faisons le choix de constituer trois sous-corpus intermédiaires, qui correspondent aux différents genres identifiés. Seuls les textes de vulgarisation scientifique (articles de revue et sites Web dédiés au domaine) sont rassemblés en un seul sous-corpus, en raison de leur

complémentarité et de leur objectif commun de transmission de connaissances spécialisées auprès d'un public non expert.

Ainsi, afin de représenter le continuum langue de spécialité-langue générale et de permettre d'appréhender le processus de déterminologisation dans le domaine de la physique des particules dans ce continuum, nous organisons les données en cinq sous-corpus. La composition de chacun des sous-corpus est résumée dans le tableau 4.2.

Sous-corpus	Degré de spécialisation	Genre textuel
<i>Spécialisé</i> (abrégé <i>SC-S</i>)	Élevé	Article de revue spécialisée Thèse de doctorat
<i>Communiqués</i> (abrégé <i>SC-C</i>)	Intermédiaire	Communiqué de presse
<i>Rapports</i> (abrégé <i>SC-R</i>)	Intermédiaire	Rapport de laboratoire « grand public »
<i>Vulgarisation</i> (abrégé <i>SC-V</i>)	Intermédiaire	Article de revue de vulgarisation Site Web de vulgarisation
<i>Presse</i> (abrégé <i>SC-P</i>)	Non spécialisé	Article de presse généraliste

Tableau 4.2 : Composition des sous-corpus en termes de genres et de degrés de spécialisation

Les différents aspects discutés dans cette section visent à clarifier certains des critères que le corpus doit satisfaire pour assurer sa pertinence pour la déterminologisation. Néanmoins, d'autres critères sont nécessaires pour maximiser les chances d'observer, dans ce corpus, des fonctionnements des termes dont nous pouvons faire l'hypothèse qu'ils relèvent de la déterminologisation. L'un de ces critères passe par la définition d'une fenêtre temporelle pertinente.

4.2.2.2. Définir une fenêtre temporelle pertinente

Dans le but de garantir la pertinence du corpus pour notre objectif, nous devons nous assurer que des termes de ce domaine sont effectivement susceptibles de s'intégrer dans la langue générale, c'est-à-dire d'être diffusés par les intermédiaires que nous avons identifiés et par les médias, donc d'être attestés dans le sous-corpus de presse généraliste. Cet aspect est pris en compte de deux manières, la seconde faisant l'objet de la section suivante. Pour ce qui est de la première, elle passe par la définition d'une fenêtre temporelle pertinente, au cours de laquelle des événements marquants se sont produits dans le domaine et sont potentiellement couverts par les médias.

Ainsi, bien que la physique des particules soit un domaine de recherche scientifique susceptible d'être présent dans les médias, il n'est pas sûr que ce domaine soit traité de manière

régulière dans le temps. De ce point de vue, il est nécessaire de déterminer une période pertinente à observer, qui conditionne la sélection de chaque texte, en fonction des paramètres définis *supra*. Cette période doit se caractériser par un traitement possiblement abondant de la physique des particules par la presse. Par ailleurs, nous avons vu que les médias, et donc la presse, ont tendance à mettre en avant ce qui a trait au spectaculaire et à la nouveauté. Dans ce contexte, nous pouvons supposer que des découvertes importantes ou des avancées technologiques particulières sont largement couvertes dans les médias. C'est pourquoi nous pensons qu'il est nécessaire d'identifier ce type d'évènements dans ce domaine et de constituer un corpus dans une fenêtre temporelle comprenant les périodes précédant et suivant ces évènements.

En effet, nous faisons à ce stade l'hypothèse que des évènements importants d'un domaine, s'ils sont largement couverts par les médias, peuvent contribuer à l'intégration dans la langue générale des termes employés pour décrire ces évènements et peuvent influencer la manière dont ces termes sont utilisés dans les textes. Ces influences seraient alors reflétées dans des changements linguistiques observables en corpus. Tout l'enjeu consiste donc à identifier ces évènements.

Une experte a été interrogée dans ce but et deux évènements ont été retenus. Ces évènements sont plutôt récents, et, d'après l'experte, sont susceptibles d'avoir été traités par les médias de manière extensive, vu leur importance dans l'évolution de la physique des particules. Ces deux évènements sont la mise en marche du LHC en 2008 et la découverte du boson de Higgs en 2012. Tous les deux se sont passés au CERN. La définition d'une fenêtre temporelle pertinente doit donc permettre d'inclure ces deux évènements. Aussi, et afin de constituer des échantillons suffisamment grands, nous faisons le choix de considérer des textes publiés à partir de 2003 jusqu'à 2016, 2016 étant l'année où le corpus a été compilé. Précisons également que la période antérieure à 2003 n'est pas prise en compte, car trop peu de données sont disponibles pour cette période, ce qui présenterait le risque de déséquilibrer le corpus. Enfin, les données collectées sont équilibrées par période plutôt que par année. En effet, bien que nous adoptions une perspective synchronique dans notre thèse, le corpus s'articule autour des deux évènements identifiés par les expertes et peut par conséquent être divisé en trois périodes avant et/ou après chacun de ces évènements (c'est-à-dire de 2003 à 2007, de 2008 à 2011 et de 2012 à 2016).

L'intérêt d'équilibrer le nombre de données par période plutôt que par année s'explique principalement par la nature parfois très différente des documents ciblés (si nous comparons par exemple la longueur d'un communiqué de presse ou d'un article de presse généraliste avec la longueur d'un rapport) et par le manque de symétrie dans la périodicité des publications (les communiqués de presse peuvent paraître chaque jour, tandis que les rapports de certains laboratoires ne sont parfois publiés que tous les trois ans), qui rend l'équilibre quantitatif par année

impossible à atteindre. Aussi, afin que chaque sous-corpus soit équilibré sur la période considérée de 2003 à 2016, nous cherchons à obtenir des données de taille comparable non pas pour chaque année prise en compte, mais plutôt pour chacune des trois périodes. Ce choix a également pour but de réduire les possibilités qu'un des événements soit surreprésenté dans les données d'un ou plusieurs sous-corpus, ce qui nuirait à la comparabilité des cinq sous-corpus (voir en particulier le tableau A.2 en annexe).

Afin de tendre vers cet équilibre et de sélectionner des textes qui soient pertinents pour la physique des particules et pour les deux événements dont nous tenons compte, la pertinence de chaque texte inclus dans le corpus est évaluée à l'aide de différents critères, que nous développons dans la section suivante.

4.2.2.3. Assurer la pertinence du corpus par rapport au domaine grâce à une procédure de sélection des textes systématique et objective

Dans cette section, nous détaillons la dernière phase de notre réflexion sur la constitution de corpus, qui concerne l'adéquation des textes avec le domaine étudié. Nous abordons cette phase d'un point de vue plus opérationnel, en nous intéressant aux critères précis d'inclusion ou d'exclusion d'un texte satisfaisant les différentes conditions énoncées *supra*. En effet, compiler un corpus pour un besoin précis implique également de sélectionner chaque texte individuellement et, afin que le corpus soit adéquat, les textes le composant doivent être pertinents pour le domaine.

Dans ce but, nous développons une procédure de sélection des textes qui répond à des critères précis et objectifs, afin d'assurer notamment la transparence et la reproductibilité de la démarche et de limiter certains biais habituellement associés à la constitution de corpus (sélection arbitraire, non-reproductibilité, etc.). Enfin, la question de la pertinence par rapport au domaine ne pouvant pas être abordée de la même manière pour les sous-corpus relevant d'un degré de spécialisation élevé ou intermédiaire et pour le sous-corpus de presse, nous traitons d'abord des quatre premiers puis du dernier.

4.2.2.3.1. Sélectionner les textes constitutifs des sous-corpus *Spécialisé, Communiqués, Rapports et Vulgarisation*

La sélection des textes constitutifs d'un corpus spécialisé repose le plus souvent sur le contenu des textes : s'il est jugé pertinent pour le domaine par l'analyste, alors les textes sont retenus. Pour Pearson (1998 : 54), l'évaluation de la pertinence du contenu d'un texte passe par l'examen de son titre, voire de la table des matières dans le cas d'un livre par exemple, ou par

l'identification des mots-clés apparaissant fréquemment dans le texte. Une approche basée sur la présence de termes considérés comme clés pour le domaine semble alors tout à fait pertinente.

Dans ce contexte, il est nécessaire de définir un nombre suffisant de termes-clés qui se réfèrent de manière aussi univoque que possible au domaine et qui soient en lien avec les deux événements servant à délimiter la fenêtre temporelle (la mise en marche du LHC et la découverte du boson de Higgs). Une experte nous a assistée dans cette tâche et des termes tels que *Modèle Standard*, *boson de Higgs*, *ATLAS*, *LHC* ou *particule élémentaire* ont été fixés. Les textes retenus sont donc ceux qui contiennent au moins l'un de ces termes dans les titres, dans les tables des matières (dans le cas des thèses ou des rapports) et même parfois dans le corps du texte si nécessaire¹⁰¹.

Parallèlement à la définition des termes-clés, l'experte joue également un rôle crucial dans l'identification des sources des textes à considérer en priorité. C'est grâce à la clarification, de la part de l'experte, de l'importance du CERN et du LHC dans la recherche du boson de Higgs, dans le cadre du Modèle Standard, que nous pouvons assurer une certaine cohérence dans la sélection des sources des textes, et donc également dans les sujets contenus dans le corpus. En effet, étant donné que le LHC se trouve au CERN et que le boson de Higgs y a été découvert, il semble pertinent de donner la priorité à des sources en lien avec le CERN. En outre, le CERN étant situé à la frontière franco-suisse, nous privilégions des sources suisses et françaises. Enfin, pour pallier le manque de publications suisses en français¹⁰² et pour garantir une certaine diversité des textes, nous considérons également des sources autres que le CERN. Les textes ainsi sélectionnés proviennent donc tous :

- d'universités françaises donnant accès aux produits de leur recherche en français, en particulier aux thèses, par le biais d'archives ouvertes,
- de la seule revue de physique spécialisée française publiant des articles en français¹⁰³,
- des sites Web des laboratoires de recherche¹⁰⁴ qui mènent des activités de recherche en physique des particules, en Suisse ou en France,

¹⁰¹ Ce dernier point est nécessaire pour les communiqués de presse. En particulier, dans les cas où les titres ne contiennent aucun des termes prédéfinis, nous jugeons qu'il est préférable de nous assurer de la non-pertinence de ces communiqués en tenant compte du corps du texte. S'agissant de textes plutôt courts, il est aisé de vérifier leur adéquation avec le domaine par une rapide lecture « en diagonale » de leur contenu.

¹⁰² La majorité des publications qui relèvent d'un degré de spécialisation élevé et qui émanent des universités suisses ainsi que du CERN sont en anglais.

¹⁰³ *Reflets de la physique* (<https://www.refletsdelaphysique.fr/>, page consultée le 7 octobre 2019).

¹⁰⁴ Dans cette dénomination, nous incluons également les organes de recherches portant le nom d'*institut* ou d'*organisation*, tels que l'IN2P3 (Institut national de physique nucléaire et de physique des particules) et le CERN (Organisation européenne pour la recherche nucléaire).

- des revues de vulgarisation françaises,
- des quotidiens suisses ou français.

À partir de ces sources, l'identification des textes pertinents se fonde essentiellement sur la présence d'au moins un des termes-clés, puis sur une validation par l'experte. En outre, afin de décider de la pertinence de ces textes, nous vérifions également leur accessibilité par les locuteurs (experts, semi-experts et non-experts). En effet, les textes ne doivent pas être destinés à une communauté restreinte de locuteurs et nous jugeons qu'il est préférable de privilégier des textes (et donc des sources) qui sont diffusés auprès du grand public, ou du moins auxquels le grand public peut accéder librement. Ce dernier critère nous amène ainsi à exclure par exemple des rapports de laboratoires internes, des comptes rendus de réunions, des spécifications, etc. En somme, nous excluons tout document possédant un caractère confidentiel et qui n'est donc pas voué à être diffusé en dehors de la communauté des utilisateurs auxquels il est dédié.

Par ailleurs, l'exploitation des termes-clés rend possible non seulement l'inclusion des textes, mais également l'exclusion de certains textes, sur une base objective. Par exemple, les communiqués de presse informant sur des sujets de nature plutôt administrative ou politique, comme l'intégration d'un nouvel État membre, sont écartés. De la même manière, vu le degré de spécialisation élevé de certains textes, une ultime étape de vérification auprès de l'experte permet de lever les derniers doutes. L'exploitation systématique des mêmes termes-clés ainsi qu'une étroite collaboration avec une experte rend donc possible la sélection des textes les plus pertinents uniquement, pour constituer les quatre premiers sous-corpus (*Spécialisé, Communiqués, Rapports et Vulgarisation*)

En revanche, vu la nature du phénomène que nous étudions ainsi que les caractéristiques de la presse, une procédure de sélection plus élaborée doit être mise en œuvre. Dans la section suivante, nous détaillons donc le protocole que nous avons mis au point pour constituer un sous-corpus de presse qui soit aussi adéquat que possible pour analyser le phénomène de déterminologisation dans le domaine de la physique des particules, tout en évitant les biais introduits par une sélection d'articles basée sur un nombre restreint de termes-clés.

4.2.2.3.2. Adapter la procédure au sous-corpus *Presse*

Ce dernier sous-corpus doit permettre l'analyse de termes de physique des particules dans une situation de déterminologisation. Il est donc capital que de tels termes soient attestés dans les textes constitutifs de ce sous-corpus. Néanmoins, la sélection de ces textes à l'aide des mêmes termes-clés que pour les autres sous-corpus n'est pas idéale, car ces termes-clés sont trop limités et

risquent d'orienter la sélection vers des articles portant sur des sujets tout à fait similaires. En effet, nous savons que la déterminologisation peut avoir des conséquences très diverses sur le fonctionnement des termes, les termes peuvent par exemple être utilisés de manière métaphorique ou dans des jeux de mots (cf. Chapitre 2, 2.2.2.1), et ces conséquences doivent apparaître dans le sous-corpus si nous cherchons à les analyser. De ce point de vue, si nous sélectionnons les articles de presse uniquement sur la base de leur thématique, avec les mêmes termes-clés que les autres sous-corpus, alors il est possible que certains des usages des termes résultant du processus de déterminologisation ne soient pas attestés du tout.

Inversement, une sélection de textes totalement aléatoire ne semble pas non plus satisfaisante. Afin d'analyser des termes de physique des particules, il faut s'assurer que de tels termes soient attestés dans les textes – et qu'ils apparaissent avec une fréquence suffisante. S'il semble évident qu'une seule occurrence ne permet pas de caractériser le fonctionnement d'une unité, il semble également difficile de définir une fréquence idéale :

« [t]he issue is how many tokens of a linguistic item are necessary for descriptive adequacy. In a typical text (or a corpus of one million words), 40 to 50 % of the word types occur only once. » (Kennedy, 1998 : 67)

Dans cette perspective, il semble préférable de s'assurer que les termes susceptibles de faire l'objet d'analyses soient suffisamment fréquents dans les textes constitutifs du sous-corpus *Presse*. Il est donc nécessaire de mettre en œuvre une procédure de sélection des articles de presse qui garantit l'attestation de termes de physique des particules dans le sous-corpus, tout en évitant autant que possible d'introduire des biais qui pourraient nuire à l'analyse. Ainsi, comme le préconise Hunston (2008 : 156), par exemple, à propos de phénomènes très spécifiques ou d'unités plutôt rares, la nécessité de collecter suffisamment de données à analyser (c'est-à-dire ici suffisamment d'occurrences de plusieurs termes de physique des particules) justifie l'orientation de la sélection des textes sur les phénomènes précis qui sont ciblés.

Dans ce but, nous développons une méthode hybride, qui tient compte de toutes ces contraintes et qui permet de constituer un sous-corpus de taille raisonnable qui soit adéquat pour le domaine ainsi que pour la déterminologisation. Cette méthode consiste à sélectionner les articles de presse à partir d'un nombre élevé de termes (191 précisément), attestés dans les quatre autres sous-corpus (nous détaillons l'identification de ces termes en 4.3.2). Concrètement, ces termes sont utilisés comme mots-clés sur la plateforme NexisUni®¹⁰⁵, pour récupérer les articles qui contiennent ces mots-clés. À la différence d'une recherche à partir de quelques mots-clés

¹⁰⁵ Anciennement LexisNexis, à laquelle nous avons accès *via* un abonnement à l'Université de Genève.

seulement, l'utilisation d'une plus longue liste permet de récupérer des articles au contenu plus diversifié. Nous évitons ainsi un des biais de sélection évoqués *supra*. En outre, afin de maximiser les chances d'observer des fonctionnements relevant de la déterminologisation, les articles sont collectés selon une stratégie précise, qui est rendue possible en particulier par le nombre élevé de mots-clés utilisés. Cette stratégie repose sur le principe d'une sélection équilibrée entre les articles les plus pertinents pour la plateforme, c'est-à-dire ceux qui comprennent un grand nombre de mots-clés, et les moins pertinents, c'est-à-dire les articles qui contiennent le moins de mots-clés, parfois même un seul.

Selon l'hypothèse sous-tendant cette stratégie, dans les articles contenant le moins de mots-clés, par exemple, nous pouvons nous attendre à observer des usages métaphoriques des termes de physique des particules ou des jeux de mots ou encore d'autres fonctionnements singuliers liés à la déterminologisation et qui restent à découvrir. *A contrario*, l'intégration d'articles dans lesquels de nombreux termes apparaissent permet dans un premier temps d'analyser l'usage de ces termes dans des textes non spécialisés et, dans un second temps, d'effectuer des comparaisons avec les sous-corpus intermédiaires et le sous-corpus spécialisé. À terme, nous pensons que la diversité que nous visons avec cette stratégie de sélection permettra d'apporter un regard complémentaire sur le processus de déterminologisation.

Ce dernier sous-corpus permet d'aboutir à un corpus totalisant un peu plus de quatre millions d'occurrences, dont la répartition dans les cinq sous-corpus est illustrée dans le tableau ci-dessous. Le détail des différentes sources incluses et de la répartition des occurrences selon les sources et selon les périodes peut être consulté dans l'annexe A.

Sous-corpus	Occurrences
<i>Spécialisé</i> (SC-S)	994 875
<i>Communiqués</i> (SC-C)	210 320
<i>Rapports</i> (SC-R)	1 141 873
<i>Vulgarisation</i> (SC-V)	620 045
<i>Presse</i> (SC-P)	1 098 708
Total	4 065 821

Tableau 4.3 : Taille du corpus en nombre d'occurrences

Dans cette section, nous avons posé les principes pour la constitution d'un corpus servant à l'exploration d'un processus continu tel que la déterminologisation. L'originalité de notre proposition réside dans l'organisation des données en un corpus comparable complexe, qui mêle différents genres textuels et différents degrés de spécialisation, afin d'approcher le continuum

existant entre la langue de spécialité de la physique des particules et la langue générale, et qui permet d'interroger cette continuité. En outre, dans un souci d'objectivité et de reproductibilité de notre démarche, nous avons veillé à définir des critères aussi objectifs que possible et à présenter nos réflexions de manière aussi transparente que possible.

Enfin, rappelons que notre démarche repose essentiellement sur l'analyse des termes. Afin de constituer un échantillon pertinent, nous développons une procédure de sélection des termes qui se base sur des critères précis.

4.3. Méthode endogène de sélection des termes

Dans notre cadre d'analyse, les termes servent non seulement de pivots autour desquels s'articulent nos observations (cf. Chapitre 3, 3.3.1), mais également de termes-clés pour la sélection des articles constitutifs du sous-corpus *Presse* (cf. 4.2.2.3.2). Afin de repérer ces termes et de constituer ces deux échantillons, nous mettons en œuvre une méthode endogène, qui se déroule en deux temps. Dans un premier temps, nous procédons au repérage semi-automatique des candidats-termes (CT) dans le corpus et, dans un second temps, nous filtrons les CT dans le but de conserver uniquement les termes les plus pertinents pour notre travail et pour le domaine.

4.3.1. Repérer des candidats-termes de manière semi-automatique

La première étape consiste à repérer les CT, à l'aide d'un logiciel dédié, qui est TermoStat dans notre cas. La figure 4.2 illustre cette première étape.

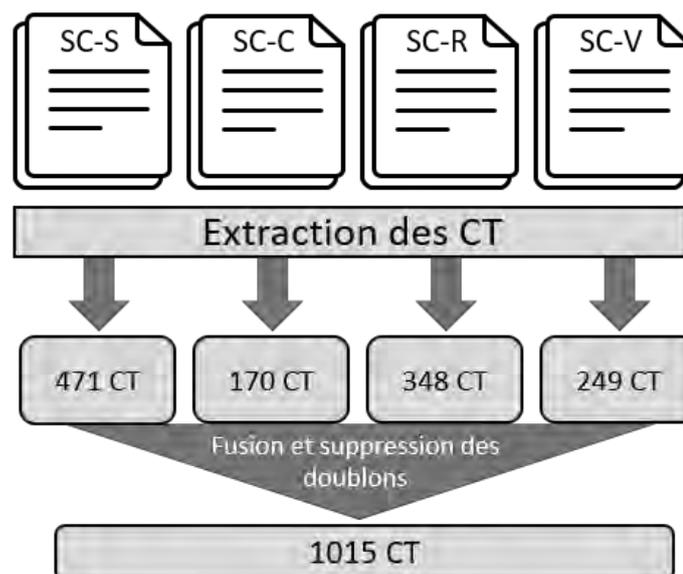


Figure 4.2 : Première étape de sélection des termes – extraction

Les sous-corpus SC-S, SC-C, SC-R et SC-V sont soumis à TermoStat séparément, dans le but d'éviter une sous-représentation des sous-corpus les plus petits. En effet, les SC-C et SC-V comportent moins d'occurrences que les SC-S et SC-R. Or, comme TermoStat fonctionne principalement à partir de la fréquence des unités dans les corpus d'analyse et de référence, il semble préférable de générer une liste de CT par sous-corpus puis de fusionner les quatre listes ainsi obtenues. De cette manière, les CT les plus spécifiques aux SC-C et SC-V peuvent être repérés. De ces quatre listes, nous retenons les CT avec un score de spécificité plus grand ou égal à 40. Ainsi, comme nous le voyons dans la figure 4.2, 471 CT sont repérés dans le SC-S, 170 dans le SC-C, 348 dans le SC-R et 249 dans le SC-V. Certains des CT sont repérés dans plusieurs sous-corpus ; c'est pourquoi, une fois les doublons supprimés, la liste se compose de 1015 CT.

4.3.2. Filtrer les résultats et constituer un échantillon pertinent

Les résultats bruts obtenus avec TermoStat doivent être nettoyés. En effet, même si certains CT sont repérés car leur fréquence relative dans le sous-corpus analysé est effectivement significativement plus élevée que dans le corpus de référence de TermoStat, tous ne sont pas forcément pertinents d'un point de vue terminologique. Dans ce but, nous établissons quatre filtres, basés sur plusieurs aspects relatifs à notre objet d'étude. Ces filtres sont développés pour répondre à des objectifs différents, que nous détaillons ci-dessous.

Ces quatre filtres sont illustrés dans la figure 4.3, qui vient compléter la figure 4.2 pour représenter l'ensemble de la procédure.

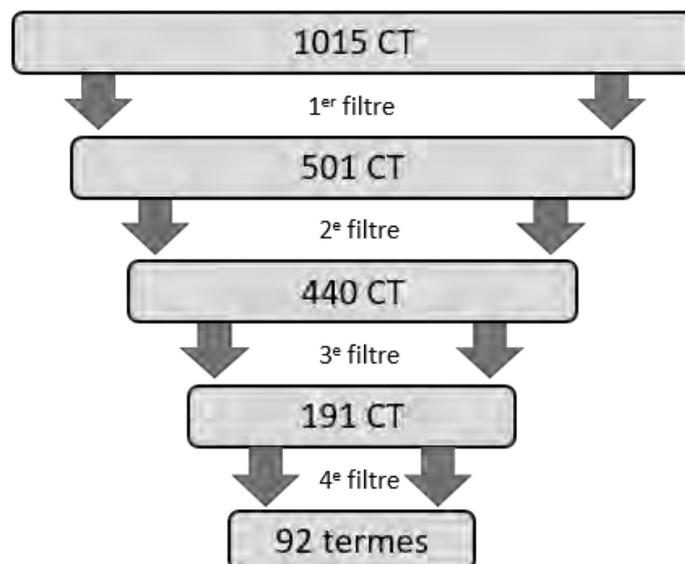


Figure 4.3 : Seconde étape de sélection des termes – filtres

Le premier filtre se base sur notre connaissance du contenu du corpus et du fonctionnement de la langue et vise à identifier les CT qui résultent d'erreurs d'étiquetage et de découpage. Différents types de CT sont alors supprimés :

- les CT qui ne sont pas des termes. Ceux-ci regroupent les prénoms et les noms de pays (par exemple, *perrine*, *rolf*, *république slovaque*, *république tchèque*), les abréviations qui apparaissent le plus souvent dans certaines références bibliographiques (par exemple, *phys*, *nucl*, *arxiv*) ou encore des éléments apparaissant dans des adresses électroniques ou des URL (par exemple, *grey@cern*, *lbc-France*, *http*). Du fait que ces éléments sont récurrents dans certains textes (par exemple, dans les communiqués, *rolf* fait référence au nom d'un ancien directeur du CERN, Rolf Heuer), leur fréquence relative est anormalement élevée par rapport à leur fréquence relative dans le corpus de référence de TermoStat, ce qui explique leur repérage par le système ;
- les CT qui présentent des erreurs de découpage, comme *détecteur de pied*, *période de prise* ou *production de paire*. Ces CT sont conformes aux matrices de TermoStat (ici nom + préposition + nom), mais une vérification en contexte montre qu'ils n'apparaissent jamais tels quels. Par exemple, *détecteur de pied* apparaît toujours dans le syntagme *détecteur de pied de gerbe*, qui est en réalité le terme attendu ;
- les CT qui sont des mots ou termes anglais, tels que *calorimeter*, *particle*, *the*, *of*, *polarization*. Encore une fois, l'observation des contextes d'apparition de ces CT permet de vérifier qu'ils apparaissent systématiquement dans des contextes en anglais (par exemple des abstracts ou des références bibliographiques). Il ne s'agit donc pas d'emprunts et ne sont donc pas pertinents pour notre analyse.

Vu la diversité des CT supprimés par ce premier filtre, soulignons que cette étape est effectuée manuellement. Notre jugement intervient donc nécessairement et une part de subjectivité peut subsister, même si nous cherchons à la limiter en vérifiant systématiquement les occurrences des CT en contexte. Par ailleurs, le nombre relativement restreint de CT obtenu après la fusion des différentes listes reste tout à fait gérable dans une optique non automatisée. Par exemple, dans le cas des abréviations présentes dans les références bibliographiques de certains textes, seule la vérification en contexte permet d'assurer que ces abréviations ne sont pas utilisées dans d'autres parties des textes et donc de confirmer leur suppression. De la même manière, certains CT considérés *a priori* comme des termes anglais sont conservés du fait de leur présence dans des contextes en français (par exemple les termes *trigger* ou *tight*). À l'issue de cette étape, 514 CT sont supprimés, soit un peu plus de la moitié.

Un second filtre est établi en vue de supprimer les CT qui ne renvoient pas à des concepts spécifiques de la physique des particules. En effet, certains des CT repérés sont en réalité caractéristiques du discours scientifique de manière générale et non de la terminologie de la physique des particules en particulier. Ils sont alors susceptibles d'apparaître dans des textes de domaines variés et sont donc peu pertinents pour notre travail.

Ces CT, puisqu'ils ne sont pas spécifiques à un domaine en particulier, renvoient en réalité à un « lexique non terminologique » qui est « propre aux écrits scientifiques » (Tutin, 2007a : 5) et que l'on appelle *lexique scientifique transdisciplinaire* (abrégé *LST*) (Drouin, 2007 ; Tutin, 2007a ; Tutin *et al.*, 2009 ; Drouin, 2010 ; Tutin & Grossman, 2013). Ce concept d'un vocabulaire qui serait commun à tous les domaines de recherche scientifique trouve son origine dans des travaux tels que ceux de Phal (1971), sur le Vocabulaire général d'orientation scientifique, abrégé *VGOS*. La différence principale entre le VGOS et les travaux plus récents sur le LST réside dans le fait que le premier prend en compte essentiellement les sciences dites « dures », excluant ainsi tous les domaines des sciences humaines. En revanche, comme l'affirme Drouin (2007 : 45), « [o]n considère que le LST transcende les domaines de spécialité et présente un noyau lexical commun significatif entre les disciplines ». Selon Tutin, l'hypothèse sous-jacente à ce lexique est double :

« [d]'une part, [...] les écrits scientifiques renvoient à des pratiques intellectuelles qui, malgré les différences inhérentes aux disciplines et aux familles de disciplines, présentent des objectifs et des procédures communs [...] qui seront nécessairement réalisés par des éléments lexicaux dans les écrits [et d']autre part, [...] les écrits scientifiques renvoient à des genres codifiés [...], qui, par-delà les disciplines, poursuivent peu ou prou les mêmes buts, et donc recourraient à des rhétoriques comparables. » (Tutin, 2007a : 5)

Le LST est donc un lexique caractéristique des genres textuels propres au discours scientifique, qui ne sert pas à exprimer les concepts propres d'un domaine, mais qui renvoie « au discours sur les objets et les procédures scientifiques » (*Ibid.* : 6) de manière générale et est commun à différents domaines de la recherche scientifique. Par conséquent, il ne renvoie pas à la terminologie d'un domaine et n'est pas pertinent pour notre travail.

Afin d'identifier les CT qui correspondent au LST nous nous servons de la liste établie dans le cadre du projet Scientext¹⁰⁶. Au total, 61 CT sont supprimés (par exemple *analyse*, *mesurer*, *interaction* ou encore *signal*).

Un troisième filtre est établi, qui vise à cibler en particulier l'échantillon de termes à conserver pour la constitution du sous-corpus *Presse*, parmi les CT extraits par TermoStat. Comme

¹⁰⁶ Projet Scientext, <https://scientext.hypotheses.org/> (page consultée le 19 septembre 2019).

nous l'expliquons en 4.2.2.3.2, ces termes-clés doivent renvoyer au domaine de la physique des particules de manière aussi univoque que possible, afin de garantir autant que possible que les articles de presse récupérés sur la base de ces termes soient pertinents pour notre étude de la déterminologisation dans ce domaine et permettent effectivement d'observer des fonctionnements particuliers des termes pour lesquels nous pouvons faire l'hypothèse qu'ils sont dus au processus de déterminologisation dans ce domaine. Dans ce but, nous cherchons à limiter l'impact de l'ambiguïté potentiellement induite par certaines formes figurant dans les CT conservés après les deux premiers filtres.

Certaines formes parmi les CT retenus peuvent en effet renvoyer à différents concepts dans des domaines distincts (par exemple *rayon* ou *vitesse*). En langue de spécialité, la question de cette ambiguïté possible ne se pose généralement pas et l'on suppose que ces formes renvoient toujours (ou du moins le plus souvent) à des termes du domaine en question. Cependant, dans des textes non spécialisés, cette hypothèse de lien fort avec le domaine en question ne peut pas être faite puisque les textes ne sont pas caractérisés comme relevant d'un domaine en particulier. Cette ambiguïté possible n'est donc pas souhaitable dans l'optique de constitution du sous-corpus *Presse*.

Aussi, afin de repérer objectivement les CT susceptibles de renvoyer à différents termes de domaines distincts, nous exploitons cinq ressources terminographiques (le Grand dictionnaire terminologique, Termium, FranceTerme, IATE et TERMDAT¹⁰⁷) et observons le nombre de domaines dans lesquels chaque forme est enregistrée. Nous considérons que les formes renvoyant à plus de cinq domaines dans plus de trois de ces banques de données terminologiques ne doivent pas être retenues. Bien que ce choix comprenne une certaine part de subjectivité, nous jugeons que ce seuil de cinq domaines constitue un bon équilibre entre un seuil trop faible, qui risque de filtrer certains termes intéressants mais enregistrés dans les ressources dans des domaines connexes¹⁰⁸, et un seuil trop élevé, qui risque de laisser passer des termes non pertinents pour les raisons explicitées.

Parallèlement, nous sommes consciente que ce choix délibéré de filtrer la liste des CT risque d'écarter des CT réellement pertinents pour le domaine (par exemple *accélération*, alors que nous

¹⁰⁷ Grand dictionnaire terminologique : <http://gdt.oqlf.gouv.qc.ca/> ;
Termium : <http://www.btb.termiumplus.gc.ca/> ;
FranceTerme : <http://www.culture.fr/franceterme> ;
Interactive Terminology for Europe (IATE) : <https://iate.europa.eu/home> ;
TERMDAT : <https://www.termdat.bk.admin.ch/>.

Ces banques de données terminologiques ont été consultées pour l'application du troisième filtre en juillet 2017.

¹⁰⁸ Par exemple « rayonnements électromagnétiques » ou « magnétisme et électromagnétisme » dans Termium, pour des termes tels que *champ électromagnétique* et *électrodynamique quantique*, fondamentaux en physique des particules (selon les expertes).

conservons *accélérateur de particules*). Cependant, ce choix permet d'appliquer des critères de sélection et de rejet transparents et, à plus large échelle, d'ancrer notre travail dans une démarche reproductible. Par ailleurs, la liste que nous définissons à ce stade, outre son utilisation pour la constitution du sous-corpus *Presse*, représente surtout un point d'entrée dans les données lors de l'analyse. Selon nos observations et en fonction des besoins de notre analyse, d'autres termes peuvent être pris en considération s'ils se révèlent pertinents (par exemple si nous sommes amenée à analyser le paradigme *accélérateur de particules*, *accélérer des particules* et *accélération de particules*).

Ce troisième filtre permet de supprimer 107 CT, parmi lesquels *accélération*, *bouchon*, *tonneau*, *grille*, *jet*, *vecteur*, *énergie*, et 333 CT sont retenus à ce stade. En outre, seuls les CT attestés dans les quatre sous-corpus (*Spécialisé*, *Communiqués*, *Rapports* et *Vulgarisation*) constituent la liste de termes-clés servant à la constitution du sous-corpus *Presse*. C'est pourquoi, parmi ces 333 CT, nous ne conservons que les 191 CT attestés dans tous les quatre sous-corpus.

Enfin, un quatrième filtre est établi, qui est pensé pour cibler les CT les plus spécifiques au domaine de la physique des particules, en comparaison notamment avec des CT plus généralement utilisés dans le « macro-domaine » de la physique. La nécessité de cette étape de tri s'explique par le constat que, malgré les trois premiers filtres appliqués, la liste de CT obtenue automatiquement avec TermoStat reste très hétérogène, voire bruitée dans une certaine mesure. Certains CT semblent en effet moins spécifiques à la physique des particules que d'autres et semblent plutôt renvoyer à la physique en général. C'est par exemple le cas de *radioactivité*, *atomique*, *ion* ou encore *scalaire*. Aussi, puisque nous nous intéressons au processus de déterminologisation tel qu'il peut se produire dans le domaine de la physique des particules, nous cherchons à analyser en priorité les termes spécifiques à ce domaine et non les termes qui apparaissent de manière transversale dans les différents domaines de la physique. Cet aspect est important si, dans l'analyse, nous souhaitons nous assurer autant que possible que nos observations peuvent être mises en lien avec la déterminologisation dans ce domaine.

Nous élaborons ce filtre à partir d'une idée développée par Drouin *et al.* (2018). Selon les auteurs, la terminologie d'un domaine particulièrement vaste comme celui de l'environnement (que l'on pourrait qualifier de « macro-domaine »), qui comprend différents sujets, peut se diviser en deux couches lexicales (Drouin, 2018), ou *lexical layers* (Drouin *et al.*, 2018). Une première couche lexicale serait commune aux différents sujets composant le macro-domaine, tandis qu'une seconde couche serait spécifique à chacun des sujets (*Ibid.* : 3419). Drouin propose donc de distinguer le *lexique transdisciplinaire de domaine* (dans le premier cas) du *lexique propre aux thématiques* (dans le second cas). Les auteurs testent cette hypothèse sur un corpus spécialisé de l'environnement, composé de sous-corpus sur des sujets diversifiés, tels que le changement climatique, les énergies renouvelables

ou encore la pollution de l'eau (*Ibid.* : 3421). Dans cette perspective, la couche lexicale qui représente le lexique transdisciplinaire de domaine se compose des termes apparaissant de manière régulière dans les sous-corpus. Corrélativement, les termes particulièrement fréquents dans un seul des sous-corpus composeraient la couche lexicale spécifique au sujet du sous-corpus (*Ibid.* : 3420).

Dans notre cas, il semble que la physique peut également être considérée comme un macro-domaine, de la même manière que l'environnement. C'est pour cette raison que nous reprenons à notre compte cette hypothèse de deux couches lexicales distinctes et cherchons à isoler les CT qui feraient partie de la couche lexicale spécifique à la physique des particules, et donc à supprimer les CT qui feraient partie de la couche lexicale commune aux différents domaines de la physique.

L'identification de ces CT représente toutefois un enjeu important de la méthodologie mise en œuvre, d'autant plus si l'on souhaite poursuivre une démarche objective. En tant qu'analyste ne possédant pas de connaissances poussées en physique, au contraire des expertes avec lesquelles nous collaborons, il nous est impossible de distinguer le lexique propre aux thématiques du lexique transdisciplinaire de domaine. Une possibilité consisterait à nous appuyer sur le *Dictionnaire de physique* (Taillet *et al.*, 2018), que nous évoquons en 4.1.2. En effet, ce dictionnaire indique une marque de domaine pour une grande partie des termes qui y sont définis. Par exemple, l'entrée du terme *boson de Higgs* contient la marque « [Phys. des part.] », qui correspond à « Physique des particules » (*Ibid.* : 86). Deux problèmes se posent néanmoins, après une vérification systématique dans ce dictionnaire pour les termes qui nous intéressent.

Le premier problème est soulevé par l'absence de marque de domaine dans certaines entrées, comme pour le terme *électron-volt*, par exemple (*Ibid.* : 254-255). Dans ce cas, faut-il le considérer comme faisant partie du lexique transdisciplinaire de domaine et donc l'exclure de notre analyse ? Sans autre indication dans ce dictionnaire à propos des termes ne comportant pas de marque de domaine, il semble difficile de répondre à cette question et de décider du traitement de ces termes dans cette étape de notre travail. Le second problème concerne les termes qui ne figurent pas dans la nomenclature de ce dictionnaire, comme le terme *champ scalaire*, par exemple. Dans ce cas, et puisque les auteurs de ce dictionnaire ne prétendent aucunement à l'exhaustivité de sa nomenclature (*Ibid.* : vii-viii), il semblerait peu judicieux d'exclure un terme de notre analyse sur cette base uniquement.

Pour pallier ces difficultés, nous faisons le choix de faire intervenir les expertes et de leur faire valider les CT restants. Plus précisément, les expertes ont eu pour tâche de classer les CT selon s'ils leur semblent appartenir au lexique transdisciplinaire de domaine (à la physique de manière générale) ou au lexique propre à la physique des particules. À l'issue de cette tâche sont

retenus les CT validés par les deux expertes, c'est-à-dire qui sont classés comme des termes de physique des particules par toutes les deux expertes, réduisant ainsi la liste à 92 termes. Soulignons enfin que, bien qu'il puisse subsister une certaine part de subjectivité dans ce classement de la part des expertes, ce dernier filtre permet tout de même de préciser la sélection des termes constituant l'échantillon sur lequel nous basons notre analyse à partir de critères transparents et reproductibles.

Nous avons consacré ce chapitre à la description des principes sous-tendant la constitution de notre objet d'étude. En particulier, l'accent a été porté sur les différentes étapes de réflexion nécessaires à la compilation d'un corpus adéquat pour l'étude de la déterminologisation, tenant ainsi compte de facteurs tels que l'importance du choix du domaine sur les phénomènes que nous cherchons à mettre en lumière, le besoin de sélection des données sur la base de critères transparents et leur organisation en cinq sous-corpus afin de créer la possibilité d'interroger le processus de déterminologisation dans le continuum langue de spécialité-langue générale. Cette méthode de compilation solide et transparente garantit alors l'objectivité de la démarche et permet de contrôler au maximum les différentes variables entrant en jeu, afin de faire l'hypothèse que les fonctionnements des termes observés dans le corpus puissent être mis en relation avec la déterminologisation et ne soient pas la conséquence de biais de constitution des données.

Afin d'asseoir notre démarche *corpus-driven*, nous avons défini une méthode endogène de sélection des termes, appuyée par la mise en œuvre de différents filtres, visant à établir de manière objective et reproductible un échantillon de termes pertinents à analyser dans le corpus. C'est précisément dans la continuité de cette démarche *corpus-driven* que nous détaillons, dans le prochain chapitre, la définition et la mise en œuvre de deux indices distributionnels permettant d'appréhender le fonctionnement des termes dans le corpus et d'interroger la question du processus de déterminologisation.

Chapitre 5 Indices distributionnels

Ce chapitre porte sur la définition et la mise en œuvre de deux indices de type distributionnel sans interprétation *a priori* tels qu'ils sont abordés en terminologie textuelle (cf. Chapitre 3, 3.2.2). Dans ce but, nous contextualisons d'abord notre réflexion en rappelant l'origine de l'analyse distributionnelle et ses différentes évolutions vers l'automatisation de tout ou partie du processus (section 5.1). Puis nous explicitons les spécificités de la déterminologisation, de l'approche que nous en proposons et des objectifs que nous poursuivons, qui rendent l'analyse distributionnelle automatique peu adéquate. Ces considérations nous amènent à détailler le développement de deux indices complémentaires, à préciser leur intérêt pour notre travail et à expliquer la stratégie globale d'analyse (section 5.2). Enfin, dans la dernière section de ce chapitre, nous précisons les paramètres dont il faut tenir compte pour mettre en œuvre ces indices dans le corpus et discutons les différents choix à effectuer, afin de faire émerger des différences pertinentes, que nous pouvons interpréter en relation avec le processus de déterminologisation.

5.1. Analyse de la distribution en corpus

5.1.1. Définition et principe de l'analyse distributionnelle

À l'origine, l'analyse distributionnelle repose sur l'idée que le fonctionnement d'une langue peut être décrit sur la base de sa structure et que le fonctionnement d'une unité peut être entièrement décrit par les contextes dans lesquels elle apparaît (Harris, 1954)¹⁰⁹. On appelle alors *distribution* l'ensemble de ces contextes : « [t]he distribution of an element will be understood as the sum of all its environments » (Harris, 1954 : 146). De ce point de vue, le sens d'une unité peut être « calculé » à partir de sa distribution dans un texte ou dans un ensemble de textes. Ce sont donc les unités qui cooccurrent avec une certaine unité qui permettent d'appréhender et de décrire son sens, ce qu'évoque d'ailleurs la célèbre citation de Firth (1957 : 179) : « [y]ou shall know a word by the company it keeps ».

En outre, il est généralement admis que le sens d'une unité n'émerge pas directement de l'observation de sa distribution dans un corpus, mais qu'il est plutôt le résultat de comparaisons : c'est en comparant la distribution de deux unités, par exemple, que l'on peut estimer

¹⁰⁹ Précisons cependant que l'origine du courant distributionnaliste est à attribuer à Bloomfield (Tognini-Bonelli, 2001 : 170-173 ; Morlane-Hondère, 2013 : 22-23 ; Condamines, 2017 : 342). Harris et Firth ont ensuite développé et opérationnalisé l'analyse distributionnelle, d'un point de vue behaviouriste et mathématique pour le premier et sociolinguistique pour le second (Condamines, 2017 : 342). Pour un aperçu plus détaillé, voir Condamines (2000 : 11-14) et Morlane-Hondère (2013 : 22-30).

le degré de similarité ou de différence de leur sens. Selon Harris (1954 : 156), en effet, « [t]he fact that, for example, not every adjective occurs with every noun can be used as a measure of meaning difference ». De ce point de vue, l'analyse distributionnelle est donc à concevoir dans une perspective contrastive. Comme le dit Sahlgren (2008 : 3) lorsqu'il commente les fondements de l'analyse distributionnelle, « linguistic meaning is inherently differential, and not referential (since that would require an extra-linguistic component): it is *differences* of meaning that are mediated by *differences* of distribution ».

Cette citation laisse d'ailleurs entrevoir que, dans cette conception, aucun élément autre que le corpus étudié n'est considéré comme nécessaire à la description du sens d'une unité, à la condition que ce corpus soit constitué de textes cohérents d'un point de vue thématique, par exemple des textes qui relèvent de la langue de spécialité d'un domaine en particulier¹¹⁰. En effet, comme l'expliquent Habert et Zweigenbaum (2002 : 206), Harris et son équipe analysent un corpus de textes restreint, composé de 25 articles scientifiques dans le domaine de l'immunologie. Ces restrictions (de taille et de domaine) garantissent alors l'homogénéité des données considérées, laquelle permet d'observer des régularités distributionnelles (Condamines, 2000 : 12-13 ; Habert et Zweigenbaum, 2002 : 206-207 ; Morlane-Hondère, 2013 : 27-29).

Ainsi, le sens est toujours fonction des contextes pris en considération dans une étude en particulier. Cela ne signifie cependant pas que l'extralinguistique n'a aucune influence sur le sens ou n'est pas à prendre en compte dans une étude du sens, mais que, du point de vue de Harris, l'extralinguistique peut se refléter et s'observer dans la distribution : « even if extralinguistic factors *do* influence linguistic events, there will always be a distributional correlate to the event that will suffice as explanatory principle » (Sahlgren, 2008 : 3).

Parallèlement, la corrélation postulée entre des différences de distribution et des différences de sens permet de supposer qu'au contraire, une distribution similaire devrait indiquer un sens similaire. C'est ce qu'on appelle désormais l'*hypothèse distributionnelle*, dans des travaux plus récents (Sahlgren, 2008 ; Fabre et Lenci, 2015 ; Boleda, 2020 parmi d'autres). Lenci la formule ainsi : « [t]he DH [Distributional Hypothesis] states that the semantic similarity of lexical items is a function of their distribution in linguistic contexts » (Lenci, 2018 : 159). Ces travaux s'inspirent de la méthode harrissienne pour proposer des méthodes d'analyse distributionnelle automatique en corpus (que nous résumons en 5.1.3).

¹¹⁰ Notons que l'analyse distributionnelle telle qu'elle a été opérationnalisée par Harris, notamment, s'inscrit plus largement dans la théorie des sous-langages, dont les fondements et la méthodologie d'analyse sont détaillés entre autres par Harris (1982, 1991), Kittredge (1982), Lehrberger (1982), Sager (1982, 1986).

Par ailleurs, à la suite de Condamines (2003 : 28) ou Gries (2015 : 50), nous pourrions même considérer que toute analyse de corpus est fondamentalement une analyse distributionnelle : l'analyse du fonctionnement d'unités lexicales dans un corpus repose en effet le plus souvent, sinon toujours, sur les contextes d'apparition de ces unités dans le corpus, autrement dit sur leur distribution. Comme le dit Lenci¹¹¹,

« in corpus linguistics there is hardly any need to motivate the DH [Distributional Hypothesis] as a methodological principle for semantic analysis. Rather to the contrary, **this is often claimed to be the unique possible source of evidence for the exploration of meaning.** » (Lenci, 2008 : 6)

Dans cette perspective, l'analyse distributionnelle ne se conçoit donc plus réellement telle qu'elle a été mise en œuvre à l'origine par Harris. Deux aspects diffèrent en particulier : la définition des contextes distributionnels pris en compte dans l'analyse et la possibilité d'automatiser certaines étapes du processus d'analyse, qui induit notamment plusieurs simplifications de l'approche harrissienne. C'est ce que nous détaillons dans les deux sections suivantes.

5.1.2. Contextes distributionnels

Dans la méthode définie par Harris, tous les cooccurrents d'une unité ne constituent pas nécessairement le contexte distributionnel qui sert à appréhender son sens. En effet, seuls les cooccurrents qui entretiennent une relation de dépendance syntaxique avec l'unité sont à prendre en compte :

« the structural property is not merely co-occurrence, or even frequent co-occurrence, but rather dependence of a word on a set: an operator word does not appear in a sentence unless a word – one or another – of its argument set is there (or has been zeroed there). » (Harris, 1991 : 332)

Ce choix implique un travail poussé d'annotation syntaxique des données pour identifier ces cooccurrents, qui peut vite devenir chronophage en l'absence d'analyseurs syntaxiques performants (Morlane-Hondère, 2013 : 30). Par ailleurs, même avec les outils adéquats, la lourdeur des traitements est inévitable, si l'on souhaite appliquer fidèlement la méthodologie de Harris. En effet, Habert et Zweigenbaum (2003) expliquent que, « [p]our réduire la variation en surface et faciliter la mise en évidence des régularités », les dépendances sont normalisées par différentes opérations de transformation, par exemple le « passage des nominalisations au verbe sous-jacent » ou le « passage du passif à l'actif » (*Ibid.*). Ces transformations, qui sont caractéristiques de la méthode de Harris, sont effectuées par les analystes, avec le concours d'experts de domaine, car

¹¹¹ Nous soulignons.

« les acceptabilités et les transformations dépendent du domaine et la seule compétence de "locuteur natif" n'est pas toujours suffisante » (*Ibid.*). Ainsi, d'après Morlane-Hondère, c'est précisément cette caractéristique qui rend l'analyse distributionnelle selon Harris particulièrement complexe à implémenter :

« [u]ne des limites à l'application à grande échelle de la méthodologie harrissienne [...] est l'importance de l'intervention humaine dans le processus d'annotation syntaxique et de transformation des phrases. » (Morlane-Hondère, 2013 : 29)

Pour cette raison principalement, ces opérations de transformation ne sont le plus souvent pas effectuées dans les approches actuelles et seules les relations de dépendance syntaxique restent prises en compte. Dans ces approches, le contexte distributionnel d'une unité est donc défini comme l'ensemble des unités qui partagent une relation de dépendance avec cette unité (Bourigault, 2002 : 76 ; Van de Cruys, 2008 : 49 ; Peirsman *et al.*, 2010 : 471 ; Morlane-Hondère, 2013 : 40*sqq.* ; Lenci, 2018 : 158). Néanmoins, cette définition du contexte distributionnel n'est pas la plus répandue dans les approches actuelles et ce sont désormais des contextes composés de l'ensemble des mots graphiques apparaissant avant et/ou après l'unité étudiée, dans une certaine fenêtre, qui sont le plus souvent exploités (Tanguy *et al.*, 2015 : 104 ; Lenci, 2018 : 154 ; Hellrich, 2019 : 11-13). Ces contextes sont donc de type « fenêtre de mots » (Habert et Zweigenbaum, 2003 ; Morlane-Hondère, 2013 : 40*sqq.*).

Dans cette seconde définition du contexte, aucune opération de transformation n'est requise, car les relations syntaxiques entre les unités ne sont pas prises en compte et la mise en œuvre d'une analyse distributionnelle sur la base de ce type de contexte s'en trouve simplifiée (Pierrejean, 2020 : 14). C'est peut-être pour cette raison que ces contextes sont désormais plus répandus, ce que semble suggérer Morlane-Hondère. L'auteur explique que l'analyse distributionnelle à partir de contextes syntaxiques

« implique une étape de prétraitement supplémentaire qui consiste à fournir au système une analyse préalable des relations syntaxiques qui opèrent entre les constituants des phrases du corpus. Étant donnée l'absence d'analyseur syntaxique opérationnel, les premières tentatives d'automatiser l'AD [analyse distributionnelle] ont nécessité une annotation manuelle longue et fastidieuse. Cette contrainte a freiné l'application de l'AD sur des corpus de grande taille. » (Morlane-Hondère, 2013 : 40)

Soulignons enfin qu'aujourd'hui, l'automatisation de l'analyse distributionnelle rend ces deux types de contextes tout à fait opérationnels (cf. section suivante). Le choix de l'un ou l'autre peut alors dépendre de l'objectif poursuivi et du type de rapprochements que l'on souhaite mettre

au jour. En effet, il est généralement reconnu que les résultats produits sont de nature différente selon le type de contexte pris en compte. Selon Heylen et Bertels (2016 : 54-55) et Fišer et Ljubešić (2019 : 166), les contextes syntaxiques permettent d'aborder le sens des unités de manière plus précise. En particulier, dans des tâches de rapprochement d'unités sur la base de leur similarité distributionnelle, ce type de contexte permet d'effectuer des regroupements d'unités entretenant une relation paradigmatique, le plus souvent de type hyperonymique (en particulier des couples de co-hyponymes) (Peirsman *et al.*, 2010 : 471-472 ; Fabre et Lenci, 2015 : 10), mais également de type synonymique et antonymique (Heylen et Bertels, 2016 : 54-55 ; Lenci, 2018 : 158). En revanche, les fenêtres de mots permettent d'opérer des rapprochements sémantiques plus « grossiers ». Fabre et Lenci (2015 : 10) parlent de relations associatives ; Peirsman *et al.* (2010 : 471-472) précisent par ailleurs que ce type de résultats est le plus souvent obtenu avec des contextes plus grands que des contextes syntaxiques. Enfin, selon Morlane-Hondère (2013 : 43) et Pierrejean (2020 : 15), les associations fournies par des contextes à fenêtres de mots sont de type thématique.

Ainsi, étant donné ces différents types de résultats, le choix de l'un ou l'autre des types de contextes dépend alors essentiellement de l'objectif d'analyse poursuivi. Soulignons également que l'automatisation de l'analyse distributionnelle rend possible l'analyse de données plus volumineuses, facilite l'analyse syntaxique des données par le recours à plusieurs outils et permet d'aborder finement certains phénomènes, selon le « grain » d'analyse choisi. La part d'automatisation de l'analyse dépend, elle aussi, de plusieurs facteurs.

5.1.3. Automatisation de l'analyse distributionnelle

Nous l'avons dit, l'un des écueils de la méthode harrissienne est généralement attribué à la lourdeur des traitements à appliquer aux données pour permettre leur analyse. En traitement automatique des langues, cependant, la possibilité de gérer de grands volumes de textes informatisés¹¹² ainsi que la puissance de calcul désormais offerte par la technologie rend possible l'implémentation de différentes méthodes d'analyse distributionnelle automatique (Bourigault et Fabre, 2000 : 133 ; Fabre et Lenci, 2015 : 8 ; Heylen et Bertels, 2016 : 51*sqq.* par exemple).

L'analyse distributionnelle profite alors de l'automatisation de certains traitements, non seulement pour l'analyse syntaxique lorsque les contextes sont de type syntaxique, mais également pour d'autres étapes dans la chaîne de traitement (repérage des contextes distributionnels, calcul de la similarité de ces contextes, rapprochement d'unités sur la base de leurs similarités

¹¹² De l'ordre de plusieurs centaines de millions d'occurrences, voire de plusieurs milliards (cf. 5.2.1.1).

distributionnelles, etc.) (Morlane-Hondère, 2013 : 30-35 ; Boleda, 2020 : 214-216). Différents outils informatiques ou langages de programmation sont exploités dans ce but (cf. Chapitre 3).

En outre, l'analyse distributionnelle automatique peut se décliner « en une variété de modèles qui se distinguent en fonction de la configuration de paramètres adoptée (utilisation de données analysées syntaxiquement, type de mesure de similarité, etc.) » (Morlane-Hondère, 2013 : 34)¹¹³. Ces modèles « combinent de façon systématique des mesures d'association et des méthodes statistiques multivariées dans le but d'explorer des structures sémantiques lexicales dans des corpus textuels de taille importante » (Heylen et Bertels, 2016 : 52), rendant ainsi possible l'extraction de divers éléments qui sont ensuite interprétés en fonction d'un objectif de recherche précis¹¹⁴.

Différentes applications sont alors visées. Par exemple, l'acquisition automatique de relations sémantiques, le plus souvent dans un but de construction de ressources, est un domaine de l'analyse distributionnelle automatique particulièrement productif. Les travaux se basent sur l'hypothèse distributionnelle pour faire émerger d'un corpus des unités qui partagent une certaine relation, telle que l'hyponymie, la synonymie ou l'antonymie (Baneyx *et al.*, 2005 ; Galy et Bourigault, 2005 ; Bannour *et al.*, 2011 ; Morlane-Hondère, 2013 ; Bernier-Colborne, 2014 ; Hazem et Daille, 2015 ; Périnet et Hamon, 2015). Tous ne poursuivent pas le même but : Périnet et Hamon (2015), par exemple, cherchent à adapter les méthodes d'analyse distributionnelle automatique à de petits corpus et l'acquisition de relations est en réalité une étape de leur méthodologie. Bannour *et al.* (2011), en revanche, mesurent la similarité entre termes complexes, et non entre mots, à partir d'un corpus spécialisé dans un objectif d'aide à la constitution d'ontologies. La particularité ici réside dans l'adaptation des méthodes d'analyse distributionnelle automatique à l'analyse des termes complexes (et non uniquement des termes simples). C'est également dans ce sens que vont les propositions d'Hazem et Daille (2015), qui cherchent à mettre au jour des synonymes de termes complexes dans un corpus de textes de spécialité.

À ce stade, il nous semble important d'insister sur le fait que, dans ces travaux, et de manière générale en analyse distributionnelle, il peut rester une certaine part d'interprétation des résultats fournis automatiquement. Par exemple, sur la base des rapprochements effectués par les modèles

¹¹³ Notons que certains modèles d'analyse distributionnelle tiennent compte non seulement des contextes distributionnels, mais également des documents dans lesquels les unités apparaissent. L'analyse sémantique latente (LSA, *Latent Semantic Analysis*) est le modèle de ce type le plus souvent cité (Sahlgren, 2008 ; Sagi *et al.*, 2009 ; Heylen et Bertels, 2016 ; Lenci, 2018). Dans notre cas, néanmoins, nous ne retenons pas ce type de modèle, vu le volume restreint de données dont nous disposons et l'hétérogénéité des différents documents inclus dans le corpus (cf. Chapitre 4).

¹¹⁴ Pour un panorama des modèles existants, voir Morlane-Hondère (2013 : 63-71) ou Lenci (2018 : 157-160).

distributionnels, ce sont bien les analystes qui identifient la relation que partagent les unités rapprochées¹¹⁵. C'est ce qu'explique Morlane-Hondère¹¹⁶ :

« le caractère imparfait des regroupements opérés montre qu'il est trop ambitieux de vouloir générer automatiquement des classes d'objets : **les résultats obtenus sont en effet loin de constituer un matériau fini. Ils peuvent cependant être exploités comme une base de départ** par les experts, pour la construction d'une ontologie, par exemple, dans une approche semi-automatique (Poibeau *et al.*, 2002). Les résultats de cette étude mettent ainsi l'accent sur le fait que la mise au jour de classes pertinentes relève avant tout d'un travail interprétatif. » (Morlane-Hondère, 2013 : 58)

Ainsi, s'il est possible de repérer automatiquement des unités qui sont sémantiquement liées, la nature des relations qu'entretiennent ces unités reste parfois difficile à caractériser. Par exemple, il n'est pas toujours possible de différencier synonymie et antonymie sur la base des résultats obtenus automatiquement, car un important degré de similarité dans les contextes distributionnels de deux unités peut tout aussi bien indiquer que les unités sont synonymes ou antonymes (Sahlgren, 2008 : 4 ; Fabre et Lenci, 2015 : 13).

Dans cette perspective, l'accent est porté sur les différences et les similarités dans la distribution de plusieurs unités dans le même corpus, qui sont interprétées en rapport avec le sens de ces unités. Mais il est également possible d'observer des différences de sens pour la même unité à travers sa distribution dans des corpus distincts. Ces travaux se situent dans différentes perspectives : d'une part, certains travaux adoptent un point de vue diachronique et se focalisent sur la mise au jour et la modélisation de changements sémantiques (Sagi *et al.*, 2009 ; Kutuzov *et al.*, 2018 ; Del Tredici *et al.*, 2019 ; Fišer et Ljubešić, 2019 ; Hellrich, 2019 ; Boleda, 2020). Par exemple, Fišer et Ljubešić (2019) cherchent à identifier l'émergence de nouveaux sens de certaines unités en slovène dans un corpus de tweets par rapport à un corpus de référence. De la même manière, Del Tredici *et al.* (2019) observent l'évolution du sens des unités dans un corpus composé de messages d'un forum de supporters de football à travers la comparaison de deux courtes périodes entre 2011 et 2017. Les changements ainsi identifiés relèvent majoritairement de métaphores, de métonymies et de jeux de mots (*Ibid.* : 2070-2072). D'autre part, Peirsman *et al.* (2010) s'intéressent au repérage

¹¹⁵ Là encore, les méthodes actuelles d'analyse distributionnelle diffèrent de l'approche proposée à l'origine par Harris. Comme l'explique Condamines, « l'ambition de Harris consiste à comprendre comment des agencements de formes (les distributions de mots) peuvent créer du sens. En principe, il s'agit donc d'évacuer le sens de l'analyse puisqu'il doit émerger de la mise au jour de régularités de forme. Ainsi, au moins dans la vision initiale de Harris, il n'y a pas d'interprétation des résultats car les dépendances relationnelles mises au jour sont le reflet de la perception humaine » (Condamines, 2003 : 27).

¹¹⁶ Nous soulignons.

et à la caractérisation de la variation lexicale diatopique, entre les variétés du néerlandais de Belgique et des Pays-Bas.

Soulignons enfin que la majorité des travaux en analyse distributionnelle automatique mettent désormais en œuvre des modèles prédictifs, basés sur des réseaux de neurones et représentant les données dans un espace vectoriel (Bernier-Colborne et Drouin, 2016 ; Bertels et Speelman, 2016 ; Lenci, 2018 ; Pierrejean et Tanguy, 2018 ; Del Tredici *et al.*, 2019 ; Boleda, 2020). L'attrait de ces modèles peut s'expliquer notamment par leurs bonnes performances (Boleda, 2020 : 215), malgré le manque de reproductibilité inhérent aux méthodes utilisées (Pierrejean et Tanguy, 2018 : 34).

Cependant, malgré des résultats jugés de bonne qualité, les modèles d'analyse distributionnelle automatique¹¹⁷ présentent certaines limites, qui peuvent freiner leur implémentation, en particulier lorsque certains phénomènes précis sont visés. Dans notre cas, les spécificités du processus de déterminologisation et de l'approche en corpus que nous proposons rendent difficile la mise en œuvre de modèles automatiques et requièrent la définition d'une méthode qui permette de saisir les différences distributionnelles que nous visons et qui sont parfois subtiles.

5.2. Définition d'une méthode d'analyse de la distribution adaptée à notre étude

Dans cette section, nous définissons deux indices observables en corpus adéquats pour notre objet d'étude et pour les données constituées. Dans ce but, nous explicitons d'abord en 5.2.1 plusieurs limites des méthodes d'analyse distributionnelle automatique pour notre étude. Ces méthodes, bien qu'elles présentent de nombreux avantages, ne semblent pas les plus adaptées à l'approche de la déterminologisation que nous proposons dans notre thèse. Nous nous inspirons néanmoins des travaux en analyse distributionnelle automatique pour définir et mettre en œuvre les deux indices. C'est ce que nous expliquons en 5.2.2. Enfin, nous explicitons en 5.2.3 la stratégie globale de comparaison des sous-corpus que nous développons, afin de permettre l'interprétation des comparaisons et la mise au jour de différences de fonctionnement pertinentes.

¹¹⁷ Que l'on appelle également *sémantique distributionnelle* (Lenci, 2008 ; Fabre et Lenci, 2015).

5.2.1. Limites de l'analyse distributionnelle automatique

5.2.1.1. Limites générales des méthodes automatiques actuelles

Nous l'avons dit, les résultats des méthodes d'analyse distributionnelle automatique ne sont généralement pas une finalité et requièrent une étape supplémentaire d'interprétation. Par ailleurs, dans des tâches d'acquisition automatique de relations, la difficulté à caractériser les relations que partagent les unités repérées est une réelle préoccupation des chercheurs (Sahlgren, 2008 : 4 ; Morlane-Hondère, 2013 : 58). En effet,

« l'évaluation des relations acquises par analyse distributionnelle reste aujourd'hui une problématique importante et il est difficile d'évaluer une méthode distributionnelle en raison de la grande variété de relations qu'elle produit. » (Périnet et Hamon, 2015 : 90)

Ces remarques mettent ainsi en évidence le fait que les représentations sémantiques issues de méthodes d'analyse distributionnelle automatique restent nécessairement partielles, imparfaites¹¹⁸. Comme le dit Lenci, « DS [Distributional Semantics] often provides a quite coarse-grained representation of semantic content » (Lenci, 2018 : 165). Même si cet aspect n'est pas toujours problématique, ces représentations peuvent tout de même être un frein à l'analyse de certaines questions. Par exemple, selon Desagulier, les méthodes de représentation vectorielle éprouvent toujours certaines difficultés à gérer la représentation des unités polysémiques et des expressions polylexicales (Desagulier, 2019 : 232-235). L'auteur prend comme exemple l'adjectif anglais *hot*, dont les 21 sens possibles enregistrés dans WordNet 3.1 ne sont représentés que par un seul vecteur. Selon l'auteur, si l'on poursuit un but d'analyse qualitative fine, alors la représentation de l'ensemble des sens possibles de cet adjectif peut se révéler problématique, car toutes les nuances qui caractérisent ces différents sens ne sont pas saisies par un unique vecteur.

En outre, bien que certains travaux se focalisent sur le repérage et la caractérisation des unités polysémiques, notamment Heylen et Bertels (2016) dans un corpus de spécialité et dans un objectif lexicographique, selon Fabre (2015 : 401), « le traitement de la polysémie constitue généralement un angle mort des travaux en sémantique distributionnelle ».

¹¹⁸ Par ailleurs, les approches prédictives actuelles, basées sur des réseaux de neurones, se caractérisent particulièrement par leur manque de reproductibilité et par l'instabilité des représentations sémantiques qu'elles génèrent. Ces aspects constituent également un point de critique important, dont il faut tenir compte lorsque l'on met en œuvre de telles approches (Pierrejean et Tanguy, 2018). Pour une discussion détaillée, voir en particulier les chapitres 4 à 6 de la thèse de B. Pierrejean (Pierrejean, 2020) et Hellrich (2019 : 57-95).

Un second point de critique important que nous pouvons soulever concerne la taille des données nécessaires à la production de résultats de qualité dans les modèles actuels d'analyse distributionnelle automatique. Il est en effet généralement reconnu que de gros volumes de données sont nécessaires (Fabre, 2015 : 397 ; Hellrich, 2019 : 49sq. ; Boleda, 2020 : 218), lesquels permettent de neutraliser, dans une certaine mesure, les biais associés à un manque de diversité au niveau des sujets traités dans les textes du corpus ou au niveau de la variété des locuteurs pris en compte, par exemple (Desagulier, 2019 : 228). Bien que les modèles actuels d'analyse distributionnelle produisent parfois des résultats de qualités différentes en fonction de la taille des corpus utilisés, il semble que des corpus de l'échelle d'un million d'occurrences restent tout de même insuffisants, surtout lorsque des modèles neuronaux sont mis en œuvre (Sahlgren et Lenci, 2016 : 977-979). Par ailleurs, les expérimentations de ces deux auteurs montrent également que, plus les données sont volumineuses, meilleure est la qualité des résultats (*Ibid.* : 978).

Pour cette raison, les corpus désormais utilisés peuvent aller jusqu'à plusieurs milliards d'occurrences (Tanguy et Fabre, 2014 : 14-16 ; Boleda, 2020 : 218), par exemple le corpus frWaC, qui contient des textes provenant du Web et qui compte environ 1,3 milliard d'occurrences, comme l'explique Wauquier (2018 : 29), ou le corpus Google Books, qui dépasse les 100 milliards d'occurrences (Tanguy et Fabre, 2014 : 14). Cependant, il semble que des corpus plus modestes restent le plus souvent utilisés. Par « modestes », nous entendons au moins 100 millions d'occurrences, comme dans le *British National Corpus* et le corpus des actes des conférences de l'*Association for Computational Linguistics (ACL Anthology)* (Pierrejean, 2020 : 69-73), mais également des corpus de 200-250 millions d'occurrences (tels que le corpus Wikipédia ou le corpus Le Monde) (Wauquier, 2018 : 29).

Ce besoin de données massives constitue par ailleurs l'un des défis actuels de la sémantique distributionnelle (Boleda, 2020 : 218). Bien que l'auteur évoque ce défi dans un contexte de sémantique diachronique et précise que les données diachroniques suffisamment volumineuses sont rares de ce point de vue, à l'exception peut-être de données pour l'anglais (*Ibid.*), cette remarque peut s'appliquer plus largement à d'autres phénomènes linguistiques. En effet, ce besoin de très gros corpus remet en question la cohérence et l'homogénéité des textes inclus dans le corpus (Tanguy et Fabre, 2014 : 15) et l'analyse distributionnelle ne sert plus nécessairement à « l'analyse de textes spécialisés aux caractéristiques lexicales et syntaxiques contraintes » (Fabre, 2015 : 401), où la cohérence et l'homogénéité du corpus sont deux critères cruciaux, mais s'applique désormais « à des textes de nature diverse, impliquant moins de redondance en termes de vocabulaire et de structures syntaxiques, et plus de polysémie » (*Ibid.*).

Ces remarques nous amènent également à souligner une limite plus générale, qui concerne la capacité des modèles automatiques à saisir certains phénomènes. En effet, il n'est pas sûr que ces modèles permettent d'aborder finement certaines questions sémantiques : « whether statistical co-occurrences alone are enough to address deep semantic questions, or whether they merely provide a shallow proxy of lexical meaning, remains an open question » (Lenci, 2018: 160). L'auteur va même plus loin et affirme que « [s]everal aspects of meaning [...] are still unexplored and may lie beyond its [Distributional Semantics] scope » (*Ibid.* : 165). De ce point de vue, les méthodes de sémantique distributionnelle ne semblent donc pas les plus adaptées à des situations où un grain d'analyse fin est requis et où l'on vise des descriptions approfondies des phénomènes repérés, dans des données plutôt restreintes. C'est ce que nous détaillons dans la section suivante, en lien avec l'approche de la déterminologisation que nous proposons.

5.2.1.2. Spécificités de notre objet d'étude pour l'analyse distributionnelle

Notre approche de la déterminologisation et les données constituées pour observer le fonctionnement des termes dans des textes de différents genres et de différents degrés de spécialisation rendent difficile la mise en œuvre de méthodes d'analyse distributionnelle automatique, en particulier les représentations vectorielles. En effet, la taille de nos sous-corpus (d'environ 200 000 occurrences à 1 million, cf. Chapitre 4, 4.2.2) semble trop limitée, comme cela peut être le cas pour les corpus de langues de spécialité, lorsqu'un domaine plutôt restreint est visé¹¹⁹. D'après Périnet et Hamon,

« [L]es modèles vectoriels ont l'avantage de permettre une quantification facile de la proximité sémantique entre deux mots. Cependant, ils souffrent d'un problème de dispersion des données, car ils s'appuient sur des espaces aux très grandes dimensions et sur la redondance des contextes partagés alors qu'il s'agit d'évènements souvent rares (Chatterjee et Mohan, 2008). [...] Lorsqu'il s'agit de corpus de spécialité, ce problème de dispersion des données est accentué par des tailles de corpus beaucoup plus petites, un faible nombre d'occurrences du vocabulaire et un nombre de contextes partagés plus faibles. » (Périnet et Hamon, 2015 : 78)

En outre, même si des travaux récents se focalisent sur l'adaptation des méthodes distributionnelles automatiques à des corpus de taille plus réduite (tels que Périnet et Hamon,

¹¹⁹ Par ailleurs, lorsque des corpus plus volumineux sont constitués en langues de spécialité, les textes qui les composent semblent couvrir des domaines plutôt étendus (par exemple le corpus Panacea qui contient environ 50 millions d'occurrences (<http://catalog.elra.info/en-us/repository/browse/ELRA-W0063/>, page consultée le 25 mars 2021) et qui, comme le précisent Bernier-Colborne et Drouin (2016 : 53), regroupe des textes liés au domaine de l'environnement en général), voire couvrent plusieurs domaines (comme dans le cas du corpus Scientext (Tutin, 2007a ; Tutin *et al.*, 2009)).

2015 ; Tanguy *et al.*, 2015 ; Bertels et Speelman, 2016), il n'est pas certain que nos données soient adaptées pour l'exploitation de ces méthodes. Ainsi, il n'est pas sûr que les fréquences des termes que nous analysons et de leurs cooccurrents soient suffisamment élevées dans les cinq sous-corpus pour mettre en évidence des différences significatives ou que les différences dans la distribution d'un terme soient suffisamment marquées entre les cinq sous-corpus pour être repérées. Par exemple, nous savons que certaines manifestations de la déterminologisation consistent en l'utilisation métaphorique de termes (cf. Chapitre 2). Comme le montrent notamment Morlane-Hondère (2013 : 139-140), Del Tredici *et al.* (2019 : 2071-2072) et Boleda (2020 : 216-219), les métaphores se manifestent dans les données par un écart significatif de distribution, mais cet écart de distribution ne peut être repéré de manière automatique que s'il est suffisamment récurrent dans le corpus, comme cela peut être le cas de métaphores (ou de métonymies) qui se stabilisent dans l'usage et qui constituent des néologismes sémantiques (cf. Chapitre 1, section 1.4). Dans ce cas, leur repérage se fonde précisément sur des emplois répétés d'une unité avec des cooccurrents qui se distinguent dans une large mesure de leurs cooccurrents « habituels ».

De la même manière, la taille des données joue un rôle important dans le repérage de ce type d'écarts de distribution, car des données très nombreuses peuvent maximiser les chances de voir se répéter un usage particulier. Or, dans le cas de la déterminologisation et vu la taille restreinte de nos données, il n'est pas sûr que les usages métaphoriques des termes que nous nous attendons à observer soient suffisamment répandus pour être repérés automatiquement. Par ailleurs, même lorsque les écarts de distribution sont mis en évidence par des méthodes automatiques, seule une interprétation humaine des résultats permet de décider que ces écarts renvoient effectivement à une métaphore (ou à une métonymie), comme l'expliquent notamment Del Tredici *et al.* (2019 : 2071*sqq.*).

Enfin, précisons que nous cherchons également à observer des changements sémantiques plus subtils, que Meyer et Mackintosh relient au degré de compréhension des concepts, qui diffère lorsque les locuteurs ne sont pas des experts (Meyer et Mackintosh, 2000a : 202-205). Dans ce cas, selon les autrices, « l'essence du sens terminologique demeure » (*Ibid.* : 202) (cf. Chapitre 2, 2.1.1.3). De la même manière, nous faisons l'hypothèse que différents types de changements peuvent survenir pour un même terme et qu'ils ne seraient pas nécessairement répartis de manière équilibrée dans le corpus. Nous supposons alors à ce stade que les écarts dans la distribution des termes ne seraient peut-être pas toujours aussi marqués que lorsque les termes sont utilisés métaphoriquement. La méthode d'analyse de la distribution que nous mettons en œuvre doit donc permettre de saisir et d'observer les différentes manifestations de la déterminologisation dans le corpus, que les usages des termes s'éloignent fortement de leurs usages par les experts ou non.

Dans ce contexte, vu les particularités de notre objet d'étude tel que nous l'abordons et vu plusieurs limites de l'analyse distributionnelle automatique pour notre objet d'étude, nous faisons le choix d'adopter une méthode d'analyse outillée de la distribution des termes qui permet de mettre au jour différents fonctionnements des termes dans les sous-corpus, que nous pouvons interpréter en relation avec le processus de déterminologisation. Plus précisément, nous mettons en œuvre deux indices de type distributionnel, dont la complémentarité permet de décrire finement les fonctionnements des termes dans le corpus. Nous définissons ces deux indices et explicitons leur intérêt pour la déterminologisation en 5.2.2 et nous détaillons leur mise en œuvre, qui s'inspire des travaux en analyse distributionnelle automatique, dans la section 5.3.

5.2.2. Définition d'indices distributionnels pertinents

Les deux indices de type distributionnel que nous exploitons pour appréhender les changements sémantiques des termes dans le corpus font intervenir les deux types de contextes distributionnels généralement utilisés en analyse distributionnelle, c'est-à-dire des fenêtres de mots et des contextes syntaxiques. Nous avons vu dans la section 5.1.2 que, dans les approches automatiques de l'analyse distributionnelle, ces deux types de contextes permettent d'opérer des rapprochements différents. Dans cette perspective, nous pensons que la combinaison des deux types de contextes fournira des informations utiles et complémentaires sur le fonctionnement des termes dans les sous-corpus. Par exemple, une analyse à partir de fenêtres de mots permettrait d'observer des changements de nature thématique dans la distribution des termes, entre les sous-corpus, ce qui pourrait nous fournir un aperçu global des éventuels points de vue différents portés sur un terme, selon le sous-corpus. De plus, la prise en compte de contextes syntaxiques permettrait d'appréhender plus finement le sens des termes, mettant ainsi en évidence des différences d'utilisation peut-être plus subtiles entre les sous-corpus. Nous visons donc à tirer profit des différents résultats produits par ces deux manières de définir le contexte, dans le but d'observer et de rendre compte de la diversité des phénomènes sémantiques qui se produisent en situation de déterminologisation.

5.2.2.1. Mesurer les différences lexicales dans la distribution des termes : repérage de *keywords*

Le premier indice de type distributionnel que nous mettons en œuvre concerne la distribution des termes dans des contextes de type « fenêtre de mots ». Dans l'approche outillée que nous développons, nous proposons d'aborder l'analyse de ces contextes par le biais d'un repérage de *keywords*. Cette méthode permet d'identifier rapidement et objectivement les différences

lexicales les plus saillantes dans la distribution des termes dans deux sous-corpus (le sous-corpus *Spécialisé* et le sous-corpus *Presse*). Dans ce but, nous ne confrontons pas les deux sous-corpus entiers, comme c'est généralement le cas avec cette méthode, mais des sous-ensembles de ces sous-corpus, composés des contextes distributionnels des termes à l'étude. Ce choix repose sur l'hypothèse que les différences les plus saillantes dans la distribution des termes permettent de mettre en évidence des fonctionnements des termes qui peuvent ensuite être mis en relation avec le processus de déterminologisation.

Dans cette section, nous définissons donc d'abord le terme *keyword* et la méthode habituellement utilisée pour repérer les *keywords* en corpus, puis nous détaillons l'intérêt que cette méthode représente pour aborder la distribution des termes dans nos données et pour nos objectifs. La mise en œuvre de cet indice ainsi que les différents paramètres dont il faut tenir compte seront détaillés en 5.3.1.

L'analyse de *keywords* est une méthode répandue en linguistique de corpus (Culpeper et Demmen, 2015 : 99 ; Bestgen, 2017 : 36 ; Egbert et Biber, 2019 : 77) surtout dans des études anglo-saxonnes, comme le montrent par exemple les ouvrages de Scott et Tribble (2006) et de Bondi et Scott (2010). En linguistique de corpus, le terme *keyword* renvoie à une unité dont la fréquence est inhabituelle dans un corpus, en comparaison avec un autre corpus. L'origine de cette définition revient à Scott (1997), pour qui

« [a] key word may be defined as a *word which occurs with unusual frequency in a given text*. This does not mean high frequency but unusual frequency, by comparison with a reference corpus of some kind. » (Scott, 1997 : 236)

Cette idée de fréquence inhabituelle renvoie en réalité aussi bien à une fréquence très élevée qu'à une fréquence très faible (Scott, 1997 : 243). Le repérage de *keywords* se fonde ainsi sur la comparaison de deux corpus, un corpus d'analyse et un corpus de référence, afin de faire ressortir du corpus d'analyse les unités dont la fréquence diffère dans une large mesure de leur fréquence dans le corpus de référence. Précisons néanmoins que le plus souvent seules les unités démontrant une fréquence plus élevée dans le corpus d'analyse que dans le corpus de référence sont étudiées (comme le montrent par exemple les travaux de Bassi (2010) ou Milizia (2010), ou la synthèse de Pojanapunya et Watson Todd (2018)), les *keywords* dits *négatifs* (Scott, 1997 : 243) n'étant pas pris en compte.

L'analyse de *keywords* permet de poursuivre de multiples objectifs de recherche. Les *keywords* peuvent en effet servir à identifier de manière générale le thème ou le domaine d'un texte ou d'un corpus, ce que l'on désigne le plus souvent par le terme *aboutness* (Scott, 1997 ; Scott et Tribble,

2006 ; Bondi, 2010 ; Gabrielatos, 2018 ; Pojanapunya et Watson Todd, 2018 ; Egbert et Biber, 2019). En analyse de discours politiques, cette méthode sert à identifier le positionnement idéologique d'une certaine personnalité ou à caractériser différents points de vue sur un sujet (Scott, 1997 ; Bassi, 2010 ; Milizia, 2010). L'utilité des *keywords* pour l'analyse et la caractérisation des genres est également mise en avant (Culpeper et Demmen, 2015 ; McEnery, 2016 ; Pojanapunya et Watson Todd, 2018 ; Egbert et Biber, 2019), ainsi que différentes applications pédagogiques, pour l'enseignement d'une langue étrangère ou d'une langue de spécialité (Scott, 1997 ; Gries, 2015 ; Pojanapunya et Watson Todd, 2018). Enfin, certains travaux mettent en œuvre cette méthode d'une manière plus originale. Par exemple, Philip (2010) propose de mettre au jour ce qu'il appelle des « thèmes métaphoriques » (2010 : 191) dans un corpus de textes politiques à partir du repérage de *keywords* et Rayson (2008) associe un corpus annoté sémantiquement aux méthodes d'extraction de *keywords* pour repérer non pas les unités individuelles, mais les champs sémantiques qui se distinguent dans un corpus par rapport à un autre.

Dans ces travaux, le repérage de *keywords* est toujours effectué par la comparaison de deux corpus et les analyses se focalisent ensuite sur les unités lexicales dont la fréquence est inhabituelle dans un corpus par rapport à l'autre. Cependant, dans notre cas, nous proposons de mettre en œuvre un repérage de *keywords* pour aborder la distribution des termes. Nous adaptons donc cette méthode pour les besoins de notre analyse et en fonction de nos données, comme le suggère par exemple Habert (2004, 2006). L'auteur explique en effet que les spécificités de chaque objet d'étude en linguistique et/ou de chaque orientation théorique dans laquelle les analyses se situent ne peuvent pas être prises en charge par les outils de traitement et d'exploration de corpus disponibles ou par les méthodes statistiques existantes (Habert, 2004 : 10). C'est au contraire à l'analyste d'adapter les outils, et même de les détourner lorsque c'est nécessaire, afin de rendre possible l'exploration de certaines données et l'interrogation de certaines problématiques. Comme le dit l'auteur, « les instruments se prêtent au détournement qui leur permet de traiter des données pour lesquelles ils n'avaient pas été prévus initialement » (Habert, 2006).

Dans ce contexte, nous pensons qu'une extraction de *keywords* présente plusieurs avantages pour l'analyse de la distribution des termes dans les sous-corpus. Tout d'abord, si nous détournons cette méthode (pour reprendre le terme employé par Habert), nous pouvons l'appliquer non pas aux sous-corpus en entier, mais à des sous-ensembles, constitués des contextes distributionnels des termes à l'étude. L'intérêt de ne considérer que les environnements distributionnels des termes dans cette étape d'identification automatique de *keywords* réside dans la possibilité de mettre au jour objectivement les unités lexicales qui apparaissent avec une fréquence particulièrement élevée dans les environnements des termes dans le SC-P par rapport au SC-S. Nous faisons ici l'hypothèse que

ces unités apparaîtraient dans des contextes démontrant un fonctionnement particulier des termes à l'étude, voire des changements sémantiques « profonds », que nous pourrions attribuer au processus de déterminologisation. De ce point de vue, les unités repérées comme *key* dans le sous-ensemble du SC-P permettraient de cibler les contextes des termes dans le SC-P où les écarts de distribution avec les contextes du SC-S sont les plus marqués.

En outre, bien qu'il soit probable que chaque terme fonctionne d'une manière particulière dans le processus de déterminologisation, nous faisons le choix de considérer les contextes distributionnels de tous les termes à l'étude dans cette étape (sur l'échantillon de termes analysés, voir 5.3.1.1). Ce choix s'explique par une volonté de mettre en lumière des phénomènes communs à différents termes, qui permettraient de mieux comprendre les changements s'opérant au cours du processus de déterminologisation. Ainsi, avec ce premier indice, nous cherchons en priorité à repérer des comportements communs aux différents termes, qui seraient observables à travers leur distribution, par exemple des thématiques particulières qui ressortent, des usages métaphoriques réguliers dans des domaines différents, un vocabulaire particulier utilisé autour de ces différents termes, etc. Le second indice dont nous traitons à la section suivante permet d'orienter nos observations plus précisément sur chacun des termes à l'étude.

Par ailleurs, l'extraction de *keywords* constitue une méthode assez peu coûteuse visant à obtenir rapidement un premier aperçu des unités lexicales qui sont significativement plus fréquentes dans les contextes des termes dans le SC-P par rapport au SC-S¹²⁰. De même, cette méthode se base uniquement sur les fréquences d'apparition des unités dans les données et les calculs habituellement utilisés n'impliquent aucune évaluation des données *a priori* de la part de l'analyste, ce qui permet d'adopter un point de vue objectif. Cette méthode s'inscrit ainsi le plus souvent dans une démarche exploratoire (Gabrielatos, 2018 : 227). Selon Culpeper et Demmen (2015 : 90), par exemple,

« [t]he advantages of this [quantitative] basis are twofold. First, it is less subject to the vagaries of subjective judgments of cultural importance. Second, it does not rely on researchers selecting items that might be important and then establishing their importance, but can reveal items that researchers did not know to be important in the first place ».

Cette citation met l'accent sur l'aspect résolument *data-driven* (ou *corpus-driven*) du repérage de *keywords*, dont l'un des avantages incontestables est l'objectivité que l'on vise dans une démarche scientifique. De plus, dès lors que les paramètres utilisés sont spécifiés, la procédure de repérage de

¹²⁰ En outre, plusieurs logiciels offrent cette fonctionnalité, comme AntConc (Anthony, 2018) ou Wordsmith Tools (Scott, 2016), ce qui facilite sa mise en œuvre (Bestgen, 2017 : 36).

keywords est tout à fait reproductible et les *keywords* ainsi obtenus constituent un point de départ solide, qui permet de cibler rapidement les unités et/ou les contextes qu'il faut analyser plus en détail par la suite (Bassi, 2010 : 299*sqq.* ; Scott, 2010 : 55-56 ; Culpeper et Demmen, 2015 : 99 ; Gabrielatos, 2018 : 227-228). En ce sens, les *keywords* représentent un point d'entrée dans les données, complémentaire aux termes eux-mêmes (cf. Chapitre 3, 3.3.1).

Ce premier indice de type distributionnel tient donc compte de contextes de type « fenêtre de mots », que nous abordons par le « détournement » d'une méthode d'analyse spécifique de la linguistique de corpus, le repérage de *keywords*. Les différents paramètres concernant la mise en œuvre de cet indice dans le corpus sont détaillés dans la section 5.3.1.

5.2.2.2. Affiner l'analyse sémantique des termes : prise en compte des dépendances syntaxiques

Le second indice que nous mettons en œuvre repose sur une analyse des contextes syntaxiques des termes à l'étude. En analyse distributionnelle, lorsque le contexte distributionnel est de ce type, l'analyse se focalise uniquement sur les cooccurrents qui sont liés à l'unité étudiée par une relation de dépendance syntaxique (Bourigault, 2002 ; Van de Cruys, 2008 ; Morlane-Hondère, 2013 ; Fabre et Lenci, 2015 ; Tanguy *et al.*, 2015 ; Lenci, 2018). Les contextes se présentent généralement sous une forme « binaire », où l'unité étudiée est reliée à chaque unité avec laquelle elle entretient une relation de dépendance, c'est-à-dire où chaque gouverneur est relié à ses dépendants (Habert et Fabre, 1998 : 114), par exemple *particule-élémentaire*, *particule-chargé*, *particule-mystérieux* dans le cas des modificateurs du terme *particule* (cf. Chapitre 7). Dans certains cas, la relation entre les unités est précisée dans la représentation et l'on parle de *triplets*, par exemple <particule ; *mod* ; élémentaire> pour reprendre la notation de Tanguy *et al.* (2015).

L'utilisation d'un corpus analysé syntaxiquement est donc requise et un analyseur syntaxique est généralement exploité dans ce but. Cependant, malgré l'existence aujourd'hui de plusieurs analyseurs syntaxiques robustes permettant le traitement de gros corpus (Tanguy *et al.*, 2019 : 51-53), l'apport effectif d'une analyse distributionnelle basée sur des contextes syntaxiques par rapport aux fenêtres de mots fait débat. Si certains auteurs démontrent en effet de meilleures performances à partir de contextes syntaxiques (par exemple Peirsman *et al.*, 2007 ; Van de Cruys, 2008), d'autres, en revanche, montrent que des « modèles par cooccurrence graphique peuvent rivaliser avec les modèles syntaxiques », pour autant qu'ils soient paramétrés avec précision (Tanguy *et al.*, 2015 : 125). Zweigenbaum et Habert (2004 : 102) ajoutent que, dès lors que le nombre de données est suffisamment grand, les méthodes basées sur des contextes à fenêtres de mots démontrent des performances aussi élevées que les méthodes avec contextes syntaxiques, tout en

étant plus rapides, du fait qu'une moindre puissance de calcul est requise. Enfin, selon Lenci (2018 : 154), « the question of whether syntactic information provides a real advantage over bag-of-words representations of contexts is still open and highly dependent on the semantic task ».

C'est ce dernier élément que nous retenons ici. En effet, les deux types de contextes présentent certains avantages et, dans les deux cas, des regroupements pertinents, bien que différents, sont effectués. C'est précisément la nature différente des résultats obtenus avec des contextes syntaxiques par rapport à des fenêtres de mots qui constitue selon nous un apport intéressant pour notre travail. De ce point de vue, l'analyse des contextes syntaxiques vient compléter les aspects relevés par l'analyse de *keywords*, permettant ainsi de dresser un panorama aussi fidèle que possible des manifestations du processus de déterminologisation observables en corpus. En ce sens, la prise en compte des dépendances syntaxiques présente au moins deux avantages.

En premier lieu, celle-ci permet de systématiser l'analyse des termes. En effet, de par l'automatisation de l'extraction des dépendances syntaxiques des termes dans le corpus, nous nous assurons de bien tenir compte de tous les contextes d'apparition des termes dans l'analyse. Il ne s'agit donc pas ici d'automatiser tout le processus d'analyse et de regroupement d'unités dont la distribution est similaire, mais uniquement d'automatiser l'étape de repérage des unités qui nous intéressent, c'est-à-dire celles qui entretiennent une relation de dépendance avec les termes à l'étude. En effet, dès lors que le corpus est analysé syntaxiquement, et quelle que soit sa taille, il est possible d'extraire automatiquement toutes les dépendances d'un terme. Ainsi, l'automatisation d'une partie du processus permet d'obtenir de manière objective, systématique et reproductible les données nécessaires à l'analyse des dépendances syntaxiques, ce qui représente une base plus fiable qu'un repérage ou une annotation manuelle, par exemple.

En second lieu, avec l'analyse des dépendances syntaxiques des termes, nous pouvons tenir compte d'un contexte qui entretient un lien plus direct avec les termes à l'étude, puisque c'est précisément la relation de dépendance qui constitue le critère de sélection du contexte. Par conséquent, nous nous assurons de la pertinence des unités retenues pour l'analyse vis-à-vis des termes, contrairement aux *keywords*, où cet aspect n'est pas aussi certain du fait que le contexte d'analyse comprend toutes les unités qui apparaissent dans une certaine fenêtre autour des termes. Ces avantages, par rapport à d'autres méthodes d'interrogation de corpus, sont par exemple mis en avant par Habert et Fabre :

« [L]es dépendances binaires offrent une représentation épurée mais précise des contextes d'emploi d'un vocable. À la différence des approches de la cooccurrence basées sur des fenêtres graphiques, elles mettent un vocable en

relation uniquement avec les mots avec lesquels il entretient une relation de dépendance syntaxique effective. Plus synthétiques que les concordances, elles fournissent comme elles des environnements éclairants et rarement ambigus. » (Habert et Fabre, 1998 : 122)

De ce point de vue, nous pouvons nous focaliser de manière plus précise et détaillée sur les fonctionnements des termes dans le corpus. En effet, étant donné que la prise en compte des dépendances uniquement implique l'analyse d'un contexte restreint, les phénomènes sémantiques en jeu dans le processus de déterminologisation peuvent être appréhendés plus finement. Ainsi, et sur la base de travaux en analyse distributionnelle qui s'appuient sur la comparaison de corpus (tels que Aussenac-Gilles *et al.*, 2003 ; Habert *et al.*, 2005 ; Halskov, 2005 ; Del Tredici *et al.*, 2019 ; Fišer et Ljubešić, 2019), nous faisons l'hypothèse que des différences dans les dépendances des termes dans les sous-corpus peuvent indiquer des fonctionnements particuliers des termes, ou même des changements sémantiques, que l'on peut mettre en lien avec la déterminologisation. En ce sens, nous nous situons dans une perspective similaire de celle de Halskov (2005), qui cherche à observer des évolutions significatives dans les collocations de termes informatiques dans un corpus de presse et qui les compare à un corpus spécialisé (Halskov, 2005 : 42-44, cf. Chapitre 2, 2.2.2).

Notre but consiste alors à repérer ces différences de distribution et à les caractériser par rapport à ce qu'elles peuvent nous apprendre quant au fonctionnement des termes en situation de déterminologisation. Nous nous attendons ainsi à observer différents types de fonctionnements, qui sont à rapprocher de changements sémantiques de type métaphorique ou métonymique décrits notamment par Meyer et Mackintosh (2000a, 2000b) et Renouf (2017) ou de phénomènes peut-être plus subtils. Dans le premier cas, à l'instar de Renouf (2017), par exemple, nous nous attendons à observer l'usage de modificateurs différents dans le SC-P par rapport au SC-S, qui peuvent pointer vers un usage métaphorique des termes. L'autrice remarque en effet l'utilisation des adjectifs *cultural* et *emotional* avec le terme *bulimia* dans des textes de presse, qu'elle interprète comme des usages métaphoriques de ce terme (Renouf, 2017 : 39-40). C'est donc par une observation fine des différentes unités partageant une relation de dépendance avec les termes à l'étude, dans les différents sous-corpus, que nous serons en mesure de mettre au jour des différences de fonctionnement attribuables au processus de déterminologisation.

Ces deux indices reposent en grande partie sur la comparaison des différents sous-corpus. Mais cette approche contrastive pose différentes questions d'ordre pratique et technique notamment, vu l'organisation des données en cinq sous-corpus. Cette particularité de notre approche impose alors de définir une méthode qui rende possibles des comparaisons entre ces

sous-corpus et qui permette d'observer des différences interprétables. Aussi, dans la section suivante, nous proposons de procéder en deux temps.

5.2.3. Définition d'une stratégie de comparaison en deux temps

L'observation de la distribution des termes dans les sous-corpus se fonde sur la comparaison des sous-corpus ; nous cherchons ainsi à faire émerger des différences dans les environnements distributionnels des termes, que nous interprétons ensuite en termes de différences de sens en rapport avec le processus de déterminologisation. Cependant, la comparaison simultanée de cinq sous-corpus reste peu envisageable, même assistée par des outils. Il n'est pas sûr, en effet, que des différences de distribution significatives soient même repérables en tenant compte d'autant de points de comparaison. Par ailleurs, si, techniquement, les outils d'exploration de corpus ne semblent pas permettre de telles comparaisons, il n'est pas non plus certain que, cognitivement, la prise en compte « simultanée » de tous les sous-corpus dans l'analyse rende possible l'interprétation des phénomènes observés.

Ces remarques renvoient notamment à la rareté des démarches où autant de sous-corpus sont pris en compte dans l'analyse, du moins à notre connaissance, et sauf peut-être dans une perspective diachronique. Il n'est pas inhabituel en effet, en linguistique diachronique, de travailler à partir de nombreux sous-corpus, s'étalant parfois sur plusieurs centaines d'années (Gries et Hilpert, 2008 ; Boleda, 2020). Mais dans ces cas, c'est précisément l'hypothèse d'une évolution observable dans des tranches successives qui est interrogée. L'exploration des sous-corpus dans un ordre qui diffère de l'ordre chronologique n'aurait alors aucun sens si l'on souhaite mettre au jour des évolutions dans le temps (Gries et Hilpert, 2008 : 63). Or, dans le cas de la déterminologisation telle que nous l'observons, il n'est pas sûr que tous les sous-corpus doivent être interrogés successivement. Au contraire, bien que le processus de déterminologisation fasse intervenir différents genres et différents degrés de spécialisation, nous ne savons pas, à ce stade, dans quel ordre interviennent les sous-corpus intermédiaires dans le processus, ni même si un ordre peut réellement être établi. Certes, il semble évident d'opposer les SC-S et SC-P, puisque nous les considérons comme les deux « pôles » du processus en corpus, mais il semble également difficile de formuler une hypothèse quant à l'ordre dans lequel les sous-corpus intermédiaires (SC-C, SC-R et SC-V) doivent être interrogés.

Dans ce contexte, il apparaît donc nécessaire de définir une stratégie de comparaison qui permette non seulement de mettre en évidence des différences dans la distribution des termes qui restent interprétables d'un point de vue aussi bien technique que cognitif, c'est-à-dire pour lesquels

nous serons en mesure de proposer des interprétations plausibles à partir des éléments tirés du corpus et en consultant les expertes lorsque cela est adéquat, mais également d'aborder les différents sous-corpus d'une manière qui soit logique et pertinente si l'on souhaite interroger la dimension continue de la déterminologisation entre langue de spécialité et langue générale. Ainsi, même si l'organisation des données en cinq sous-corpus a été pensée pour tenir compte de cette dimension, ce choix reste peu opérationnel dès lors qu'il faut analyser tous ces sous-corpus. Pour pallier cette difficulté, nous proposons de procéder par étape et de mettre au point un schéma de comparaison qui permette de mettre au jour des éléments pertinents pour l'analyse. C'est pourquoi nous procédons à une analyse en deux temps. Dans un premier temps, nous comparons d'abord les SC-S et SC-P et, dans un second temps, nous tenons compte des sous-corpus intermédiaires.

Plus précisément, dans une première étape, nous nous concentrons sur les deux pôles du processus de déterminologisation tels que nous pouvons les observer dans le corpus, à l'instar des auteurs qui ont abordé la déterminologisation d'un point de vue contrastif en corpus (Halskov, 2005 ; Dury, 2008 ; Condamines et Picton, 2014b). Notre but étant d'observer des changements sémantiques dans la continuité entre le spécialisé et le non spécialisé, une étape passe nécessairement par le repérage des différences distributionnelles les plus significatives dans le SC-P par rapport au SC-S. Il s'agit donc de caractériser le fonctionnement des termes dans le SC-P, tout en insistant particulièrement sur les aspects qui diffèrent du SC-S.

Dans une seconde étape, nous nous basons sur ces différences pour interroger la dimension continue de la déterminologisation à travers différents degrés de spécialisation. Dans ce but, nous tenons compte des sous-corpus *Communiqués* (SC-C), *Rapports* (SC-R) et *Vulgarisation* (SC-V) et observons dans quelle mesure les usages qui semblent caractéristiques du SC-P sont réellement caractéristiques du SC-P. En d'autres termes, nous cherchons à vérifier si certains des usages, ou même tous, sont également attestés dans un ou plusieurs des sous-corpus intermédiaires. À ce stade, nous faisons l'hypothèse qu'une certaine part des fonctionnements caractéristiques du SC-P par rapport au SC-S sont en réalité également présents dans les sous-corpus intermédiaires, ce qui permet de formuler une hypothèse de changement progressif du sens des termes dans les degrés de spécialisation.

Dans le même temps, nous cherchons également à observer dans quelle mesure les situations de communication représentées dans ces trois sous-corpus jouent un rôle dans les usages des termes qui sont attestés dans le SC-P par rapport au SC-S, à partir de l'hypothèse que les termes sont progressivement transférés dans le lexique général à travers ces intermédiaires. De même, nous faisons l'hypothèse que les changements de sens les plus souvent attribués au processus de déterminologisation (tels que ceux décrits par Meyer et Mackintosh (2000a, 2000b), par exemple)

sont également progressifs dans ces textes et qu'ils sont donc observables dans les sous-corpus intermédiaires. Par ailleurs, nous visons non seulement à observer le rôle des intermédiaires dans la déterminologisation, mais également à déterminer si certains de ces intermédiaires sont plus susceptibles de jouer un rôle que d'autres. Par exemple, nous faisons l'hypothèse à ce stade que les communiqués de presse influencent particulièrement la manière dont sont utilisés les termes dans la presse, du fait que ces textes sont expressément destinés aux journalistes (cf. Chapitre 4, 4.2.2.1.1).

Cette stratégie de comparaison de la distribution des termes en deux temps nous permet donc non seulement de produire des résultats qui restent interprétables, puisqu'ils se basent sur l'observation d'un nombre plus restreint de sous-corpus à la fois, mais également d'aborder cette dimension continue de la déterminologisation d'une manière qui soit peut-être plus « gérable » cognitivement que la prise en compte des cinq sous-corpus à la fois. Au terme de l'analyse, nous serons donc en mesure de véritablement construire une interprétation des phénomènes en lien avec la déterminologisation dans la continuité entre le SC-S et le SC-P. Cette stratégie est appliquée dans les deux méthodes d'analyse distributionnelle que nous mettons en œuvre. Nous les détaillons dans la section suivante.

5.3. Mise en œuvre des indices

La mise en œuvre en corpus des deux indices de type distributionnel définis pour notre étude requiert de déterminer certains paramètres, relatifs notamment à la taille des contextes à récupérer, au score de *keyness* à utiliser et aux relations de dépendance à cibler. Nous discutons ces paramètres dans les deux sections qui suivent, d'abord pour le repérage de *keywords*, puis pour l'extraction des unités qui entretiennent une relation de dépendance avec les termes ciblés.

5.3.1. Méthodologie pour le repérage de *keywords*

Différents choix méthodologiques interviennent afin de repérer automatiquement des *keywords*, à partir du point de vue particulier que nous adoptons. Par exemple, puisque l'on sait qu'en analyse distributionnelle, la taille des contextes peut influencer la nature et la qualité des résultats (par exemple Fišer et Ljubešić, 2019 : 166), il convient de déterminer une taille pertinente pour notre but, à partir de différents critères. De la même manière, il est nécessaire d'établir le calcul utilisé pour repérer les *keywords*. Aucun calcul ne fait l'unanimité dans les travaux de linguistique de corpus qui se fondent sur une analyse de *keywords* et différentes positions sont

exprimées. Il convient alors d'examiner ces positions afin de prendre une décision éclairée et de produire des résultats pertinents.

5.3.1.1. Extraction des contextes distributionnels des termes

Comme nous le disons *supra*, nous procédons à l'extraction des *keywords* dans un sous-ensemble du SC-P par rapport à un sous-ensemble du SC-S, lesquels sont constitués des contextes distributionnels des termes dans ces deux sous-corpus. Soulignons toutefois que tous les termes retenus à la suite de l'extraction semi-automatique effectuée avec TermoStat (cf. Chapitre 4, 4.3) ne font pas partie de l'échantillon utilisé ici. De fait, certains de ces termes apparaissent avec une fréquence très réduite dans l'un ou l'autre de ces sous-corpus, parfois quelques occurrences seulement. Avec une fréquence aussi réduite, les contextes de ces termes seraient également fortement réduits, ce qui ne nous semble pas pertinent dans le but de mettre au jour des régularités distributionnelles. C'est pourquoi, afin d'assurer que l'analyse se base sur des données suffisamment nombreuses, un seuil de fréquence d'occurrence des termes dans les SC-P et SC-S a été fixé à 10¹²¹. Avec cette restriction, l'échantillon à analyser se compose de 52 termes (la liste peut être consultée dans l'annexe B).

La mise en œuvre de cet indice suppose différents choix, à savoir la taille des contextes et la manière de les récupérer. Précisons cependant que la taille des contextes n'est généralement pas une question que se posent les chercheurs travaillant à partir d'analyses de *keywords*, du fait que cette méthode est le plus souvent exploitée pour mettre en évidence les unités les plus spécifiques d'un corpus par rapport à un autre corpus (cf. 5.2.2.1). Dans ce cas, comme ce sont les corpus en entier qui sont confrontés, il n'y a pas lieu de définir une taille de contexte particulière autour d'une unité ou d'un ensemble d'unités. Mais l'utilisation particulière que nous faisons de cette méthode, afin d'accompagner notre analyse de la distribution des termes dans le corpus, rend nécessaire la définition d'une taille de contexte à exploiter autour des termes à l'étude.

En analyse distributionnelle, cette question est d'ailleurs récurrente et revêt une certaine importance (Sahlgren, 2008 : 11 ; Fabre et Lenci, 2015 : 10). Ainsi, les approches tenant compte de fenêtres de mots définissent le contexte de deux manières (Van de Cruys, 2008 : 47 ; Morlane-Hondère, 2013 : 42-43) : il s'agit soit d'une unité textuelle, comme la phrase dans le cas de Bannour *et al.* (2011) par exemple, soit d'un certain nombre de mots graphiques avant et/ou après l'unité

¹²¹ Précisons qu'il n'existe pas de fréquence minimum idéale qui fournirait avec certitude des données en nombre suffisant pour décrire adéquatement les usages d'une unité (Kennedy, 1998 : 67-68). Il semble ainsi généralement admis qu'une fréquence trop faible est insuffisante (Hunston, 2008 : 156), mais il n'existe aucun seuil « idéal ». Notre choix de le fixer à 10 comporte donc nécessairement une part d'arbitrarité, dont nous sommes consciente.

d'intérêt. Cette seconde option est la plus souvent utilisée, selon Sahlgren (2008 : 10) et Lenci (2018 : 154), pour des raisons évidentes de simplicité à mettre en œuvre. Dans notre cas, nous estimons également qu'un nombre fixe de mots graphiques à récupérer garantit qu'un volume de données comparable soit extrait pour chaque occurrence de chaque terme.

La question de la taille des contextes semble plus délicate, car il n'existe pas de taille idéale et les travaux ne sont pas consensuels sur ce point. La taille peut en effet varier dans une très large mesure, allant de quelques mots graphiques à plusieurs centaines (Morlane-Hondère, 2013 : 43 ; Fabre et Lenci, 2015 : 10). Cela dit, même si, comme l'affirme Sahlgren (2008 : 11), « [t]here is very seldom a theoretical motivation for a particular window size », la taille du contexte n'est pas pour autant fixée de manière totalement arbitraire. Celle-ci doit en effet être déterminée en fonction de l'objectif de recherche et du volume des données à disposition (Tanguy *et al.*, 2015 : 108).

Pour ce qui est du volume des données, les auteurs rapportent que « la meilleure combinaison globale [est] une petite fenêtre avec un grand corpus » (*Ibid.*), ce qui reviendrait à considérer essentiellement des collocations (Sahlgren, 2008 : 8-11). Tanguy *et al.* tiennent donc compte de contextes composés d'une, de trois et de cinq unités autour des unités à l'étude. De ce point de vue, du fait que les sous-corpus dont nous extrayons les contextes sont de taille modeste (de 200 000 à 1 million d'occurrences environ), ils semblent peu adaptés à l'utilisation de petits contextes. Par ailleurs, nous estimons qu'une taille de contexte plus grande peut mieux correspondre à notre objectif de recherche et à la méthode mise en œuvre. En effet, selon Morlane-Hondère « [u]ne fenêtre plus étendue a tendance à fournir des contextes plus lâches qui sont liés au mot cible par une association de type thématique » (Morlane-Hondère, 2013 : 43). L'extraction de *keywords* étant une méthode qui a généralement tendance à mettre en évidence les thèmes spécifiques du corpus d'analyse par rapport au corpus de référence (cf. 5.2.2.1), alors il semble judicieux dans notre cas d'exploiter un contexte dont la taille plus large permet de repérer des « association[s] de type thématique ».

Dans cette perspective, nous faisons le choix de récupérer des contextes composés de 20 mots graphiques avant et après le terme. Cette taille de contexte constitue ainsi un compromis entre un contexte suffisamment grand, qui privilégie le repérage d'associations thématiques, et un contexte suffisamment réduit, qui maximise les probabilités que les *keywords* extraits soient effectivement en lien avec les termes analysés, car plus le nombre de mots graphiques est élevé, plus il est possible que les unités les plus éloignées du terme à l'étude ne soient pas pertinentes.

Enfin, pour obtenir des *keywords* de meilleure qualité, nous travaillons à partir des versions étiquetées et lemmatisées avec TreeTagger (Schmid, 1994) des SC-S et SC-P, nous inspirant ainsi de Bannour *et al.* (2011) et de Tanguy *et al.* (2015). L'extraction des contextes à proprement parler a été effectuée avec AntConc à l'aide d'expressions régulières. Ces contextes constituent des sous-ensembles des SC-S et SC-P (désormais *SC-S-KW* et *SC-P-KW*).

5.3.1.2. Établir le score de *keyness* : enjeu pour l'identification de *keywords* pertinents

Les *keywords* sont identifiés à l'aide d'un score de *keyness* : plus ce score est élevé, plus une unité est considérée comme *key*. Ce score peut être établi à partir de différents calculs, qui se basent tous au moins sur la fréquence des unités dans les échantillons comparés. Plusieurs facteurs entrent en jeu pour définir le calcul à utiliser, nous les discutons dans cette section.

Premièrement, différentes statistiques sont exploitées dans le but de mesurer la significativité des différences de fréquence observées entre le corpus d'analyse et le corpus de référence. À l'origine, la significativité statistique est calculée avec le χ^2 (Scott, 1997 : 236), mais d'autres mesures sont également exploitées aujourd'hui, telles que le *log-likelihood* (Brezina, 2018 : 83-84). Ces deux calculs sont d'ailleurs les plus souvent utilisés d'après l'étude de Pojanapunya et Watson Todd (2018 : 144), probablement du fait qu'ils sont aisément exploitables de par leur implémentation dans WordSmith Tools, comme le suggère Gabrielatos (2018 : 230). Néanmoins, pour certains auteurs (Kilgarriff, 1996, 2005 ; Gries, 2006 ; Bestgen, 2017 ; Brezina, 2018 ; Egbert et Biber, 2019), ces calculs ne sont pas toujours pertinents pour identifier les *keywords*, car ils ne mesurent pas l'importance des différences de fréquence des unités dans le corpus d'analyse par rapport au corpus de référence, mais la significativité statistique de ces différences. Cet aspect est problématique selon Gabrielatos (2018 : 230), car

« statistical significance does not reveal the size of a frequency difference, but, indirectly, the level of confidence we can have that the difference we have observed (however large or small) is dependable. »

Aussi, parmi les cas où l'utilisation de la significativité statistique comme mesure de l'importance des différences de fréquence peut produire des résultats inadéquats, l'auteur donne un exemple où « [a] very large frequency difference may have very low statistical significance » (*Ibid.* : 232). Dans ce cas, étant donné qu'une faible significativité statistique engendre un faible score de *keyness* et que les unités avec les scores les plus faibles ne sont généralement pas conservées pour analyse, alors cela signifie que les unités correspondant à ce cas de figure sont ignorées, bien que les différences de fréquence constatées soient importantes. En outre, l'inverse est également vrai : une faible différence de fréquence peut être tout à fait significative d'un point de vue

statistique (*Ibid.*). Il est donc possible que certains *keywords* ne soient en réalité pas représentatifs des véritables particularités du corpus d'analyse par rapport au corpus de référence.

Prenons un exemple : le repérage des *keywords* dans SC-P-KW par rapport à SC-S-KW, que nous effectuons avec AntConc, affiche le verbe *avoir* en quatrième position, aussi bien avec le *log-likelihood* qu'avec le χ^2 , avec un score respectivement de 1499,23 et 1591,07 (les données présentées dans cet exemple sont résumées dans le tableau 5.1). Les scores ne sont pas les éléments les plus importants ici, mais bien la position à laquelle figure le verbe *avoir*. La quatrième position indique qu'il s'agit d'une unité particulièrement *key*, c'est-à-dire que la différence de fréquence de cette unité entre les deux sous-ensembles comparés est tout à fait significative d'un point de vue statistique, selon les calculs utilisés. Or, d'un point de vue linguistique, ce verbe figure parmi les plus fréquents dans la langue, tout comme les verbes *être*, *pouvoir* ou *faire* par exemple. Le statut de *keyword* que ce verbe obtient ici semble donc plutôt étrange à première vue.

En réalité, ce statut est à attribuer au fonctionnement des calculs du *log-likelihood* et du χ^2 . En effet, dans le cas du verbe *avoir*, les résultats de ces deux calculs ne signifient pas que la différence de fréquence est particulièrement importante, mais simplement que, d'un point de vue statistique, l'hypothèse nulle peut être rejetée en toute confiance et qu'il existe donc bien une différence de fréquence de l'unité entre les deux sous-ensembles, quelle que soit cette différence (Gabrielatos, 2018 : 230-231). Le problème soulevé par les auteurs que nous citons *supra* est justement que l'importance des différences de fréquence n'est pas toujours bien prise en compte par les calculs de ce type, bien qu'elle soit pourtant à l'origine de la définition de *keywords* (voir la définition de Scott (1997) en 5.2.2.1). Dans le cas du verbe *avoir*, en effet, la fréquence relative dans SC-P-KW n'est « que » le double que dans SC-S-KW, alors que d'autres unités qui obtiennent des scores plus faibles présentent des différences bien plus importantes proportionnellement (cf. tableau 5.1). C'est par exemple le cas de l'adjectif *immense*, qui figure à la 141^e position avec le *log-likelihood* et à la 153^e avec le χ^2 , avec des scores entre 8 et 9 fois plus faibles que le verbe *avoir*, tandis que les différences de fréquence entre SC-P-KW et SC-S-KW sont proportionnellement plus importantes. Cela signifie que, malgré une différence de fréquence relative plus importante de l'unité *immense* entre les deux échantillons que de l'unité *avoir*, l'adjectif *immense* est considéré comme moins *key* avec le *log-likelihood* et le χ^2 qu'*avoir*.

	<i>avoir</i>	<i>immense</i>
Fréquence absolue dans le sous-ensemble du SC-P	5031	87
Fréquence relative dans le sous-ensemble du SC-P	1414,9	24,5
Fréquence absolue dans le sous-ensemble du SC-S	4752	3
Fréquence relative dans le sous-ensemble du SC-S	644,4	0,4
Score obtenu avec le <i>log-likelihood</i>	1499,23	171,46
Position dans la liste	4	141
Score obtenu avec le χ^2	1591,07	168,66
Position dans la liste	4	153

Tableau 5.1 : Fréquences et scores de *keyness* pour les unités *avoir* et *immense*

Ces remarques renvoient en réalité à une limite des calculs du *log-likelihood* et du χ^2 notamment. Ces calculs sont en effet sensibles à la taille des corpus et aux fréquences des unités dans les corpus (Gabrielatos, 2018 : 233 ; Pojanapunya et Watson Todd, 2018 : 146 ; Egbert *et al.*, 2020 : 40). Comme l'explique Gabrielatos :

« given a frequency difference, the higher the raw frequencies of an item in the two corpora and/or the larger the two corpora, the higher the statistical significance value will be. » (Gabrielatos, 2018 : 233)

Ainsi, plus la fréquence des unités est élevée dans les corpus qui sont comparés, plus il est probable que l'hypothèse nulle soit rejetée, ce qui a tendance à survaloriser, dans les résultats, les unités très fréquentes, au détriment des unités moins fréquentes. C'est ce que nous montrons avec les exemples d'*avoir* et d'*immense*. Dans ce contexte, ce type de calcul n'est donc probablement pas le plus adéquat pour identifier les unités qui sont proportionnellement plus fréquentes dans SC-P-KW par rapport à SC-S-KW.

Deuxièmement, afin de pallier ces biais possibles et de mieux tenir compte des différences de fréquence observées entre le corpus d'analyse et le corpus de référence, d'autres calculs sont possibles, qui n'ont pas une base probabiliste comme le χ^2 et le *log-likelihood*. Ceux-ci se fondent sur l'ampleur de l'effet (*effect size*). Comme l'affirment Pojanapunya et Watson Todd (2018 : 146), « [a]n alternative set of statistics, effect size statistics, focuses on how large the difference between the two frequencies of a word is ». Ce type de calcul permet donc de mettre en évidence non pas les unités dont les différences de fréquences entre les corpus d'analyse et de référence sont les plus significatives d'un point de vue statistique (comme dans l'exemple d'*avoir*), mais bien celles dont les différences sont les plus grandes, proportionnellement (comme dans l'exemple d'*immense*). Le plus souvent, les calculs de ce type utilisent les fréquences relatives des unités dans les corpus et non les fréquences brutes.

Les points de vue à propos de l'utilisation de l'un ou de l'autre de ces types de calculs ne sont pas consensuels. Certains auteurs, et notamment Gabrielatos, adoptent une position tout à fait tranchée. Pour lui, en effet, c'est précisément l'importance des différences de fréquence d'une unité entre deux corpus qui détermine la *keyness* de cette unité, et non la significativité statistique des différences. L'auteur argumente que cette conception de *keyword* est celle qui se rapproche le plus de la définition originale proposée par Scott :

« [v]ery early on, keywords were defined as 'words whose frequency is unusually high in comparison with some norm' (Scott 1996 : 53). It is straightforward to derive from this definition that a keyword is identified by way of a frequency comparison. It should clearly follow, then, that an appropriate metric for keyness would reflect the size of the frequency difference, and that the larger the difference, the more 'key' a word would be. » (Gabrielatos, 2018 : 229)

D'autres auteurs, en revanche, sont plus nuancés. Par exemple, Pojanapunya et Watson Todd (2018) reconnaissent certaines limites du *log-likelihood*, notamment l'influence de la taille des corpus sur les valeurs retournées par ce calcul, mais argumentent sur le fait que les résultats obtenus avec les deux types de calculs sont intéressants dans les deux cas, bien qu'ils soient de nature différente. Selon leur comparaison (2018 : 160*sq.*), le *log-likelihood* serait plus utile pour identifier les caractéristiques d'un genre textuel en particulier, alors que le calcul du *odds ratio* (un calcul de l'ampleur de l'effet) rendrait plutôt compte des thèmes spécifiques du corpus d'analyse. Les auteurs ajoutent par ailleurs que dans le premier cas, il est plus probable d'observer des unités plutôt répandues dans la langue et qui figurent parmi les plus fréquentes, tandis que, dans le second cas, les unités repérées sont moins communes et moins fréquentes dans les corpus analysés.

De la même manière, bien qu'il soit également critique quant à l'application d'un calcul probabiliste, Kilgarriff (2005 : 264) insiste surtout sur la nécessité de considérer les résultats avec une certaine prudence, afin de ne pas tirer de conclusions trop rapides, voire erronées ou surinterprétées. Cette mise en garde rejoint d'ailleurs l'utilisation des *keywords* qui est faite dans la plupart des travaux exploitant cette méthode. En effet, comme nous le disons *supra*, les *keywords* représentent le plus souvent un point de départ, une première manière d'entrer dans les données, suivie par la prise en compte des contextes du corpus dans une perspective plus qualitative.

Troisièmement, insistons sur le fait que le choix d'un calcul dépend fondamentalement de l'objectif de chaque recherche, comme le soulignent notamment Culpeper et Demmen (2015 : 99). Dans notre cas, nous nous concentrons sur les différences de fréquence les plus importantes entre nos deux échantillons, mais nous ne visons pas exclusivement les unités les plus fréquentes. À ce stade, dans une démarche essentiellement exploratoire, même les événements les plus isolés sont dignes d'intérêt. Parallèlement, il semble également intéressant d'obtenir une information sur la

significativité de ces différences. C'est pourquoi, suivant la proposition de Pojanapunya et Watson Todd (2018 : 161), nous combinons deux calculs et ne conservons que les *keywords* dont la différence de fréquence est la plus importante entre les deux échantillons et est également statistiquement significative.

Plus précisément, nous faisons le choix d'utiliser le *relative frequency ratio* (ratio des fréquences relatives, abrégé *RFR*) (Damerou, 1993) comme mesure de l'ampleur de l'effet, qui est une des possibilités énumérées par Pojanapunya et Watson Todd (2018) et Gries (2015, 2016) notamment¹²², ainsi que le *log-likelihood* comme mesure de la significativité statistique. Ce dernier choix s'explique par le fait que selon Culpeper et Demmen (2015 : 98) et Rayson *et al.* (1997 : 148-149), ce calcul permet d'obtenir globalement de meilleurs résultats que le χ^2 , en particulier lorsque les données et les fréquences sont limitées. En outre, ces calculs sont implémentés dans AntConc, que nous utilisons pour extraire les *keywords*.

La comparaison des deux échantillons avec AntConc produit une liste de 9 863 *keywords*. Tous ne sont évidemment pas à considérer comme *key*. Les valeurs obtenues avec le RFR varient entre 904,2034 et 1,0003. Plus la valeur est élevée, plus l'unité correspondante est considérée comme *key* (Gries, 2015 : 55). Aussi, une valeur proche de 1 indique que la fréquence relative de l'unité concernée est très proche dans les deux échantillons. Les unités pour lesquelles la valeur de l'ampleur de l'effet est faible ne sont donc pas à prendre en compte en priorité. Cependant, il reste à déterminer le nombre de *keyword* à prendre en compte pour la suite de l'analyse.

De manière générale, seul un certain nombre de *keywords* apparaissant comme les plus *key*, c'est-à-dire avec les valeurs les plus élevées, sont conservés pour une analyse approfondie (Gabrielatos, 2018 : 238 ; Egbert et Biber, 2019 : 80). Néanmoins, dans les travaux synthétisés par Pojanapunya et Watson Todd (2018 : 135*sqq.*), le nombre de *keywords* retenus semble le plus souvent arbitraire, les *n keywords* obtenant les scores les plus élevés étant généralement sélectionnés. Ce nombre varie entre 10 et 1 000. Les auteurs décrivent également une autre option, qui consiste à retenir tous les *keywords* dont le score est supérieur à un certain seuil de significativité statistique, par exemple $p < 0,05$ ou $p < 0,001$ (*Ibid.*). Dans notre cas, cette option semble peu envisageable, car cela représente encore plus de 5 000 *keywords* dans le premier cas et plus de 2 300 dans le second cas, ce qui reste conséquent dans une approche qualitative. Par ailleurs, ces seuils de significativité

¹²² Vu le nombre de calculs de l'ampleur de l'effet existants (voir par exemple Kilgarriff, 2001 ; Gries, 2016 ; Gabrielatos, 2018 ; Pojanapunya et Watson Todd, 2018) et vu qu'ils sont tous appropriés pour le repérage de *keywords* selon ces auteurs, nous sommes consciente que le choix du RFR comporte une certaine part d'arbitrarité, mais comme le souligne Brezina (2018 : 85) à ce propos, « the question of which statistic best suits the identification of keywords is an open one. »

ne garantissent pas que la différence de fréquence relative calculée avec le RFR soit réellement importante. Pour ces raisons, et pour que le nombre de données reste raisonnable dans une perspective d'analyse qualitative, nous faisons le choix de retenir les 500 *keywords* dont la différence de fréquence est la plus grande, c'est-à-dire où la valeur de l'ampleur de l'effet est la plus élevée, et pour lesquels le score obtenu avec le *log-likelihood* est significatif¹²³.

Ces 500 *keywords* constituent alors la base d'une analyse plus fine, d'abord des *keywords* en eux-mêmes, puis de leurs contextes, dans les SC-S et SC-P et dans les sous-corpus intermédiaires. Le Chapitre 6 est dédié à l'exploration de cet indice en corpus. Dans la section suivante en revanche, nous détaillons les paramètres déterminés pour la mise en œuvre du second indice de type distributionnel.

5.3.2. Méthodologie pour l'analyse des contextes syntaxiques

L'analyse des contextes syntaxiques des termes requiert en amont différentes étapes de préparation et d'extraction des données. Aussi, dans cette section, nous précisons d'abord les contraintes de sélection de l'échantillon de termes à analyser, sur la base de leur fréquence dans les cinq sous-corpus, puis nous spécifions plusieurs aspects relatifs à l'extraction des contextes, tels que les relations de dépendance sur lesquelles nous focaliser en priorité, le filtrage (ou non) de certaines catégories grammaticales, l'application (ou non) d'un filtre de fréquence, le traitement des résultats bruts afin de les rendre comparables entre les sous-corpus.

5.3.2.1. Contrainte de fréquence sur le choix des termes à analyser

Contrairement à l'analyse de *keywords*, où toutes les unités apparaissant dans la fenêtre spécifiée autour du terme sont incluses dans l'analyse, dans le cas des contextes syntaxiques, seules les unités qui partagent une relation de dépendance syntaxique avec les termes sont prises en compte. La taille des contextes syntaxiques est donc par définition plus petite que la taille des contextes à fenêtres de mots, comme le souligne d'ailleurs Morlane-Hondère (2013 : 44) lorsqu'il

¹²³ Si nous reprenons les exemples d'*avoir* et d'*immense* que nous avons présentés *supra*, nous remarquons qu'avec ce choix de ne conserver que les 500 premiers résultats, *immense* aurait tout de même été sélectionné, aussi bien avec le *log-likelihood* qu'avec le χ^2 , puisque cette unité figure à la 141^e position dans le premier cas et à la 153^e dans le second cas. Néanmoins, d'autres unités pertinentes auraient été ignorées, telles que *fantastique*, *cancer* ou *mythique*, par exemple (nous y reviendrons dans le Chapitre 7). Mais le choix de trier les résultats par RFR permet surtout d'éviter de conserver des unités qui seraient statistiquement significatives uniquement du fait de leur fréquence particulièrement élevée, comme dans le cas d'*avoir*, et qui seraient alors des « faux positifs » pour l'objectif que nous poursuivons. Ce choix nous permet donc de nous concentrer sur les unités qui sont *a priori* les plus pertinentes.

compare les modèles d'analyse distributionnelle automatique à partir de ces deux types de contextes. En outre, selon Habert et Zweigenbaum (2003),

« [l]es contextes syntaxiques sont "maigres" et diminuent donc les éléments de rapprochement entre mots. Seuls les mots très fréquents entrent dans suffisamment de contextes pour que cet étalage ne soit pas fatal. »

De ce point de vue, et afin d'obtenir suffisamment de données, il semble qu'il soit nécessaire de privilégier les termes les plus fréquents et donc de spécifier une fréquence minimum d'apparition des termes dans le corpus plus élevée que pour le repérage de *keywords*.

Ce choix contraint fortement la sélection des termes à analyser pour au moins deux raisons. D'abord, il semble évident que le choix d'un seuil de fréquence plus élevé aura pour conséquence un nombre plus faible de termes à considérer. Ensuite, contrairement à l'indice fondé sur une extraction de *keywords*, ce seuil de fréquence doit être respecté dans les cinq sous-corpus, et non uniquement dans les SC-P et SC-S. Ainsi, du fait que le contenu des textes peut différer dans une certaine mesure selon les sous-corpus, même s'ils portent tous sur la physique des particules et qu'ils ont été sélectionnés à partir de critères développés pour garantir la cohérence du corpus (cf. 4.2), il est peu probable que les termes retenus pour notre travail de manière générale soient attestés dans les mêmes proportions dans chacun des sous-corpus.

Prenons un exemple afin d'illustrer notre propos. Les termes *collisionneur* et *hadron* apparaissent respectivement à 1 008 et 1 027 reprises dans le corpus en entier. Ils semblent donc être deux candidats tout aussi intéressants l'un que l'autre pour notre analyse des dépendances. Néanmoins, la répartition de ces termes dans les sous-corpus montre que la fréquence absolue du terme *hadron* est plus faible dans les SC-C et SC-P que celle du terme *collisionneur*. Le terme *hadron* n'y apparaît en effet que 13 et 14 fois respectivement, contre 69 et 108 fois pour *collisionneur* dans ces mêmes sous-corpus. Aussi, il n'est pas sûr qu'avec de telles fréquences, l'extraction des dépendances du terme *hadron* dans les SC-C et SC-P permette de récupérer suffisamment de données pour notre analyse.

Parallèlement, bien qu'il n'existe pas non plus de solution idéale ici et que notre choix comporte nécessairement une certaine part d'arbitrarité, il est aussi – et surtout – guidé par la nécessité d'effectuer une analyse à partir d'un nombre de termes qui semble suffisant. De manière générale, en effet, notre but n'est pas de caractériser le fonctionnement d'un seul terme en lien avec la déterminologisation, mais d'observer dans quelle mesure il est possible de mettre au jour des fonctionnements réguliers, qui peuvent par la suite conduire à une meilleure compréhension des manifestations du processus de déterminologisation dans les textes. Le seuil de fréquence doit donc

garantir que non seulement des contextes syntaxiques suffisamment nombreux, pour chaque terme et dans chaque sous-corpus, sont récupérés, mais aussi que ces informations sont récupérées pour un nombre suffisant de termes.

De la même manière, il est également nécessaire de limiter autant que possible les cas où nous ne pourrions pas déterminer si l'absence d'une unité dans les contextes d'un terme dans un sous-corpus est attribuable à un fonctionnement particulier du terme dans ce sous-corpus ou à un manque de données. Au vu de ces différents éléments, nous faisons donc le choix de conserver uniquement les termes qui apparaissent à 50 reprises au moins dans chacun des sous-corpus. Nous pensons que ce seuil, plutôt élevé si l'on considère la taille et l'hétérogénéité des sous-corpus, permet de limiter les biais mentionnés. L'échantillon ainsi défini est composé de 15 termes nominaux¹²⁴.

5.3.2.2. Repérage et extraction des contextes syntaxiques en corpus

Dans cette section, nous déterminons d'abord les relations de dépendance à considérer en priorité, puis, à partir du corpus analysé syntaxiquement en dépendances, nous expliquons la méthode de repérage mise en œuvre et le traitement appliqué aux données extraites pour permettre leur analyse quantitativement et qualitativement.

Le premier aspect concerne les types de relations de dépendance sur lesquelles nous concentrer. En effet, toutes les relations ne sont pas nécessairement prises en compte dans les travaux existants portant sur une analyse distributionnelle (automatique) à partir de contextes syntaxiques¹²⁵. Habert et Fabre (1998 : 114*sqq.*), par exemple, retiennent uniquement les relations entrant en jeu dans des syntagmes nominaux, telles que nom-modifieur, et les comparent dans deux parties d'un corpus. Les auteurs cherchent ainsi à observer des similitudes et des différences de sens, à partir des dépendances spécifiques à chacune des parties et de celles qui sont partagées.

Cette vision plutôt restrictive du contexte syntaxique, comme l'affirment les auteurs (*Ibid.* : 114), ne semble pas être la plus répandue aujourd'hui. Grefenstette (1996 : 209) définit par exemple le contexte syntaxique d'un nom comme l'ensemble des adjectifs, noms et verbes qui

¹²⁴ Nous décrivons cet échantillon plus en détail au Chapitre 7. Soulignons cependant que les verbes et les adjectifs extraits par TermoStat et conservés à l'issue des différentes étapes de filtrage des résultats bruts de TermoStat (cf. Chapitre 4, 4.3) n'apparaissent pas avec une fréquence suffisamment élevée dans tous les sous-corpus pour être retenus ici. Pour cette même question de fréquence, la sélection de cet échantillon a été opérée sur la liste obtenue après le troisième filtre et, dans le cas de plusieurs termes, le recours à un hyperonyme a été nécessaire (voir en particulier l'annexe B).

¹²⁵ Dans certains cas (par exemple Van de Cruys, 2008), les relations de dépendance prises en compte ne sont tout simplement pas précisées, nous ne pouvons alors que supposer que toutes les relations de dépendance sont sélectionnées.

entretiennent une relation syntaxique avec lui. Il s'agit également du choix de Tutin (2007b), dans une tâche de classification automatique des noms du lexique scientifique transdisciplinaire. Halskov (2005), en revanche, ne s'intéresse qu'aux modificateurs adjectivaux et nominaux des termes nominaux, afin d'observer des évolutions majeures dans la distribution des termes, qui refléteraient des évolutions de sens dues au processus de déterminologisation. Tanguy *et al.* (2015 : 109*sqq.*), qui évaluent plusieurs approches distributionnelles tenant compte de contextes syntaxiques et de contextes à fenêtres de mots, vont plus loin. Ils distinguent ainsi l'exploitation des dépendances « brutes », c'est-à-dire telles que les présentent un analyseur syntaxique et où toutes les relations sont considérées, d'une « exploitation plus sophistiquée » où « des transformations et normalisations [sont appliquées] au réseau de dépendances fourni par l'analyseur ». Dans ce dernier cas, certaines relations sont exclues, telles que la relation *det* (déterminant), afin de privilégier les unités les plus pertinentes d'un point de vue sémantique. Enfin, Morlane-Hondère (2013 : 42) cite différents travaux où seules les relations sujet-verbe et verbe-objet (voire uniquement verbe-objet) définissent le contexte syntaxique d'un nom. Il poursuit en concluant que

« [l]a question de savoir quel type de relation conserver dans un modèle distributionnel syntaxique et quelles en sont les conséquences sur les rapprochements générés n'a pas encore de réponse claire. » (*Ibid.*)

En dépit de cela, il ressort des travaux que nous citons que les contextes syntaxiques des noms sont le plus souvent composés des verbes dont ils sont sujets ou objets ainsi que des adjectifs qui les modifient. De même, les relations qui impliquent des mots grammaticaux (telles que la relation *det*) sont plus rarement prises en compte (Lenci, 2018 : 154). De manière générale, il semble que cette sélection ait pour but de focaliser l'analyse sur les unités les plus pertinentes d'un point de vue sémantique, en évitant donc les mots grammaticaux, et de produire, à terme, les résultats les plus significatifs.

Pour ces raisons, et étant donné que tous les termes sélectionnés dans cette section sont des termes nominaux, nous faisons le choix de nous focaliser sur les relations nom-modifieur, sujet-verbe et verbe-objet. Précisons que l'étiquetage opéré par Talismane met en relation chaque unité d'une phrase avec une seule autre unité (Urieli, 2013 : 26). Les contextes syntaxiques que nous récupérons automatiquement sont donc contraints également par l'étiqueteur utilisé. Par ailleurs, nous considérons la relation nom-modifieur, et non pas nom-adjectif comme dans les travaux cités *supra*, du fait que les unités qui partagent une relation de type « modifieur » avec un nom ne sont pas nécessairement toujours des adjectifs au sens strict, du moins pas pour Talismane. Aussi, si nous ne tenons pas compte des autres catégories grammaticales pouvant entrer dans cette relation, alors nous courons le risque d'introduire un certain silence dans les résultats. Par conséquent, pour

chaque terme, nous récupérons tous les verbes dont il est objet et/ou sujet et ses modificateurs, pour autant que ces derniers soient des adjectifs, des noms (propres ou communs) ou des verbes au participe passé¹²⁶.

Afin de récupérer ces unités, nous travaillons à partir d'une version du corpus analysée syntaxiquement en dépendances avec Talismane et nous systématisons cette étape grâce à l'exploitation d'une base de données MySQL¹²⁷. Cette solution nous permet de créer une base de données relationnelle à partir des sorties de Talismane, où les unités partageant une relation de dépendance peuvent aisément être mises en correspondance. Il suffit alors d'effectuer une requête SQL, où le terme en question et la relation visée sont spécifiés, pour récupérer toutes les unités qui correspondent à nos critères.

Les résultats bruts issus de ces requêtes SQL consistent en des listes d'unités, correspondant à chaque relation, par exemple, la liste des verbes lorsque le terme *particule* est sujet. Afin de rendre ces données exploitables pour l'analyse, nous les prétraitons à l'aide d'un script R¹²⁸ (R Core Team, 2018) qui permet de regrouper les unités selon leur fréquence et de présenter les informations extraites dans tous les sous-corpus dans un seul tableau. Cette première étape de traitement des données brutes facilite les comparaisons et l'identification de différences entre les sous-corpus. Plus précisément, et à partir des sorties brutes de la base de données (comme dans l'exemple du tableau 5.2), un tableau est généré (exemple du tableau 5.3). Les exemples illustrés dans les tableaux 5.2 et 5.3 sont issus d'un extrait des verbes dont le terme *particule* est sujet.

¹²⁶ Il semble en effet qu'il y ait quelques erreurs d'étiquetage dans le corpus, où des adverbes modifient des noms. Ces cas ne sont pas pertinents pour notre analyse et nous les écartons d'emblée à cette étape.

¹²⁷ Nous remercions sincèrement Philippe Baudrion pour son aide précieuse dans la conception, la gestion et l'interrogation de cette base de données.

¹²⁸ Ce script provient de l'atelier *Multivariate statistics in corpus-linguistic analyses* donné par Guillaume Desagulier lors de l'école doctorale *Current methods in corpus linguistics 2018* de la Conférence universitaire de Suisse occidentale (CUSO), que nous avons adapté pour correspondre à nos données.

traverser	SC-C
interagir	SC-C
peser	SC-P
traverser	SC-P
décrire	SC-P
interagir	SC-P
traverser	SC-R
décrire	SC-R
interagir	SC-R
interagir	SC-S
décrire	SC-S
traverser	SC-S
définir	SC-S
interagir	SC-S
traverser	SC-S
décrire	SC-V
interagir	SC-V
décrire	SC-V
interagir	SC-V
transférer	SC-V
traverser	SC-V
interagir	SC-V
traverser	SC-C

Tableau 5.2 : Extrait des résultats bruts issus de la relation sujet-verbe pour le terme *particule*

	SC-S	SC-C	SC-R	SC-V	SC-P
décrire	1	0	1	2	1
définir	1	0	0	0	0
interagir	2	1	1	3	1
peser	0	0	0	0	1
transférer	0	0	0	1	0
traverser	2	1	1	2	1

Tableau 5.3 : Extrait des résultats présentés sous forme de tableau

La première colonne du tableau 5.3 regroupe les verbes extraits dans la base de données MySQL et les cinq colonnes suivantes indiquent la fréquence de chacun de ces verbes dans les sous-corpus. Ce traitement appliqué aux données brutes représente également une première étape

d'analyse, car il est désormais possible de repérer assez rapidement les unités qui n'apparaissent que dans un sous-corpus, par exemple. Nous y reviendrons dans le Chapitre 7.

Ce traitement est appliqué aux résultats issus des 15 termes sélectionnés pour cette section, pour chacune des trois relations identifiées. Il en résulte 45 tableaux, sur le modèle du tableau 5.3.

Ce chapitre nous a permis de détailler le choix de deux indices visant à aborder de manière complémentaire la distribution des termes dans le corpus, afin de faire émerger, à partir des comparaisons entre les sous-corpus, des différences que nous pouvons interpréter en relation avec le processus de déterminologisation. Nous avons décrit la mise en œuvre de ces indices, qui repose sur différents paramètres et fait intervenir plusieurs outils.

La description de ces deux indices complète le cadre méthodologique que nous mettons en place dans notre thèse. Dans cette deuxième partie, nous avons donc posé les bases pour une méthodologie outillée d'analyse des termes dans le processus de déterminologisation, tel qu'il peut être observé en corpus. Nous avons ancré notre travail dans une démarche de recherche outillée en corpus. Par la mise en place d'un cadre méthodologique solide et transparent, nous avons montré l'importance d'appuyer la démarche sur des données réelles et d'appréhender leur analyse à partir d'indices précis.

Dans la partie suivante, nous nous concentrons sur l'exploration de ces deux indices dans le corpus. Dans le Chapitre 6, nous détaillons les différents phénomènes mis en lumière par l'exploration des *keywords* et, dans le Chapitre 7, nous montrons ce que l'analyse des contextes syntaxiques révèle sur le fonctionnement des termes dans le processus de déterminologisation. Enfin, nous consacrons le Chapitre 8 à la synthèse des différents éléments repérés par l'exploration des indices et montrons en quoi ils permettent d'alimenter la réflexion sur la déterminologisation et de mieux définir ce processus.

Troisième partie

Appréhender la déterminologisation par
la distribution des termes en corpus

Chapitre 6 Exploration des *keywords* en corpus

Ce chapitre est consacré à l'exploration dans le corpus du premier indice de type distributionnel défini dans notre travail. Dans ce but, nous nous concentrons sur les 500 premiers *keywords* issus d'une méthode de repérage essentiellement *corpus-driven*, lesquels représentent les différences les plus saillantes dans la distribution des termes dans le SC-P par rapport au SC-S (cf. Chapitre 5, 5.3.1). Dans une optique exploratoire, nous cherchons d'abord à dresser un premier portrait de ces différences à partir de l'observation globale des *keywords*. Ensuite, nous nous servons de ces unités comme des « marqueurs » afin de cibler l'analyse sur les contextes du SC-P où elles cooccurrent avec les termes à l'étude. Nous faisons ici l'hypothèse que, dans ces contextes, le fonctionnement des termes se distingue le plus de leur fonctionnement dans le SC-S, ce qui peut révéler des phénomènes qui sont à interpréter en relation avec la déterminologisation. Enfin, nous explorons les sous-corpus intermédiaires et observons de quelles manières ces fonctionnements se manifestent dans les étapes intermédiaires du processus représentées par ces sous-corpus. Nos observations mettent en lumière le fait que certains fonctionnements repérés dans le SC-P sont également attestés dans les sous-corpus intermédiaires, ce qui nous permet de nuancer nos hypothèses sur la déterminologisation et de proposer de nouveaux éléments de réflexion.

6.1. Observation globale des *keywords*

6.1.1. Regroupement des *keywords*

L'observation des 500 premières unités identifiées comme des *keywords* révèle une certaine diversité de fonctionnements auxquels ces unités peuvent renvoyer. Par exemple, parmi les premiers *keywords*, nous repérons aisément des noms propres (*Rolf* et *Heuer* par exemple), des noms de lieux (*Meyrin*), des adjectifs axiologiques (*fantastique* ou *phénoménal* par exemple) ou encore des termes de physique des particules (*électronvolt* par exemple). Cependant, afin de limiter les biais de surinterprétation possibles associés à l'observation d'une liste d'unités décontextualisées aussi hétérogène, les contextes d'apparition de ces unités dans le SC-P sont systématiquement vérifiés et il apparait que certaines unités renvoient à des fonctionnements de même nature et peuvent, de ce fait, être regroupées en plusieurs catégories homogènes.

6.1.1.1. Repérage de catégories homogènes

Avant de détailler les différents regroupements opérés, soulignons l'importance de la prise en compte des contextes dans cette étape à l'aide d'un exemple. Parmi les 500 unités retenues, nous

trouvons *guerre*. Sans tenir compte des contextes dans le SC-P, il serait aisé de supposer un lien entre la recherche en physique des particules et des applications militaires, par exemple par analogie avec la recherche en physique nucléaire et la bombe atomique pendant la Seconde Guerre mondiale (cf. Chapitre 4, 4.1.2). Or, les contextes¹²⁹ ne permettent pas de vérifier cette hypothèse. Ceux-ci montrent au contraire que l'unité *guerre* est utilisée dans des expressions figées (exemple 2 ci-dessous), dans des références culturelles (exemple 3), en référence à la période où le CERN a été créé, après la Seconde Guerre mondiale (exemple 4) ou encore dans des contextes évoquant les collaborations entre scientifiques de nationalités différentes pendant la guerre froide (exemple 5).

2. Est-ce de bonne **guerre** de la part du CERN? Oui. (SC-P, Le Temps, 2014)
3. "Dis papa, à quoi ça sert un accélérateur de particules ? C'est comme dans la **guerre des étoiles** ?" (SC-P, Les Échos, 2011)
4. Créé [sic] au lendemain de la **guerre**, l'organisation compte aujourd'hui 20 Etats membres engagés pour construire et opérer des accélérateurs de particules [...] (SC-P, Les Échos, 2012)
5. Il fut une époque où Américains et Soviétiques se rencontraient à la cafétéria du CERN pour tenter de rendre la **guerre froide** un peu moins froide. (SC-P, Le Monde 2016)

L'observation des contextes des unités dans le SC-P nous permet d'identifier plusieurs thématiques ou fonctionnements. Dans cette perspective, nous faisons l'hypothèse que les unités pouvant être regroupées dans une catégorie sur la base de leurs contextes dans le SC-P peuvent mettre en évidence des fonctionnements similaires ou des groupes de phénomènes à propos de l'usage des termes dans ce sous-corpus, que nous pouvons lier à la déterminologisation. À ce stade, nous identifions cinq catégories, composées d'unités :

- qui sont typiques des discours journalistiques,
- qui relèvent de la physique des particules,
- qui apparaissent dans des contextes rapportant une anecdote ou un évènement particulier,
- qui suscitent l'imagination,
- ou encore qui renvoient à d'autres domaines.

Unités typiques des discours journalistiques. La première catégorie regroupe des unités que nous jugeons typiques des discours journalistiques. Nous nous basons ici essentiellement sur le *Manuel de l'agencier* de l'Agence télégraphique suisse (ATS, 2008), ainsi que sur le *Manuel de l'agencier*

¹²⁹ Sans précision de notre part, l'unité *contexte* est employée pour désigner les contextes d'apparition des unités ou des termes dans le corpus.

de l'Agence France-Presse (AFP, 2004), qui précisent que, dans les dépêches, les informations doivent être correctement référencées, c'est-à-dire que les sources, y compris les personnes et leur nationalité lorsque c'est approprié, les lieux et les dates d'un évènement particulier doivent être mentionnés. Ces indications se retrouvent également dans des manuels dédiés à l'écriture journalistique, qui expliquent que les six questions « qui, quoi, où, quand, comment, pourquoi » doivent toujours être traitées (Brisson *et al.*, 2016 : Chapitre 1 ; Martin-Lagardette, 2016 : 40-42). Si ce type d'information tient du standard de rédaction journalistique, ou du moins de « bonnes pratiques », alors nous pouvons en déduire que la présence d'unités dénotant des noms propres ou des noms de pays, par exemple, sert à répondre à ces questions. Il est donc naturel que ce type d'unités apparaisse avec une fréquence plus élevée dans le SC-P que dans le SC-S, puisque le premier sous-corpus est composé d'articles de presse. Par ailleurs, il est également précisé que, selon le type d'informations concerné, un « langage plus imagé » peut être utilisé (ATS, 2008 : 24), le recours à des expressions figées est encouragé (Martin-Lagardette, 2016 : 62-65) et observé de manière récurrente (Watine, 2006 : 87-90). Enfin, nous remarquons que l'utilisation du pronom *vous*, notamment, est parfois conseillée pour se rapprocher du lectorat ou pour adopter un style plus familier (Martin-Lagardette, 2016 : 60-61). À partir de ces éléments, nous déduisons que les unités citées ci-dessous et que nous qualifions de typiques des discours journalistiques relèvent bien des spécificités du style journalistique. Nous observons ainsi :

- des noms propres (*Rolf, Heuer, Berners* (pour Tim Berners-Lee), *Etienne* (pour Etienne Klein), *Hawking* (pour Stephen Hawking) par exemple),
- des noms de pays, de communes ou de régions (*Caen, Meyrin, Europe, Russie, Grenoble, Suède* par exemple),
- des unités qui consistent en des compléments de temps dans le corpus (*lundi, mardi, matin* par exemple),
- des pronoms et des possessifs (*vous, votre*),
- des unités qui font partie d'expressions figées (*feu* dans *sous les feux des projecteurs* ou *donner le feu vert, filet* dans *le filet se resserre* ou *les mailles du filet, bataille* dans *cheval de bataille* par exemple).

Unités relevant de la physique des particules. La deuxième catégorie englobe des unités qui évoquent le domaine de la physique des particules. Bien que le SC-P ne soit pas un corpus de spécialité mais d'articles de presse, il apparaît que certaines formes renvoient à des termes de physique des particules. Ces formes sont considérées comme des *keywords* lors de l'extraction, car

leur fréquence est significativement plus élevée dans le SC-P que dans le SC-S. Il s'agit le plus souvent de variantes dénominatives de termes de physique des particules (c'est-à-dire des dénominations qui renvoient au même concept¹³⁰) qui ne sont pas réparties de manière équilibrée entre ces deux sous-corpus. Par exemple, la forme *antiélectron* est plus caractéristique du SC-P que du SC-S (12 occurrences dans le SC-P mais une seule dans le SC-S, soit 92,3 % et 7,7 % du total des occurrences de cette forme dans les deux sous-corpus), raison pour laquelle elle est repérée. En revanche, sa variante *positron*, qui n'est pas identifiée comme un *keyword*, présente une répartition plus équilibrée (79 occurrences dans le SC-P et 91 occurrences dans le SC-S, soit 46,5 % et 53,5 % du nombre total d'occurrences dans ces deux sous-corpus). De la même manière, la forme *électronvolt* apparaît uniquement dans le SC-P, en particulier dans les syntagmes *million d'électronvolts* et *milliard d'électronvolts*, alors que, dans le SC-S, les abréviations *eV*, *MeV* et *GeV* sont plus généralement utilisées.

Unités se rapportant à une anecdote. La troisième catégorie regroupe des unités qui apparaissent dans des contextes narrants une anecdote ou un événement particulier. Cette interprétation est possible uniquement par l'examen minutieux des contextes dans le sous-corpus, qui nous permet de remarquer que, lorsqu'une anecdote plutôt cocasse ou un événement particulier sont traités dans la presse, ils le sont avec une certaine récurrence. C'est précisément cette récurrence qui permet de repérer, avec la méthode des *keywords*, certaines des unités utilisées pour rendre compte de ces anecdotes ou de ces événements. Par exemple, dans les contextes 6 et 7 ci-dessous, un léger incident survenu au LHC est raconté. Cet incident est repris dans plusieurs articles et les unités *miette*, *pain* et *oiseau* apparaissent avec une fréquence suffisamment élevée par rapport au SC-S pour qu'elles soient considérées comme des *keywords*. Le même raisonnement peut être suivi pour d'autres unités, par exemple *EPFL*¹³¹ et *terroriste*. Ces unités apparaissent en effet à plusieurs reprises dans des articles portant sur l'affaire d'un chercheur du CERN, travaillant également à l'EPFL, qui a été accusé d'actes terroristes, comme le montre le contexte 8. La récurrence de la cooccurrence de ces deux unités avec des termes de physique des particules soumis à l'analyse permet au système de les considérer comme des *keywords*.

6. Des **miettes** de **pain** lâchées par un **oiseau** sur une installation extérieure de l'accélérateur de particules du Centre européen de recherche nucléaire (CERN) ont déclenché un "faible" réchauffement de la machine. (SC-P, Le Temps, 2009)

¹³⁰ Sur la question de la variation dénominative en terminologie, voir en particulier Freixa (2006), Dury et Lervad (2008), Picton (2009), Pelletier (2012), Kim (2017), Freixa et Fernández-Silva (2017) ou encore Picton et Dury (2017).

¹³¹ École polytechnique fédérale de Lausanne, <https://www.epfl.ch/fr/> (page consultée le 18 mai 2021).

7. Les réparations terminées, le collisionneur a connu une autre mésaventure, plus cocasse que grave : des **miettes de pain**, peut-être tombés [sic] du bec d'un **oiseau**, ont causé un court-circuit dans un transformateur extérieur. (SC-P, Le Monde, 2010)
8. Le Franco-Algérien, arrêté le 8 octobre, travaillait au CERN et à l'**EPFL**. Il conversait par Internet avec des **terroristes** d'Al-Qaida au Maghreb islamique. (SC-P, Le Temps, 2009)

Unités suscitant l'imagination. La quatrième catégorie regroupe des unités qui suscitent l'imagination ou qui évoquent des aspects mystérieux ou extraordinaires. Il peut s'agir d'adjectifs évaluatifs axiologiques, comme *fascinant*, *énigmatique*, *élusif*, *fantastique* ou *phénoménal*, ou de noms propres véhiculant une référence culturelle, notamment *Graal* et *Arlésienne*. Ces unités renvoient par ailleurs au fait que la physique des particules est un sujet susceptible d'être traité d'un point de vue sensationnaliste dans la presse, comme nous l'expliquons au Chapitre 4 (4.1.1.1), et marquent donc cette idée de sensationnalisme dans les contextes des termes observés.

Unités renvoyant à d'autres domaines. La dernière catégorie concerne des unités qui évoquent d'autres domaines que la physique des particules, par exemple *cancer*, *IRM*, *ITER*, *combustible*, *fossile*, *art*, *musée*, *satellite*, *financier*, *budgétaire*, *biologie*, *climat*. À ce stade, nous jugeons qu'il est préférable de considérer ces unités comme des *unités qui renvoient à d'autres domaines*, plutôt que comme de véritables termes. En effet, nous nous appuyons ici essentiellement sur nos propres connaissances ainsi que sur la distribution de ces unités dans les contextes du SC-P pour les associer effectivement à des domaines différents de la physique des particules. Par exemple, dans le contexte 9 ci-dessous, plusieurs unités, mises en évidence en gras, nous permettent de comprendre que l'unité *IRM*, identifiée comme un *keyword*, est utilisée dans un contexte lié au domaine médical. Néanmoins, dans la mesure où nous ne procédons à aucune validation de ces unités par d'éventuels experts de ces domaines, contrairement aux termes de physique des particules, nous ne pouvons pas affirmer qu'il s'agit de termes appartenant à ces domaines.

9. Pour ce qui est de la prise en charge des **malades**, dans le plan **Hôpital** 2007 (qui recoupe aussi celui contre le **cancer**), il est prévu d'améliorer l'équipement français en **radiologie** de pointe, en installant, d'ici à quatre ans, notamment 111 accélérateurs de particules, 54 *IRM* et 39 **scanners**, ainsi que 75 **tomographies à positon**. (SC-P, Le Figaro, 2004)

Précisons enfin que toutes les unités apparaissant dans les 500 premiers résultats ne sont pas classées dans une de ces catégories, certaines sont en effet plus difficiles à regrouper. Les contextes de ces unités sont très hétérogènes et il ne semble pas que nous puissions en tirer profit dans cette étude. Par exemple, des unités telles que *langue*, *annonce*, *femme*, *papier* et *salle* apparaissent dans les résultats, mais leurs contextes restent plus difficilement interprétables, comme le montrent

les exemples 10 à 14 ci-dessous. De fait, nous ne les considérons pas dans la suite de nos observations.

10. A l'Organisation européenne pour la recherche nucléaire, une seule **langue** se parle : celle du savoir. Y est né le LHC, instrument quasi parfait pour comprendre l'univers. (SC-P, Le Monde, 2010)
11. L'**annonce** a été faite par le directeur général de l'Organisation européenne pour la recherche nucléaire (CERN) de Genève, Rolf Heuer, à la Conférence européenne sur la physique des hautes énergies, qui se tenait du 21 au 27 juillet à Grenoble. (SC-P, Le Monde, 2011)
12. Vous êtes la première **femme** à assurer la direction générale du CERN. (SC-P, Le Temps, 2016)
13. Postulée sur le **papier**, elle a finalement été observée. Et créée dans les grands accélérateurs de particules. (SC-P, Le Figaro, 2005)
14. Enfouies à 1400 mètres sous terre pour être à l'abri des particules des rayons cosmiques, les **salles** ont été construites expressément dans la direction de Genève. (SC-P, Le Temps, 2010)

6.1.1.2. Catégories retenues pour notre travail

Bien que différents regroupements soient opérés à partir des unités identifiées comme des *keywords*, seuls deux de ces regroupements sont conservés pour notre étude : les unités suscitant l'imagination et celles renvoyant à d'autres domaines. Les autres catégories ne sont pas retenues pour les raisons suivantes.

Premièrement, en ce qui concerne les unités qualifiées de typiques de la presse, dans la mesure où, d'après les sources citées *supra*, ces unités peuvent être rattachées au style et aux contraintes rédactionnelles des discours journalistiques de manière générale, il semble qu'elles ne soient pas caractéristiques des discours de la presse autour de la physique des particules en particulier. Ces unités n'apporteraient alors que peu d'éléments à propos du fonctionnement des termes de ce domaine en situation de déterminologisation dans le SC-P.

Deuxièmement, pour ce qui est des unités qui sont en réalité des variantes dénominatives de termes de physique des particules, bien que la répartition des variantes dans les sous-corpus constitue un angle intéressant, ce type d'observations dépasse le cadre de l'analyse de la distribution des termes telle que nous la mettons en œuvre dans notre thèse et devrait être abordée dans une étude à part entière.

Troisièmement, nous ne conservons pas non plus les unités qui sont liées à une anecdote ou à un évènement particulier, car leur attestation dans le SC-P ne semble pas renvoyer à un changement durable ou un fonctionnement stable qui pourrait être mis en relation avec la déterminologisation. Ce type d'observations est un écueil plutôt commun en analyse

distributionnelle, où, selon Habert *et al.* (2005 : 313), par exemple, un changement conséquent dans la distribution d'une unité ne renvoie pas nécessairement à un changement sémantique. La présence d'un évènement particulier peut en effet engendrer certaines modifications importantes dans la distribution d'une unité, mais sans grande incidence sur leur fonctionnement sur le long terme. C'est ce que soulignent par exemple Del Tredici *et al.* (2019 : 2072-2073), qui, dans leur étude de l'évolution du sens en diachronie courte, constatent ce qu'ils appellent des « faux positifs », c'est-à-dire des unités pour lesquelles les changements importants observés dans leur distribution sont en réalité liés à une personne ou à un évènement particulier d'une ampleur modérée. Les auteurs donnent notamment l'exemple de l'unité *independence* (en anglais), dont les différences de distribution renvoient au contexte politique de la Catalogne dans leur sous-corpus le plus récent, et non à une évolution sémantique stable.

De manière similaire, ce type de changement non significatif du point de vue sémantique peut également être révélateur de la centralité d'un évènement ou d'une thématique à un moment donné. Ainsi, l'augmentation de la fréquence de certaines unités dans la distribution d'une unité, comme dans l'étude de Del Tredici *et al.*, peut refléter le fait que les personnes ou les évènements auxquels ces changements renvoient se trouvent en réalité au centre des préoccupations des locuteurs émetteurs et destinataires des textes pris en considération dans le corpus, à un instant T. Nous empruntons cette notion de centralité à Jacques (2006), qui explique que lorsque le référent d'un terme complexe est central dans un domaine, alors une variante réduite de ce terme est plus souvent utilisée, car la centralité du référent permet une interprétation correcte de la variante réduite, dans le domaine considéré. Cette notion peut aussi être mise en œuvre pour rendre compte de certaines différences distributionnelles, qui ne sont pas le reflet d'un sens qui aurait évolué dans le temps, mais qui sont plutôt le fait d'une attention accrue dans les textes du corpus sur un évènement particulier, comme dans l'étude de Del Tredici *et al.* C'est également en ce sens que vont les propositions de Picton sur la centralité (2009 : 208-209).

Ainsi, deux catégories d'unités uniquement sont conservées. La première concerne les unités qui suscitent l'imagination et qui renvoient à la notion de sensationnalisme. Soulignons cependant que, bien que ces unités puissent relever des caractéristiques stylistiques des discours de presse et que leur attestation dans le SC-P ne soit probablement pas spécifique au domaine de la physique des particules, il nous semble intéressant de poursuivre nos observations avec ces unités. En effet, celles-ci suscitent l'imagination des lecteurs ou renvoient à des aspects mystérieux ou

extraordinaires de la recherche en physique des particules¹³². Autrement dit, elles semblent particulièrement évocatrices. Or il est possible que la cooccurrence de ces unités avec certains termes, si elle est récurrente, puisse mener à véhiculer certaines connotations dans le sens de ces termes, du moins est-ce notre hypothèse à ce stade. En outre, d'après Meyer et Mackintosh (2000b : 115), l'observation de nouvelles connotations fait partie des conséquences sémantiques de la déterminologisation. Partant de cette affirmation, il nous semble pertinent de poursuivre l'analyse avec ces unités afin de vérifier notamment si elles ajoutent véritablement certaines connotations aux termes que nous étudions et si ces connotations contribuent aux changements sémantiques survenant dans le processus de déterminologisation. Dans la suite de notre travail, nous faisons référence à ces unités comme appartenant à la catégorie « sensationnalisme ».

La seconde concerne les unités renvoyant à d'autres domaines que la physique des particules. Il nous semble pertinent de nous intéresser de plus près à ces unités, car d'une part elles ne sont *a priori* pas caractéristiques des spécificités stylistiques ou des contraintes rédactionnelles des discours journalistiques et, par conséquent, pourraient nous éclairer sur le fonctionnement des termes de physique des particules en situation de déterminologisation. D'autre part, nous observons une réelle récurrence dans la cooccurrence de ces unités et de plusieurs des termes à l'étude. Cette récurrence suggère que des changements se produisent dans l'environnement distributionnel des termes de manière plus durable et plus stable que les changements liés à une anecdote ou à un événement particulier, comme discuté *supra*. Il nous semble que ce type de changement est particulièrement pertinent pour notre étude, dans la mesure où il correspond dans une plus large mesure aux conséquences de la déterminologisation décrites dans la littérature (cf. Chapitre 2, 2.1.1.3 et 2.2.2.1 notamment). Nous faisons donc l'hypothèse à ce stade que les contextes où ces unités coexistent avec des termes de physique des particules sont susceptibles d'indiquer la présence de phénomènes attribuables au processus de déterminologisation.

Parallèlement, c'est également dans une optique exploratoire que nous abordons l'analyse des contextes de ces unités. Nous cherchons ainsi à comprendre ce que révèle la cooccurrence de ces unités avec les termes soumis à l'analyse à propos de ces termes. Autrement dit, nous poursuivons le but non seulement de comprendre les raisons de leur cooccurrence avec les termes de physique des particules, mais également d'identifier la nature précise des phénomènes mis en évidence par cette cooccurrence et ce qu'ils signifient du point de vue du processus de déterminologisation. En outre, il est possible que les phénomènes repérés de cette manière diffèrent

¹³² Les unités classées dans cette catégorie sont les suivantes : *élusif, énigmatique, énorme, fameux, fantastique, fantomatique, faramineux, fascinant, fugace, furtif, immense, insaisissable, magnifique, minuscule, monstre, mystérieux, mythique, phénoménal, révolutionnaire, secret, vertigineux, Arlésienne, Graal, mystère et révolutionner*.

selon le domaine que les unités évoquent, du moins souhaitons-nous tenir compte de cette possibilité dans l'analyse. C'est pourquoi nous faisons le choix de distinguer les différents domaines (ou groupes de domaines lorsque ceux-ci sont particulièrement proches) que nous pouvons identifier à partir des unités et donc de classer les unités en catégories plus restreintes. Les catégories ainsi obtenues sont¹³³ :

- **économie et finances**, pour les unités *franc, euro, coûter, marché, financier, dollar, crise, financer, investir, budgétaire, financement, économique, banque* et *salaires* ;
- **spatial**, pour les unités *ISS, astronome, observatoire, astronomie, navette, astrophysicien, astre, Antares, Voie lactée, astronomique, satellite, supernova* et *cosmos* ;
- **médical**, pour les unités *imagerie, médical, tumeur, hadronthérapie, radiothérapie, tissu, cancer, IRM, médecine, santé, cancéreux, protonthérapie, tomographie* et *AAA* ;
- **énergie nucléaire**, pour les unités *Iter* et *combustible* ;
- **arts et culture**, pour les unités *livre, musée, Louvre* et *art* ;
- **archéologie et paléontologie**, pour les unités *verre* et *fossile*.

Avec les unités de la catégorie « sensationnalisme », 74 unités sont retenues à l'issue de cette étape. Précisons enfin que les ambiguïtés possibles selon l'unité et son rattachement à une catégorie ont systématiquement été levées par la prise en compte des contextes d'apparition de l'unité dans l'échantillon du SC-P utilisé pour l'extraction de *keywords* (cf. Chapitre 5, 5.3.1.1). Par exemple, dans les contextes compris dans cet échantillon, l'unité *tissu* apparaît toujours en lien avec le domaine médical, et notamment lorsque des traitements des tumeurs par rayonnement sont évoqués, comme dans les exemples suivants :

15. Il s'agit en majorité (170 000 cas) de traitements ionisants (photons et électrons) dont le dosage est délicat, car ces rayonnements détruisent certains **tissus** adjacents à la tumeur. (SC-P, Les Échos, 2004)
16. La thérapie par protons permet de délivrer des doses d'irradiation très importantes sur une tumeur, tout en respectant les **tissus** sains avoisinants. (SC-P, Le Temps, 2006)

¹³³ Précisons que nous retenons uniquement les catégories pour lesquelles au moins deux unités ont été relevées.

6.1.2. Limitation de la subjectivité par le recours à un accord inter-annotateurs

Malgré la prise en compte systématique des contextes d'apparition des unités dans le SC-P, le classement des *keywords* en différentes catégories peut sembler subjectif. Cette part de subjectivité est toutefois inévitable dès lors que le jugement humain intervient dans l'interprétation des données (cf. Chapitre 3, 3.1.3). Cependant, comme le propose Jacques (2016 : 94-96), et lorsque les données s'y prêtent, cette part de subjectivité peut être limitée, par exemple par l'interrogation de personnes extérieures à l'étude. À partir du jugement de ces personnes, un accord inter-annotateurs peut être calculé, qui fournit une mesure objective de la fiabilité du jugement de l'analyste. C'est de cette manière que nous proposons d'estimer la fiabilité de notre jugement et, partant, d'estimer notre propre subjectivité.

Plus précisément, le recours à une mesure de l'accord inter-annotateurs constitue un moyen de calculer le degré d'accord entre plusieurs annotateurs qui émettent un jugement à propos d'un jeu de données (Hallgren, 2012 : 23). Selon le résultat de cette mesure, le jugement peut être considéré comme fiable, et la subjectivité attribuée à l'analyste est moindre. Ce dernier point n'est pas dénué d'importance, car, si l'on se place dans une optique de recherche reproductible, alors il devient nécessaire de démontrer que les jugements émis par l'analyste sur certaines données, et sur lesquels la recherche se fonde, ne reflètent pas uniquement l'avis individuel de l'analyste (Carletta, 1996 : 249). En ce sens, la mesure de l'accord inter-annotateurs permet de donner plus de poids aux résultats issus de ce type de jugements. De même, le calcul d'un accord inter-annotateurs « éclaire sur la part de variation individuelle dans l'appréhension des phénomènes de langue et dans la compréhension linguistique » (Jacques, 2016 : 95) et, si l'accord est bon, alors nous pouvons considérer que cette « part de variation individuelle » n'a qu'une incidence moindre sur l'analyse et l'interprétation des données.

Dans notre cas, l'intérêt de procéder à un accord inter-annotateurs repose sur la possibilité de valider le classement des unités tel que nous l'avons effectué, c'est-à-dire d'observer dans quelle mesure les catégories que nous avons choisies pour chaque unité sont également choisies par des personnes extérieures à l'étude et si ces choix sont homogènes. Dans ce but, nous proposons une tâche de classification de ces unités à 10 participants, qui partagent un profil similaire. En effet, les participants sélectionnés sont des langagiers, qui ne sont pas experts du domaine de la physique des particules, dont l'âge varie entre 21 et 35 ans et dont le français est la langue maternelle.

6.1.2.1. Tâche de classification

La tâche proposée aux participants¹³⁴ a consisté à classer 114 unités dans une des 7 catégories ou dans une catégorie « autre », s'ils estimaient qu'une unité ne faisait partie d'aucune des catégories. Parmi ces 114 unités, 40 ont été ajoutées par nos soins, afin d'équilibrer quantitativement chaque catégorie. En effet, étant donné que le nombre d'unités par catégorie diffère, parfois dans une large mesure (par exemple, la catégorie « sensationnalisme » comprend 25 unités alors que la catégorie « arts et culture » n'en comprend que 4), il est préférable de limiter ces différences dans la tâche de classification. Ce choix permet d'éviter que le nombre inégal d'unités par catégorie ne soit un biais pour les participants. Ces 40 unités ne sont pas prises en compte dans le calcul de l'accord inter-annotateurs.

En outre, chaque unité a été accompagnée d'un exemple tiré du SC-P pour tenir compte de la polysémie et de l'homonymie possibles hors contexte. L'ajout d'un exemple a également permis de rapprocher la réalisation de la tâche par les participants de nos propres conditions de classement, car, comme nous l'expliquons *supra*, nous avons veillé à tenir compte systématiquement des contextes en effectuant les regroupements. Un seul exemple a néanmoins été ajouté pour chaque unité, dans le but de ne pas alourdir la tâche inutilement.

Concrètement, cette tâche a été proposée aux participants par le biais d'un questionnaire LimeSurvey sous la forme illustrée dans le tableau 6.1 ci-dessous. Pour chaque unité, les participants devaient cocher la case de la catégorie qui leur semblait convenir, ou bien celle de la catégorie « autre ». S'ils estimaient qu'une unité pouvait appartenir à plusieurs catégories, les participants avaient la possibilité de le signaler dans un commentaire à la fin de la tâche.

¹³⁴ Nous remercions sincèrement Karla Orihuela pour son aide indispensable et ses précieux conseils dans la conception et la réalisation de cette tâche.

	Sensationnalisme (imaginaire, mystère, extraordinaire)	Économie et finances	Spatial	Arts et culture	Médical	Archéologie et paléontologie	Énergie nucléaire	Autre
<p>franc</p> <p>L'instrument - le plus grand et plus complexe du monde, devisé à 4,7 milliards de francs - aura pour but principal de révéler, avec ses quatre détecteurs hauts comme des cathédrales (ALICE, ATLAS, CMS et LHCb), l'élusif "boson de Higgs".</p>								
<p>tissu</p> <p>Il s'agit en majorité (170 000 cas) de traitements ionisants (photons et électrons) dont le dosage est délicat, car ces rayonnements détruisent certains tissus adjacents à la tumeur.</p>								
<p>mystérieux</p> <p>Les chercheurs du CERN ont l'espoir de mettre la main sur d'autres objets aussi mystérieux : des particules supersymétriques qui pourraient constituer la " matière noire " de l'Univers.</p>								
<p>combustible</p> <p>Cette technologie très innovante consiste à utiliser un accélérateur de particules pour produire les neutrons qui déclenchent les réactions de fission dans le combustible nucléaire.</p>								

Tableau 6.1 : Exemple de la tâche de classification soumise aux participants

6.1.2.2. Calcul de l'accord inter-annotateurs et résultats

Le choix d'une mesure de l'accord inter-annotateurs dépend essentiellement du type de données sur lequel les annotateurs émettent un jugement, sur le nombre d'annotateurs participant à la tâche et sur la manière dont celle-ci a été conçue. Nous faisons le choix d'utiliser le kappa de Light, pour plusieurs raisons. D'abord, selon Hallgren (2012 : 26), le kappa de Cohen (Cohen, 1960) et ses variantes sont les mesures les plus appropriées pour estimer le degré d'accord entre différents annotateurs pour des variables catégorielles, c'est-à-dire des variables qualitatives qui ne peuvent pas être ordonnées (Brezina, 2018 : 7), ce qui correspond aux données évaluées dans notre cas. Par ailleurs, les mesures du kappa comportent l'avantage, par rapport à d'autres mesures similaires, de tenir compte des cas où l'accord entre les annotateurs serait uniquement le fruit du hasard (Cohen, 1960 : 40 ; Hallgren, 2012 : 26).

Enfin, nous exploitons le kappa de Light (Light, 1971) et non une autre variante de cette mesure en raison du nombre d'annotateurs et du fait que chaque annotateur émet un jugement sur chaque unité. En effet, le kappa de Cohen permet de calculer le degré d'accord entre deux annotateurs uniquement (Hallgren, 2012 : 27) et le kappa de Fleiss (Fleiss, 1972) entre plusieurs annotateurs dans des tâches où toutes les données ne sont pas évaluées par tous les annotateurs. Comme le dit Hallgren (2012 : 28), le kappa de Fleiss convient pour des études « where any constant number of m coders is randomly sampled from a larger population of coders, with each subject rated by a different sample of m coders ». De ce point de vue, seul le kappa de Light est adéquat pour les besoins de notre étude.

Le kappa de Light est calculé avec R et le package *irr* (Gamer *et al.*, 2012). Pour l'ensemble des unités, nous obtenons un kappa de 0,718, qui correspond à un accord substantiel, selon Landis et Koch (1977 : 165). Cela signifie que les participants ont de manière générale assigné la même catégorie à chaque unité, leur degré d'accord est donc satisfaisant. En revanche, cela ne signifie pas nécessairement que, pour chaque unité, la catégorie sur laquelle les participants sont d'accord correspond à la catégorie que nous avons définie. En effet, il est possible que, pour une unité, tous les participants choisissent une autre catégorie que la catégorie attendue. Un tel cas de figure ne serait pas visible dans le résultat de l'accord inter-annotateurs, car, fondamentalement, les participants seraient d'accord entre eux. Une étape de vérification supplémentaire est donc nécessaire.

À l'issue de cette étape, seule une unité, *verre*, n'a pas été assignée à la catégorie attendue par plus de la moitié des participants. À l'origine, cette unité a été classée dans la catégorie « archéologie et paléontologie », car les contextes sont le plus souvent en lien avec des questions de datation de

verre, comme le suggère l'exemple proposé dans la tâche et reproduit ci-dessous. Or, le fait que la majorité des participants n'a pas établi ce lien démontre que celui-ci n'était pas aussi évident que ce que nous avons estimé. Cette unité est donc retirée de l'analyse, de même que la catégorie « archéologie et paléontologie ». Sans l'unité *verre*, cette catégorie ne serait composée que d'une seule unité, *fossile*. Aussi, de la même manière que nous n'avons pas créé de catégories pour des unités seules, il semble peu pertinent de conserver cette catégorie avec une seule unité. L'échantillon d'unités sur lequel nos analyses portent est désormais composé de 72 unités.

17. Après la datation du vin par mesure de la radioactivité en césium 137, c'est au tour du **verre** des bouteilles de passer au crible de l'accélérateur de particules. (SC-P, Les Échos, 2008)

Grâce à un accord inter-annotateurs substantiel pour l'ensemble des unités classées par les participants et après avoir retiré une des catégories, l'échantillon est constitué de 6 catégories. Dans la suite de ce chapitre, et conformément au principe d'analyse énoncé au Chapitre 5 (5.2.3), nous caractérisons d'abord les contextes de ces unités dans le SC-P, lorsqu'elles apparaissent dans l'environnement distributionnel des termes à l'étude, puis nous tenons compte des sous-corpus intermédiaires (SC-C, SC-R et SC-V).

6.2. Identification de différents types de contextes dans le SC-P

Dans cette section, nous mettons au jour différents types de contextes dans lesquels les termes à l'étude sont attestés dans le SC-P. Plus précisément, nous adoptons une démarche essentiellement exploratoire, où les unités retenues *supra* servent de marqueurs et permettent de cibler l'analyse sur les contextes des termes dans le SC-P qui se distinguent *a priori* le plus des contextes dans le SC-S. Ainsi, à l'aide des unités des six catégories retenues, nous dégagons sept types de contextes illustrant des fonctionnements des termes différents du SC-S, que nous interprétons ensuite en relation avec la déterminologisation.

6.2.1. Contextes explicitant l'utilité ou l'utilisation pour d'autres domaines

Certains contextes indiquent une utilisation possible ou une utilité particulière de « découvertes » en physique des particules, ou de concepts de physique des particules, dans d'autres domaines. Dans ce premier type de contextes, l'utilité ou l'utilisation sont signalées par la présence

de différents marqueurs¹³⁵, comme l'illustrent les exemples suivants. Dans ces exemples, les termes de physique des particules sont indiqués en gras, les unités classées dans les catégories sont en italiques et les marqueurs sont soulignés.

18. Le PSI, seul centre suisse à pratiquer l'*hadronthérapie*, traite actuellement une centaine de patients par an grâce à des **faisceaux de protons**. (SC-P, Le Temps, 2014)
19. La *radiothérapie* classique utilise des **photons** (des "particules" de lumière), alors que la *protonthérapie* fait appel à des **protons**, autrement dit des noyaux d'hydrogène. (SC-P, Les Échos, 2006)
20. Ce mode de traitement utilise un **accélérateur de particules (synchrotron)** pour irradier avec une grande précision les *tumeurs* grâce à des **faisceaux de protons** et d'ions carbone (photo). (SC-P, Les Échos, 2008)

Ainsi, dans les exemples 18 à 20, qui illustrent des unités classées dans la catégorie « médical », les marqueurs sont *grâce à* (exemples 18 et 20), *utiliser* (exemples 19 et 20), et *faire appel à* (exemple 19). Ces exemples nous permettent alors de comprendre que certains concepts ne sont pas utilisés dans la recherche en physique des particules uniquement, mais ont une certaine utilité également en médecine.

Cette interprétation est par ailleurs consolidée avec l'observation de contextes explicitant clairement le fait que la recherche en physique des particules profite à d'autres disciplines, notamment la médecine. L'exemple 21 explique en effet que les recherches menées au Tevatron, un accélérateur de particules qui a été exploité jusqu'en 2011¹³⁶, ont contribué au développement de l'imagerie par résonance magnétique, dont le sigle, *IRM*, a été identifié comme un *keyword* et classé dans la catégorie « médecine ». De la même manière, l'exemple 22 explicite le bénéfice que la médecine, et notamment l'imagerie médicale, retire des recherches en physique des particules.

21. Les avancées réalisées grâce au **Tevatron** ont aussi eu des retombées pratiques révolutionnaires dont l'imagerie par résonance magnétique (*IRM*). (SC-P, Le Temps, 2011)
22. Internet, *médecine* : les étonnantes retombées des recherches menées au **Cern** [...] La recherche en physique des particules a également permis l'émergence de l'imagerie médicale. Sans oublier l'hadronthérapie, qui guérit certaines tumeurs en les bombardant avec des protons ou des ions carbone. (SC-P, Le Figaro, 2008)

Ce type de contexte est également repéré à l'aide d'unités des catégories « arts et culture » et « énergie nucléaire », il ne s'agit donc pas d'un fonctionnement spécifique à la catégorie « médecine ». Ainsi, dans les deux exemples suivants, les marqueurs *dédié à* et *grâce à* nous permettent

¹³⁵ Pour davantage de détails sur ces marqueurs d'utilité, voir Condamines *et al.* (à paraître).

¹³⁶ Tevatron, Fermilab, <https://www.fnal.gov/pub/tevatron/tevatron-accelerator.html> (page consultée le 18 mai 2021).

d'observer que les accélérateurs de particules peuvent aussi être utilisés dans des domaines liés à l'art, dans « la recherche sur les œuvres d'art », comme l'explique en particulier l'exemple 23.

23. Mais que va devenir l'**accélérateur de particules** du *Louvre*, Aglae, dédié à la recherche sur les œuvres d'art, dont l'installation a coûté 43 millions d'euros ? (SC-P, Les Échos, 2010)
24. L'analyse élémentaire de la terre cuite et de la polychromie a été, quant à elle, réalisée grâce à l'accélérateur de particules Aglae (situé sous les jardins du *Louvre*), en collaboration avec le Centre de recherche pétrographique et géochimique de Nancy. (SC-P, Les Échos, 2003)

L'exemple ci-dessous illustre un contexte évoquant le domaine de l'énergie nucléaire : la cooccurrence du terme *accélérateur de particules* avec l'unité *combustible*, qui a été classée dans la catégorie « énergie nucléaire », permet de constater l'utilisation d'accélérateurs de particules dans le domaine de l'énergie nucléaire également.

25. Cette technologie très innovante consiste à utiliser un **accélérateur de particules** pour produire les neutrons qui déclenchent les réactions de fission dans le *combustible* nucléaire. (SC-P, Les Échos, 2009)

Enfin, un dernier cas, quelque peu différent, peut être constaté. Dans ce cas, la présence d'un marqueur n'indique pas l'utilité ou l'utilisation d'un terme de physique des particules dans un autre domaine, mais plutôt l'utilisation d'un terme à la fois en physique des particules et dans un autre domaine. C'est le cas par exemple dans le contexte suivant, où l'on constate qu'une propriété de la supraconductivité est utile aussi bien dans le développement des accélérateurs de particules que dans le fonctionnement des IRM.

26. En supraconductivité, ce n'est pas le cas. Les courants électriques circulent bel et bien parfaitement. Leur résistance n'est pas proche de zéro, elle est tout simplement nulle. Cette propriété remarquable a permis de fabriquer des électroaimants de très grande puissance utilisés aujourd'hui dans les *IRM* ou dans les **accélérateurs de particules**. (SC-P, Le Figaro, 2015)

À ce stade, nous pensons que l'accent porté sur les utilisations ou les applications possibles pour d'autres domaines peut relever d'une stratégie de la part des médias pour susciter l'intérêt des lecteurs, en leur montrant de manière peut-être plus concrète comment les découvertes effectuées par la recherche en physique des particules ou les concepts développés dans ce domaine peuvent être utiles à la société, ou du moins à mettre en avant les rapports entre la science et la société, comme le suggère Peynaud (2014 : §9). Nous détaillerons davantage cet aspect dans le Chapitre 8.

6.2.2. Contextes évoquant d'autres domaines

D'autres contextes montrent que plusieurs domaines semblent entretenir des liens privilégiés avec la physique des particules. À la différence du type de contextes précédent, l'utilité ou l'utilisation ne sont pas marquées et les contextes évoquent plutôt une certaine interdisciplinarité dans la recherche en physique des particules et dans les autres domaines identifiés. Ce type de contexte est repéré à l'aide des unités des catégories « spatial », « arts et culture » et « énergie nucléaire ».

Différents phénomènes peuvent être observés pour les unités classées dans la catégorie « spatial ». Ainsi, l'exemple 27 montre que la recherche de la matière noire s'effectue par le biais d'un télescope, qui est installé dans un observatoire, ce qui indique que la matière noire est un sujet de recherche en astronomie. Néanmoins, l'exemple 28 reproduit ci-dessous suggère que, si l'« énigme de la "matière sombre" » est « venue de l'astronomie » et est désormais étudiée au LHC, alors la recherche de la matière noire¹³⁷ est un sujet de recherche qui semble partagé par l'astronomie et la physique des particules.

D'une manière similaire, l'exemple 29 nous apprend que, pour détecter de l'antimatière, un détecteur doit être envoyé dans l'espace par une navette, ce qui laisse supposer certaines collaborations ou, dans tous les cas, une certaine interdisciplinarité dans la recherche en physique des particules. Cette hypothèse a par ailleurs été confirmée par une experte qui, lors d'un entretien, nous a expliqué que l'étude de l'antimatière s'effectue en l'observant soit en laboratoire par le biais de simulations, soit dans l'espace, là où elle se trouve « naturellement ». De la même manière, d'après l'experte, il est possible d'observer des interactions entre particules telles qu'elles se produisent « naturellement » dans l'espace ou de les produire *via* les accélérateurs de particules, dans le cadre de collisions et de désintégrations, par exemple. C'est par ailleurs ce que montrent les exemples 30 à 32, où il est explicité que l'annihilation entre des électrons et des antiélectrons a été observée par un satellite et que les photons et les neutrinos sont créés par des réactions nucléaires dans les étoiles.

27. Le but premier d'OGLE, dont le télescope est installé à l'*observatoire* de Las Campanas au Chili, est la recherche de **matière noire** ; mais ce programme a permis d'autres découvertes inattendues comme des planètes extrasolaires. (SC-P, Le Figaro, 2016)
28. Et si les particules prédites par la **supersymétrie** étaient celles qui constituent cette "**matière sombre**", dont personne n'a pu comprendre la nature ? Une chance, cette énigme de la "matière sombre" venue de l'*astronomie*, pour le **LHC**. Elle permet d'étoffer le cahier des charges de la machine, elle lui ouvre la possibilité de répondre à

¹³⁷ La dénomination *matière sombre* est une variante du terme *matière noire*.

- une question qui ne se posait même pas quand son chantier a été ouvert. (SC-P, Le Monde, 2010)
29. Vers des phénomènes jusque-là indétectables : un volumineux détecteur d'**antimatière**, qui attend son tour depuis des années, va enfin prendre place dans le dernier vol de *navette*. (SC-P, Le Monde, 2010)
 30. De plus, le *satellite* a observé, dans la zone centrale de la *Voie lactée*, l'annihilation entre des **électrons** et leurs **antiparticules**, les **antiélectrons**. (SC-P, Le Temps, 2007)
 31. Dans le Soleil, **photons** et **neutrinos** sont créés au cours des réactions nucléaires qui ont lieu au cœur de notre *astre*. (SC-P, Le Temps, 2013)
 32. Le 23 février 1987, ils réussirent à détecter une averse de **neutrinos** issus de l'explosion d'une *supernova* éloignée. (SC-P, Les Échos, 2003)

Un autre cas de figure se présente dans des contextes où un terme de physique des particules et des unités qui renvoient au domaine de l'énergie nucléaire sont attestés. Dans les exemples ci-dessous, le terme *neutron* cooccure non seulement avec les unités *combustible* et *Iter*¹³⁸, de la catégorie « énergie nucléaire », mais également avec d'autres unités que l'on peut associer à l'énergie nucléaire, comme *uranium* (exemple 33) ou *réacteur* (exemple 35) par exemple. Ces contextes semblent ainsi suggérer que le terme *neutron* occupe une place particulière en énergie nucléaire¹³⁹.

33. En permanence, la fission d'un noyau d'uranium dégage un **neutron** qui part lui-même casser un autre noyau de *combustible*. (SC-P, Les Échos, 2007)
34. Deux isotopes de l'hydrogène y sont introduits (deuterium et tritium). Comme dans les étoiles, leur fusion produit un atome d'hélium et un **neutron** très énergétique. C'est le principe d'*Iter* dont le maillon faible est la densité peu élevée de *combustible* dans le milieu. (SC-P, Les Échos, 2010)
35. L'hélium 3 fait aussi rêver nombre d'ingénieurs nucléaires : n'émettant aucun **neutron**, il serait le *combustible* idéal d'un futur réacteur à fusion. (SC-P, Le Monde, 2004)

Cette idée de recherche faisant intervenir différentes disciplines se manifeste cependant différemment selon le domaine auquel renvoient les unités qui cooccurrent avec les termes à l'étude. Par exemple, lorsqu'*accélérateur de particules* et/ou *synchrotron* cooccurrent avec *Louvre*, comme dans les deux exemples ci-dessous, il semble que les contextes dénotent l'utilisation de ces instruments, issus de la physique des particules, pour la recherche sur les œuvres d'art, comme nous l'avons par ailleurs constaté *supra* avec les exemples 36 et 37, mais sans que l'on puisse observer de marqueurs liés à cette utilisation. Il semble alors bien que certains concepts de physique des particules profitent à la recherche dans d'autres disciplines.

¹³⁸ International Thermonuclear Experimental Reactor, <https://www.iter.org/> (page consultée le 11 août 2020).

¹³⁹ Ces exemples renvoient d'ailleurs aux liens très étroits existant entre les domaines de la physique des particules, de la physique nucléaire et de la physique atomique, comme nous l'expliquons au Chapitre 4 (4.1.2).

36. Les enseignements passionnants tirés de l'analyse des œuvres par l'**accélérateur de particules** situé sous le Carrousel du *Louvre* plaident en faveur d'un nouveau projet. (SC-P, Les Échos, 2009)
37. Aussi prometteurs soient-ils, les instruments portables n'ont pas supplanté les grandes installations d'imagerie hautement performante, comme l'**accélérateur de particules** Aglaé dans les entrailles du *Louvre* ou les **synchrotrons**. (SC-P, Le Temps, 2014)

Au vu de ces différents exemples, il semble par ailleurs que nous puissions distinguer au moins deux phénomènes selon la nature du concept impliqué et le domaine auquel renvoient les unités attestées. Dans le cas de termes qui désignent des techniques ou des instruments, comme *accélérateur de particules* et *synchrotron* dans les exemples 36 et 37, l'observation des contextes dans le SC-P montre que ces techniques ou ces instruments sont exploités pour la recherche dans d'autres domaines que la physique des particules, et notamment ici dans la recherche sur les œuvres d'art. En revanche, lorsque les termes désignent des objets d'étude de la physique des particules, comme les composants de la matière (*matière noire, antimatière*), des particules ou des types de particules (*électron, antiparticule, photon, neutrino, neutron*) ou encore des théories (*supersymétrie*), les contextes du SC-P semblent indiquer que l'étude de ces objets peut être appréhendée par plusieurs domaines. Il peut s'agir de cas où la recherche est interdisciplinaire ou de cas où un terme occupe une place centrale dans différents domaines connexes.

6.2.3. Contextes insistant sur des aspects financiers

Certains contextes, où des termes de physique des particules cooccurrent avec des unités classées dans la catégorie « économie et finances », renvoient en particulier aux coûts engendrés par la recherche en physique des particules, notamment lorsqu'il est question de la construction des instruments dédiés à la recherche. Ce cas est tout à fait particulier en comparaison avec les cas cités *supra*, car il ne s'agit pas ici de constater une certaine interdisciplinarité dans la recherche ou l'utilité de certains concepts de physique des particules pour d'autres domaines. Nous constatons plutôt que toute une thématique liée à la recherche en physique des particules, qui relève de l'aspect financier, est récurrente dans le SC-P en comparaison avec le SC-S. Ainsi, dans les exemples 38 à 42, nous observons la cooccurrence des termes *LHC*, *CERN* et *boson de Higgs* avec les unités *budgétaire*, *euro*, *coûter*, *coût*, *investir*, *dollar* et *financer*.

38. Cette rallonge *budgétaire* s'ajoute aux 26 millions d'*euros* que le **CERN** a dépensé dans la réparation du **LHC** cette année. (SC-P, Les Échos, 2009)
39. L'enjeu de cet énorme effort scientifique et technologique, qui *coûtera* au total plus de 3 milliards d'*euros*, est une mystérieuse particule, le **boson de Higgs**. (SC-P, Le Figaro, 2004)

40. Les partenaires industriels du **CERN** ont promis d'*investir* 1,5 million de *dollars* chacun sur trois ans. (SC-P, Le Temps, 2003)
41. Et c'est lui [le LHC] qui, promet-on, devrait identifier le fameux **boson de Higgs**. Son *coût* : 3,9 milliards d'*euros*. (SC-P, Le Figaro, 2010)
42. L'investissement est de 3,7 milliards d'*euros*, dont 2,8 milliards pour l'accélérateur et 0,7 milliard pour les détecteurs. Les vingt Etats européens membres du **CERN** en *financent* 90 %, d'autres pays 10 %. (SC-P, Le Monde, 2008)

6.2.4. Contextes mettant en évidence des aspects culturels

Enfin, d'autres contextes, repérés à l'aide d'unités classées dans la catégorie « arts et culture », indiquent que certains termes de physique des particules s'intègrent dans des contextes littéraires ou artistiques.

43. Comment croire à l'avion hypersonique du **Cern**, fusée qui amène en 64 minutes le héros du *livre* de Boston à Genève? (SC-P, Le Temps, 2005)
44. Cette pièce pour six danseurs et quatre lampes géantes en rotation s'inscrit dans un programme culturel du **CERN**, *Collide@CERN*, qui crée des ponts entre sciences et *art* et permet à un artiste confirmé de consacrer plusieurs mois à la recherche fondamentale sans impératif de production. (SC-P, Le Temps, 2013)

Ces deux exemples illustrent par ailleurs des aspects différents : dans l'exemple 43, il semble que le CERN ainsi que les applications (imaginaires) des recherches qui y sont menées (un « avion hypersonique » ici) contribuent au développement de l'intrigue d'un roman, tandis que l'exemple 44 traite d'un véritable « programme culturel du CERN » qui permet à des artistes de développer des projets (de danse ici) dans le contexte des activités du CERN.

6.2.5. Contextes évoquant des aspects extraordinaires

Certains contextes, où les termes à l'étude cooccurrent avec des unités de la catégorie « sensationnalisme », mettent en évidence plusieurs aspects extraordinaires ou hors norme de la recherche en physique des particules. Par exemple, la cooccurrence des unités *immense* et *énorme* avec les termes *accélérateur de particules*, *CERN*, *proton* et *particule élémentaire* dans les exemples 45 et 46 semble attester la taille hors norme des équipements dédiés à la recherche en physique des particules. En revanche, dans les exemples 47 et 48, les adjectifs *faramineux* et *phénoménal*, qui modifient les noms *énergie* et *vitesse*, semblent plutôt démontrer la puissance extraordinaire des équipements.

45. Cet *immense* **accélérateur de particules** de 27 km de circonférence est installé au **CERN**, près de Genève. (SC-P, Le Temps, 2008)

46. A plein régime, celle du **LHC** devrait produire 600 millions de collisions par seconde. Elles auront lieu quatre fois par tour, au sein d'*énormes* détecteurs capables de reconnaître chacune des **particules élémentaires** ainsi libérées. (SC-P, Le Monde, 2007)
47. En juillet, dans l'anneau de 27 km du **LHC**, les scientifiques ont lancé les uns contre les autres des paquets de particules (**protons**) et avec une énergie *farouche* (3,5 TeV). (SC-P, Le Temps, 2010)
48. Une semaine plus tard, les équipes du Centre européen de recherches nucléaires (**Cern**), qui exploite le **LHC**, parviennent à injecter dans son anneau de 27 kilomètres de circonférence des faisceaux de protons lancés à des vitesses *phénoménales* jusqu'à atteindre une énergie encore jamais atteinte de 1,18 téraélectronvolts (TeV). (SC-P, Le Figaro, 2009)

6.2.6. Contextes évoquant des aspects mystérieux

Lorsque les termes cooccurrent avec d'autres unités de la catégorie « sensationnalisme », telles que *mystère*, *secret*, *mystérieux* ou encore *fantomatique*, il semble que les contextes contribuent à entretenir une certaine part de mystère autour de la recherche en physique des particules, et notamment lorsque les termes *boson de Higgs*, *matière noire*, *neutrino*, *antimatière*, *proton*, *quark* ou encore *interaction forte* sont impliqués. C'est ce qu'illustrent les exemples suivants.

49. En brisant des particules, les physiciens espèrent enfin capter l'empreinte du *fantomatique* **boson de Higgs**. (SC-P, Le Monde, 2008)
50. Les premières découvertes, notamment celles du fameux boson de Higgs ou de la *mystérieuse* **matière noire** qui représente 23 % de la masse de l'Univers, ne sont toutefois pas attendues avant la fin de 2009. (SC-P, Le Figaro, 2008)
51. Les chercheurs n'ont pas fini pour autant d'expliquer tous les *mystères* qui entourent l'**antimatière**. (SC-P, Les Échos, 2008)
52. Ces particules fondamentales qui composent les **protons** et les **neutrons** sont au cœur des *mystères* de l'**interaction forte**, l'une des quatre forces fondamentales de notre univers avec la gravité, la force électromagnétique et l'**interaction faible**. (SC-P, Les Échos, 2004)
53. Le *secret* de la **matière noire** sera peut-être levé cet été, grâce à un instrument lancé par un des derniers vols d'une navette spatiale américaine. (SC-P, Les Échos, 2010)
54. Mais c'est en fait un des rares moyens dont disposent les physiciens pour essayer de percer le *mystère* du **proton** (SC-P, Le Figaro, 2003)
55. Car le *secret* des **quarks** s'avère très difficile à percer. (SC-P, Le Figaro, 2003)

6.2.7. Contextes dénotant des emplois métaphoriques des termes

Le dernier type de contexte observé à l'aide des *keywords* est un peu différent et concerne des cas où les termes de physique des particules sont utilisés de manière métaphorique. Ces derniers cas sont particulièrement difficiles à repérer, car les métaphores ne sont pas récurrentes et la cooccurrence des termes avec les unités analysées ne peut pas être systématiquement associée à un cas de métaphore. Ces emplois sont repérés aussi bien à l'aide des unités classées dans les catégories

évoquant d'autres domaines que la physique des particules qu'à l'aide des unités de la catégorie « sensationnalisme ». Ainsi, dans les exemples 56 et 57, les unités *art* et *livre* de la catégorie « arts et culture » permettent de repérer des usages métaphoriques des termes *matière noire*, *anti-matière* ; dans l'exemple 58, l'unité *économique* de la catégorie « économie et finances » permet de repérer un usage métaphorique d'*accélérateur de particules* ; dans les exemples 59 et 60, ce sont les unités *mystère* et *mystérieux* qui mettent en évidence les usages métaphoriques du terme *matière noire*. Notons que les guillemets dans l'exemple 60 constituent un indice supplémentaire en ce sens.

56. Aux antipodes des white boxes de règle dans l'*art* contemporain, ce sauvage a préféré la "**matière noire**" de la nuit. (SC-P, Le Figaro, 2012)
57. Quand on écrit des nouvelles, il y a cet avant et cet après que l'on dit pas mais que l'on suppose. Ce n'est pas écrit, et pourtant cela compte : c'est la **matière noire** de la nouvelle. Son **anti-matière**. Sa narration invisible. Dans mon *livre*, je voulais que l'enchaînement des chapitres rende visibles cet avant et cet après. (SC-P, Le Monde, 2009)
58. "Elle [la Bourse] n'est plus envisagée comme un **accélérateur de particules économiques** pour les pays développés, mais comme un péril pour les économies émergentes qui en dépendent", ajoute Fidelity. (SC-P, Les Échos, 2016)
59. A la réflexion, le vertige vient d'ailleurs. De ce que l'on ne raconte pas, d'un vide et d'un *mystère* que le récit évite, d'une **matière noire** dont il fait son centre de gravité. (SC-P, Le Monde, 2010)
60. La "**matière noire**" du génome, pas si *mystérieuse* (SC-P, Le Monde, 2010)

Par ailleurs, les termes *matière noire* et *accélérateur de particules* sont particulièrement susceptibles de donner lieu à des métaphores, comme le montrent les exemples 61 à 63. Il semble ainsi qu'une autre hypothèse puisse être avancée à ce stade quant à la récurrence d'usages métaphoriques pour ces deux termes, qui concerne plus spécifiquement le rôle de la presse dans le processus de déterminologisation. Nous l'avons dit *supra* (6.1.1.1), les journalistes sont tenus de capter l'attention de leurs lecteurs et sont encouragés à utiliser un langage imagé lorsque celui-ci est approprié. De ce point de vue, il est donc possible que les occurrences où ces termes sont utilisés de manière métaphorique relèvent d'une volonté de la part des journalistes de mettre en œuvre un style plus imagé par le recours à des métaphores, dans le but de maintenir l'intérêt des lecteurs. Nous reviendrons sur ce point dans le Chapitre 8.

61. On peut parler d'une "**matière noire**" de la biodiversité, qu'on commence tout juste à explorer grâce à la génomique. (SC-P, Les Échos, 2009)
62. Aujourd'hui, on vit dans un **accélérateur de particules** avec toutes ces informations, ces distances raccourcies par les avions et les trains. (SC-P, Le Figaro, 2013)
63. **Accélérateur de particules**, animateur sur le terrain et en coulisses, Ribéry s'impose de plus en plus comme le véritable leader de l'équipe de France. (SC-P, Le Figaro, 2009)

Les différents types de contextes repérés dans cette section et les exemples proposés permettent d'illustrer différents cas de figure identifiés par la cooccurrence de termes de physique des particules avec les unités repérées par une extraction de *keywords* et classées dans différentes catégories. Ces contextes permettent également de rendre compte de la diversité des fonctionnements des termes dans le SC-P.

En effet, nous avons observé des contextes :

- où l'utilité de certains concepts de physique des particules pour d'autres domaines est explicitée ;
- où les liens que différents domaines entretiennent avec la physique des particules sont précisés, qu'il s'agisse d'interdisciplinarité, de collaborations, ou encore de la centralité d'un terme dans plusieurs domaines connexes ;
- illustrant l'importance de la thématique financière autour de la recherche en physique des particules dans le SC-P ;
- où les termes de physique des particules s'intègrent dans des contextes artistiques ou littéraires ;
- insistant sur plusieurs aspects extraordinaires et hors norme de la recherche en physique des particules ;
- mettant en évidence la part de mystère et de fascination qui entoure la recherche en physique des particules ;
- attestant plusieurs emplois métaphoriques des termes.

À la lueur de ces éléments, il semble donc que les manifestations de la déterminologisation sont multiples dans ce sous-corpus. D'un point de vue sémantique, il nous semble évident que les occurrences où les termes de physique des particules relèvent d'un usage métaphorique contribuent aux changements sémantiques dus à la déterminologisation. Il s'agit par ailleurs de la conséquence la plus connue de ce phénomène, discutée notamment par Meyer et Mackintosh (2000a, 2000b ; cf. Chapitre 2, 2.1.1.3). Néanmoins, il est possible que les autres types de contextes jouent également un rôle dans les changements sémantiques résultant du processus de déterminologisation. Par exemple, nous pouvons faire l'hypothèse à ce stade que la cooccurrence récurrente de termes de physique des particules avec des unités renvoyant aux aspects financiers de la recherche participe à la création de l'image d'une recherche qui coûte cher à la société. Il semble en effet que la mention récurrente d'importantes sommes, dans les contextes du terme *LHC* notamment, peut participer à

l'ajout de certaines connotations dans le sens de ce terme qui est véhiculé dans le SC-P. De la même manière, il est possible que les contextes de ce terme mettant en avant les aspects extraordinaires, hors norme, voire merveilleux, de la recherche dans ce domaine, ou explicitant l'utilité des accélérateurs de particules pour d'autres domaines, contribuent à véhiculer une image plus positive pour la société.

Ces hypothèses renvoient par ailleurs à un des aspects caractérisant le premier type de changement sémantique lié à la déterminologisation discuté par Meyer et Mackintosh (2000b : 115). Les autrices expliquent en effet que « the layperson's understanding of a term may be coloured by connotations that were not present in the experts' original understanding » et donnent notamment l'exemple du terme *atomic* qui a tendance à être associé à Hiroshima et au terme *atomic bomb* et non au sens « terminologique » relatif au terme *atom* (*Ibid.*). Dans cette perspective, il est possible que la diversité des contextes observés dans le SC-P favorise la diffusion de connotations qui sont absentes du SC-S et qui contribuent à la construction du sens des termes concernés dans un contexte non spécialisé.

Afin d'étayer ces hypothèses et d'examiner de quelles manières ces aspects se manifestent dans les sous-corpus intermédiaires, nous détaillons nos observations dans ces sous-corpus dans la section suivante.

6.3. Exploration des *keywords* dans les sous-corpus intermédiaires

Dans cette section, nous cherchons à interroger le rôle des sous-corpus intermédiaires dans le processus de déterminologisation, à partir des fonctionnements des termes repérés dans le SC-P à l'aide de l'exploration des *keywords* et détaillés dans la section précédente. Nous cherchons ainsi, d'une part, à vérifier si les unités classées dans différentes catégories sont également attestées dans les sous-corpus intermédiaires et, d'autre part, à observer ce que les contextes de ces unités peuvent le cas échéant nous apprendre sur le rôle des intermédiaires dans le processus de déterminologisation.

6.3.1. Répartition des unités des catégories dans les sous-corpus

Afin de vérifier si les unités considérées comme des *keywords* dans le SC-P sont véritablement caractéristiques de ce sous-corpus, nous observons dans un premier temps si elles

sont attestées dans les sous-corpus intermédiaires. Dans un second temps, nous comparons les fréquences des unités attestées dans ces sous-corpus avec leur fréquence dans le SC-P¹⁴⁰.

La répartition des unités des catégories dans les sous-corpus montre que, sur l'ensemble des unités, seules trois d'entre elles ne sont attestées dans aucun des sous-corpus intermédiaires (*Arlésienne*, *AAA*, *Louvre* ; nous y reviendrons *infra*). Cela signifie que 95,8 % des unités apparaissent au moins une fois dans au moins un des trois sous-corpus intermédiaires, suggérant ainsi qu'elles ne sont pas tout à fait propres au SC-P uniquement.

En outre, il est intéressant d'observer le nombre d'unités attestées par catégorie dans chacun des trois sous-corpus. Les graphiques présentés dans la figure 6.1 ci-dessous¹⁴¹ représentent le nombre d'unités apparaissant au moins une fois dans chaque sous-corpus intermédiaire (indiqué dans chaque barre correspondante), par rapport au nombre d'unités de chaque catégorie (indiqué entre parenthèses dans le titre de chaque graphique).

¹⁴⁰ La fréquence des unités dans les sous-corpus intermédiaires a été calculée automatiquement avec R et les packages *gsubfn* (Grothendieck, 2018) et *languageR* (Baayen et Shafei-Bajestan, 2019), à partir des versions étiquetées et lemmatisées de ces sous-corpus.

¹⁴¹ Ces graphiques ont été réalisés avec R, avec le package *ggplot2* (Wickham, 2009 ; Wickham *et al.*, 2020a).

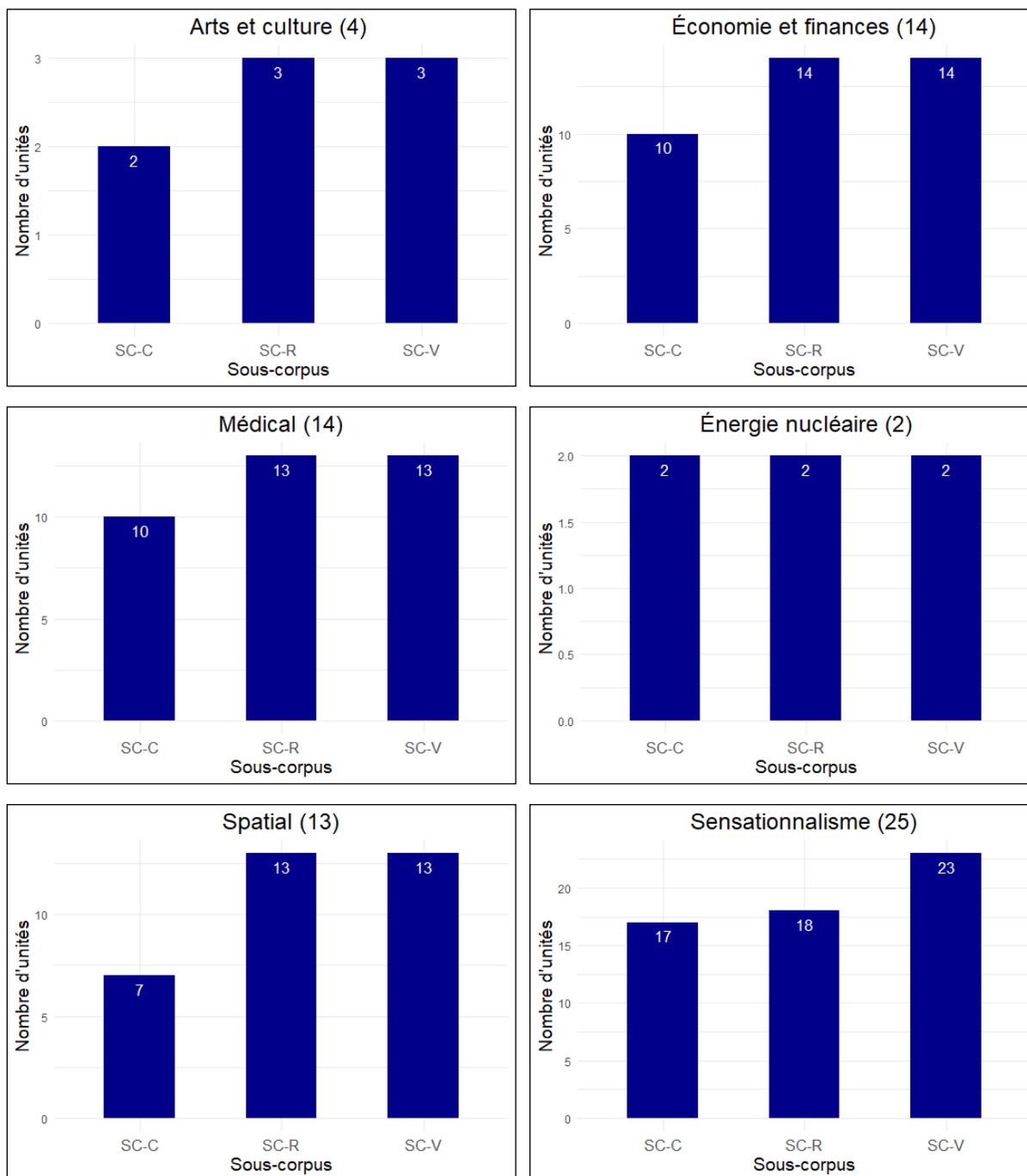


Figure 6.1 : Nombre d'unités attestées dans les sous-corpus intermédiaires, par catégorie

Selon nous, l'intérêt de ces graphiques réside dans la possibilité de comparer aisément les données à la fois selon les différentes catégories et selon les sous-corpus. Ainsi, ces graphiques mettent en évidence non seulement le fait que, de manière générale, le SC-C est le sous-corpus dans lequel le moins d'unités sont attestées, mais également qu'à l'exception des unités de la catégorie « sensationnalisme », presque toutes les unités de chaque catégorie sont attestées dans les SC-R et SC-V.

Cette information n'est toutefois pas suffisante, elle doit être complétée par l'observation des fréquences de ces unités dans les sous-corpus intermédiaires. Afin de permettre des comparaisons avec le SC-P, nous illustrons d'abord dans la figure 6.2 la fréquence relative de l'ensemble des unités de chaque catégorie dans le SC-P et dans les sous-corpus intermédiaires, puis, dans les figures 6.3 et 6.4, la fréquence relative des unités des catégories dans ces sous-corpus¹⁴².

La figure 6.2 permet non seulement de comparer les fréquences des unités des catégories dans le SC-P et dans les sous-corpus intermédiaires (indiquées dans les barres correspondantes), mais également d'identifier quelles catégories sont les plus représentées dans ces sous-corpus. Ainsi, si les fréquences relatives des unités des catégories « arts et culture » et « économie et finances » sont plutôt faibles, en comparaison avec le SC-P, et stables, ce n'est pas le cas pour toutes les autres catégories. Par exemple, selon le sous-corpus, nous remarquons que certaines catégories sont particulièrement présentes, comme la catégorie « médical » dans le SC-R, ou encore les catégories « énergie nucléaire » et, dans une moindre mesure, « spatial » dans le SC-V. Nous émettons toutefois quelques réserves et interprétons les données pour la catégorie « énergie nucléaire » avec prudence, car celle-ci n'est composée que de deux unités, ce qui reste difficilement généralisable.

De la même manière, bien que le SC-C soit le sous-corpus dans lequel le moins d'unités de la catégorie « sensationnalisme » sont attestées, il s'agit du sous-corpus intermédiaire où la fréquence relative de ces unités est la plus élevée, suivi de près par le SC-V.

¹⁴² Les fréquences relatives sont calculées sur 100 000 occurrences.

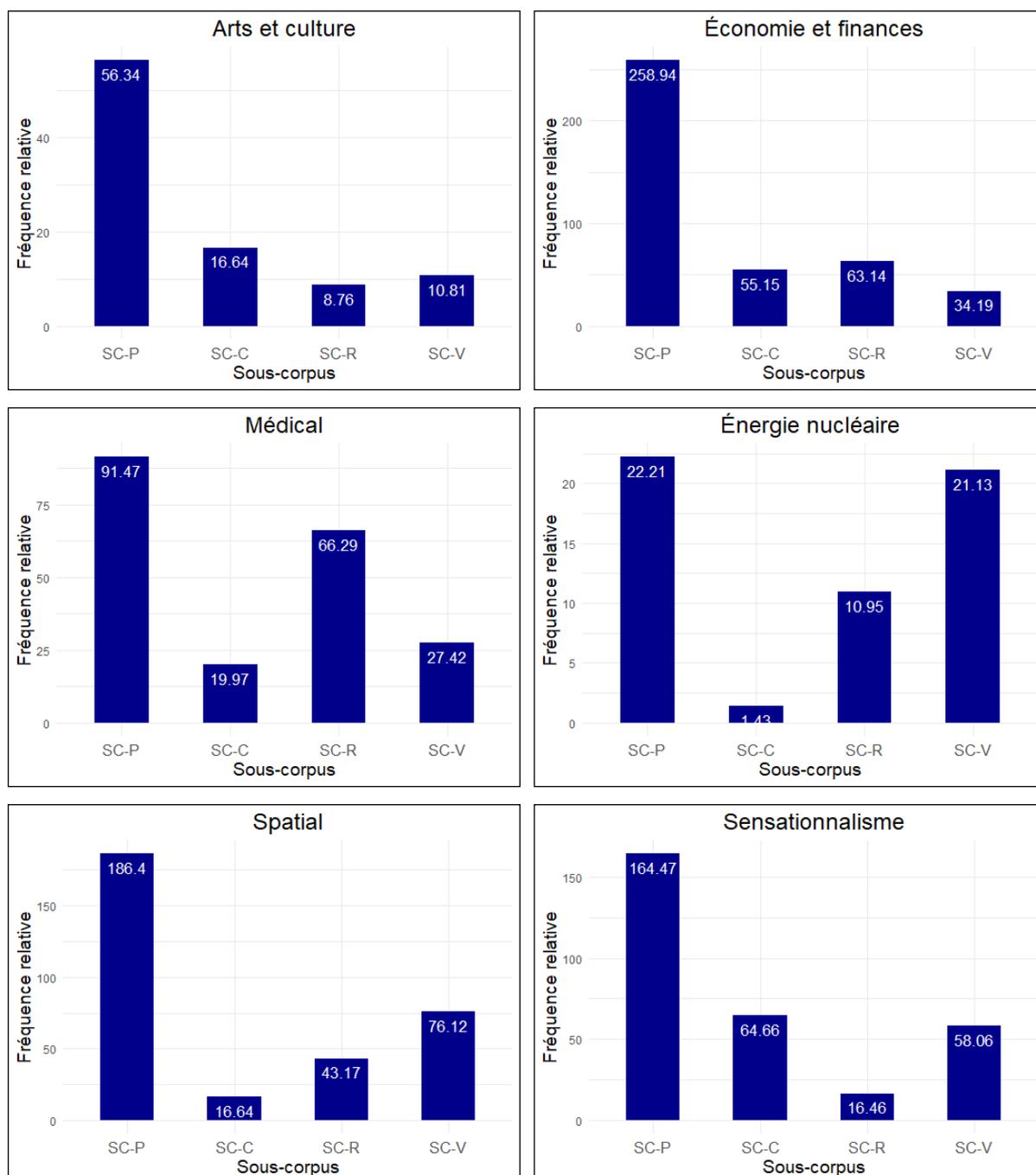


Figure 6.2 : Fréquence relative de l'ensemble des unités de chaque catégorie, dans les sous-corpus intermédiaires et le SC-P

À ce stade, il semble donc que les unités renvoyant à d'autres domaines que la physique des particules sont plutôt bien réparties dans les sous-corpus intermédiaires, avec plusieurs différences selon les catégories, alors que ces sous-corpus sont composés de textes relevant de la physique des particules. En outre, bien que la notion de sensationnalisme soit généralement attribuée au style de l'écriture journalistique, les unités de cette catégorie ne sont pas aussi caractéristiques du SC-P que ce que nous avons supposé. Ces premières explorations dans les sous-corpus intermédiaires permettent alors de nuancer les constats effectués en 6.2 et montrent, globalement, que les

catégories retenues sont présentes dans les sous-corpus intermédiaires, avec une fréquence toutefois plus faible que dans le SC-P.

Dans le but d'affiner ces observations, nous tenons également compte de la fréquence de chaque unité dans les sous-corpus intermédiaires et dans le SC-P. Ainsi, la figure 6.3 illustre la fréquence relative des unités de la catégorie « sensationnalisme » dans les sous-corpus intermédiaires et dans le SC-P et la figure 6.4 la fréquence relative des unités de la catégorie « médical »¹⁴³. Pour chaque unité, les quatre sous-corpus considérés sont représentés sur l'axe des abscisses et la fréquence relative est illustrée sur l'axe des ordonnées. Précisons toutefois que l'ordre des sous-corpus sur l'axe des abscisses ne reflète pas un ordre précis ou une hypothèse de notre part. Il s'agit plutôt de présenter les données sous une forme qui rend leur interprétation plus aisée. Ainsi, le SC-P est placé en première position afin de faciliter les comparaisons. Par ailleurs, vu la taille de chaque graphique, seuls les graphiques générés pour les catégories « sensationnalisme » et « médical » sont reproduits dans cette section.

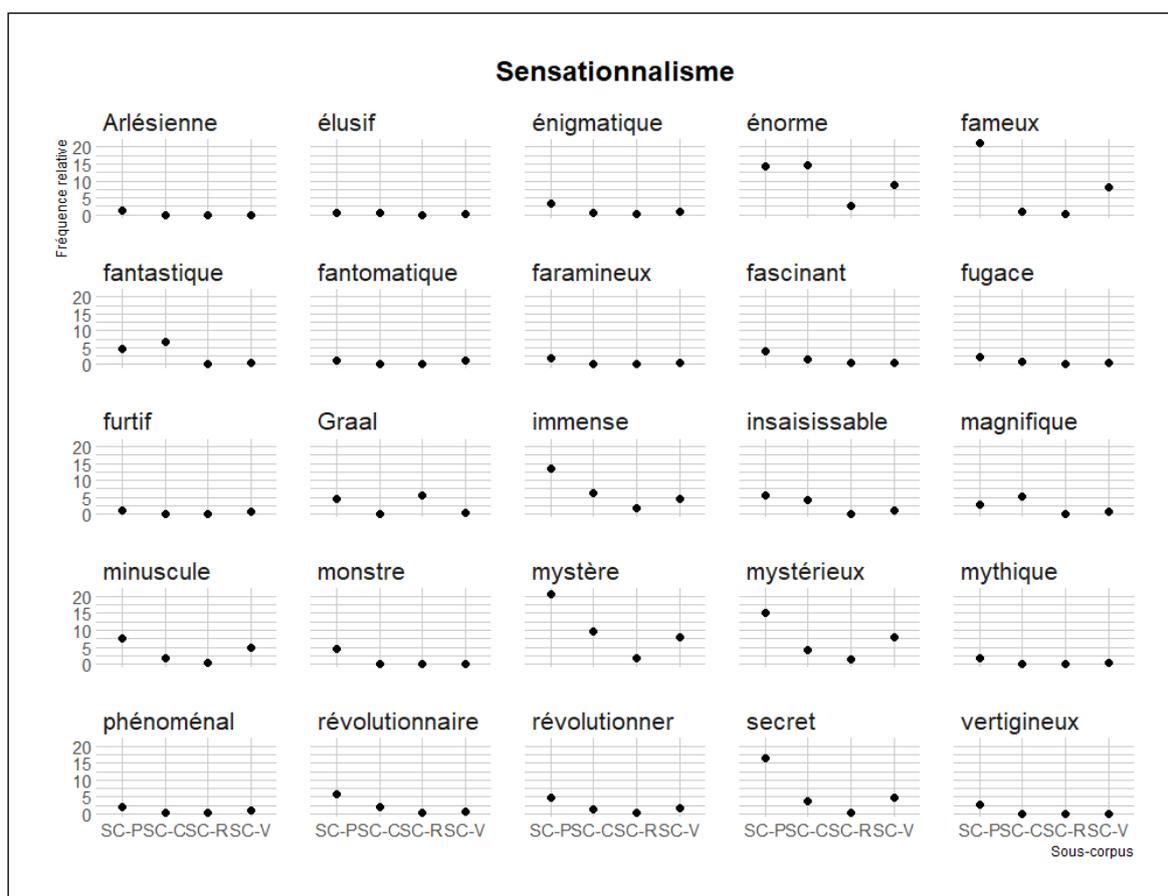


Figure 6.3 : Fréquence relative des unités de la catégorie « sensationnalisme » dans les sous-corpus intermédiaires et dans le SC-P

¹⁴³ Ces graphiques ont été générés avec R, avec les packages *ggplot2* (Wickham, 2009 ; Wickham *et al.*, 2020a), *dplyr* (Wickham *et al.* 2020b) et *hrbrthemes* (Rudis *et al.*, 2020).

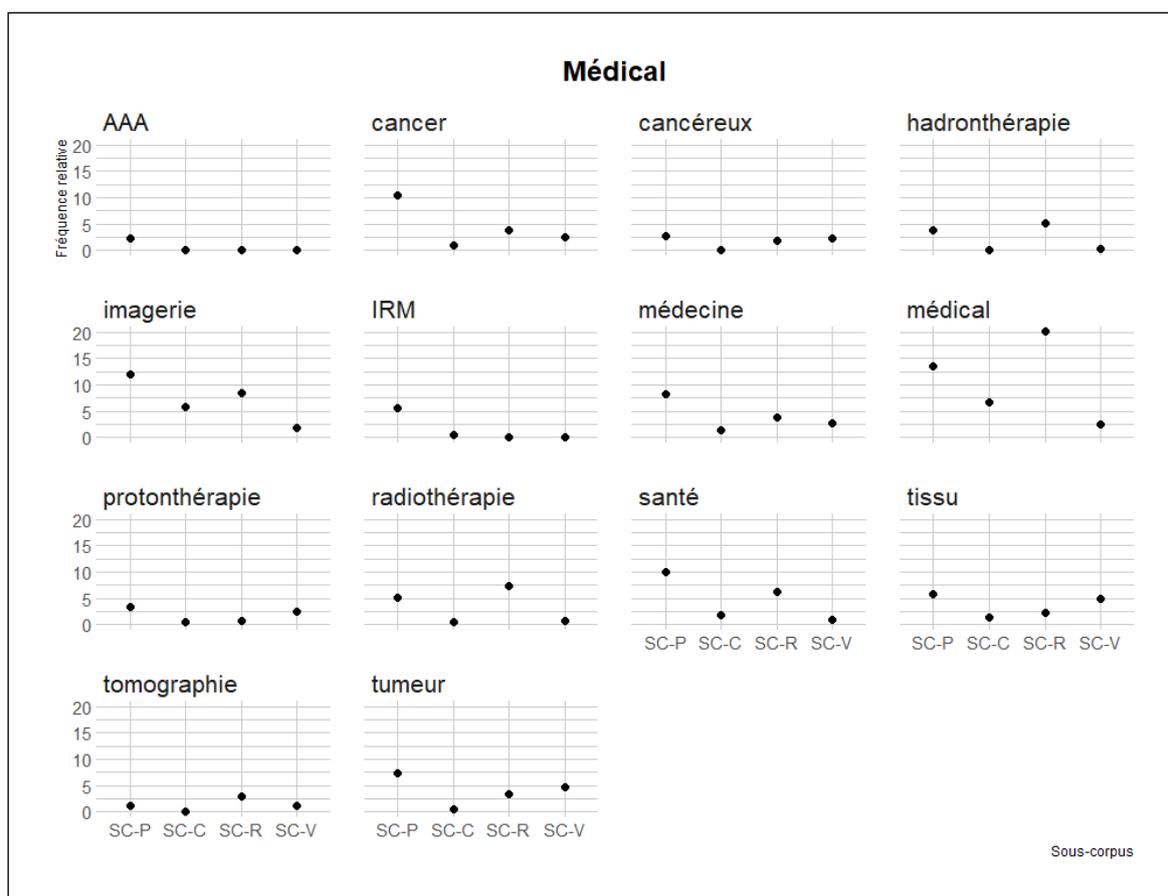


Figure 6.4 : Fréquence relative des unités de la catégorie « médical » dans les sous-corpus intermédiaires et dans le SC-P

Selon nous, cette méthode de visualisation comporte au moins deux avantages. D’abord, ce type de graphique permet de comparer dans le même temps les fréquences relatives des unités les unes par rapport aux autres et selon chacun des sous-corpus. Ensuite, nous pensons qu’il est beaucoup plus aisé de repérer certaines différences de fréquence à l’aide de ces graphiques plutôt que par l’observation de tableaux de fréquence, parfois longs, où les données numériques s’enchaînent. L’une de ces différences concerne les unités dont la fréquence relative est plus élevée dans au moins un sous-corpus intermédiaire que dans le SC-P. Ainsi, pour la catégorie « sensationnalisme », la fréquence relative des unités *énorme*, *fantastique* et *magnifique* est plus élevée dans le SC-C que dans le SC-P, celle de l’unité *fantomatique* est légèrement plus élevée dans le SC-V que dans le SC-P et enfin la fréquence relative de *Graal* est plus élevée dans le SC-R que dans le SC-P¹⁴⁴. Pour la catégorie « médical », les unités *hadronthérapie*, *médical*, *radiothérapie* et *tomographie* ont une fréquence relative plus élevée dans le SC-R que dans le SC-P.

¹⁴⁴ Précisons cependant qu’il s’agit ici de l’acronyme d’une collaboration (Grenoble Anneau Accélérateur Laser, <http://www.lnf.infn.it/~levisand/graal/graal.html>, page consultée le 14 septembre 2020) qui a impliqué l’un des laboratoires dont les rapports ont été inclus dans le SC-R. Cet acronyme apparaît sous la forme *GRAAL* dans le sous-corpus, mais a été lemmatisé en *Graal* ou *graal*. Il s’agit donc d’une erreur due

À partir de ces éléments, nous pensons que l'analyse doit être poursuivie sur ces unités en priorité, d'abord car leur fréquence plutôt élevée garantit qu'un nombre suffisant d'occurrences pourra être analysé, et ensuite car le fait que ces unités sont plus fréquentes dans un sous-corpus intermédiaire par rapport au SC-P semble particulièrement intéressant. En effet, ces unités ont à l'origine été repérées car elles se distinguent dans une large mesure dans le SC-P en comparaison avec le SC-S de par leur fréquence. Nous avons alors ainsi émis l'hypothèse que ces unités sont caractéristiques de ce sous-corpus. Or, étant donné que la fréquence relative de certaines de ces unités est plus élevée dans au moins un sous-corpus intermédiaire par rapport au SC-P, cette première hypothèse ne semble pas entièrement valide et il faut observer les contextes d'apparition de ces unités dans les sous-corpus intermédiaires pour mieux saisir les phénomènes en jeu. Par exemple, à ce stade nous supposons que la fréquence relative plus élevée de quatre unités de la catégorie « médical » dans le SC-R en comparaison avec le SC-P n'est pas un phénomène anodin, mais pourrait révéler des aspects importants et intéressants à propos du processus de déterminologisation (ce que nous montrons *infra*, 6.3.2.2).

Enfin, précisons que, pour les autres catégories, les unités ayant une fréquence relative plus élevée dans au moins un sous-corpus intermédiaire par rapport au SC-P sont :

- *financement* et *financer* pour la catégorie « économie et finances » (dans les SC-R et SC-C respectivement) ;
- *combustible* pour la catégorie « énergie nucléaire » (dans les SC-R et SC-V) ;
- *supernova* pour la catégorie « spatial » (dans le SC-V)¹⁴⁵.

6.3.2. Prise en compte des contextes

Dans cette section, nous nous focalisons sur l'observation des contextes des unités retenues dans les sous-corpus intermédiaires, afin de consolider ou de nuancer les interprétations proposées à partir de l'exploration du SC-P et d'approfondir notre réflexion sur le processus de déterminologisation. Ainsi, nous montrons que la plupart des phénomènes identifiés dans le SC-P sont également attestés dans au moins un sous-corpus intermédiaire, avec toutefois certaines nuances, que nous précisons le cas échéant. Dans les sections qui suivent, nous décrivons d'abord les différents fonctionnements repérés, en relation avec les fonctionnements observés dans le SC-

à l'étiqueteur qui n'est pas pertinente pour notre analyse. Nous n'en tenons pas compte dans la suite de notre travail.

¹⁴⁵ Pour ce qui est de la catégorie « arts et culture », aucune unité n'est retenue, car la fréquence relative d'aucune de ces unités n'est plus élevée dans un sous-corpus intermédiaire par rapport au SC-P.

P, puis nous expliquons de quelle manière ces observations contribuent à alimenter notre réflexion sur la déterminologisation.

6.3.2.1. Importance de la thématique financière

Les contextes des unités *financement*, dans le SC-R, et *financer*, dans le SC-C, montrent que les questions financières occupent une place particulière dans ces deux sous-corpus, de la même manière que nous l'avons remarqué pour le SC-P. C'est ce qu'illustrent les exemples ci-dessous.

64. Une demande de *financement* a été soumise à l'Union européenne dans le contexte de son 6e programme-cadre pour ce projet, appelé "HIPPI" [injecteur de **protons** pulsé de haute intensité], qui s'inscrit dans l'activité coordonnée CARE [Recherche coordonnée sur les accélérateurs en Europe]. (SC-R, CERN, 2003)
65. Par ailleurs, au titre du 7e programme-cadre de l'Union européenne, lancé en 2007, le **CERN** a pu obtenir un *financement* pour plusieurs bourses Marie Curie et pour le projet SLHC, qui jouera un rôle majeur dans le relèvement de la luminosité du **LHC**. (SC-R, CERN, 2007)
66. [...] directeur du Bureau de la science du Ministère de l'énergie qui *finance* la contribution américaine aux aimants du **LHC**. (SC-C, CERN, 2003)
67. En tant que membre du **Cern**, la France participe pour 15 % au budget de l'organisation. Mais dans le cadre de sa "contribution exceptionnelle", en tant que pays hôte, la France a *financé* 16,33 % du LHC. (SC-C, IN2P3, 2013)

Cependant, si nous observons les contextes de ces unités plus en détail, nous pouvons remarquer deux différences intéressantes. D'une part, nous constatons que, dans les SC-C et SC-R, les coûts exacts sont plus rarement mentionnés que dans le SC-P. D'ailleurs, les unités *euro*, *franc* et *dollar* apparaissent toujours avec une fréquence relative plus faible (et même presque nulle) dans les SC-R et SC-C que dans le SC-P, comme l'illustre la figure 6.5. Notons que la différence est particulièrement frappante pour l'unité *euro*.

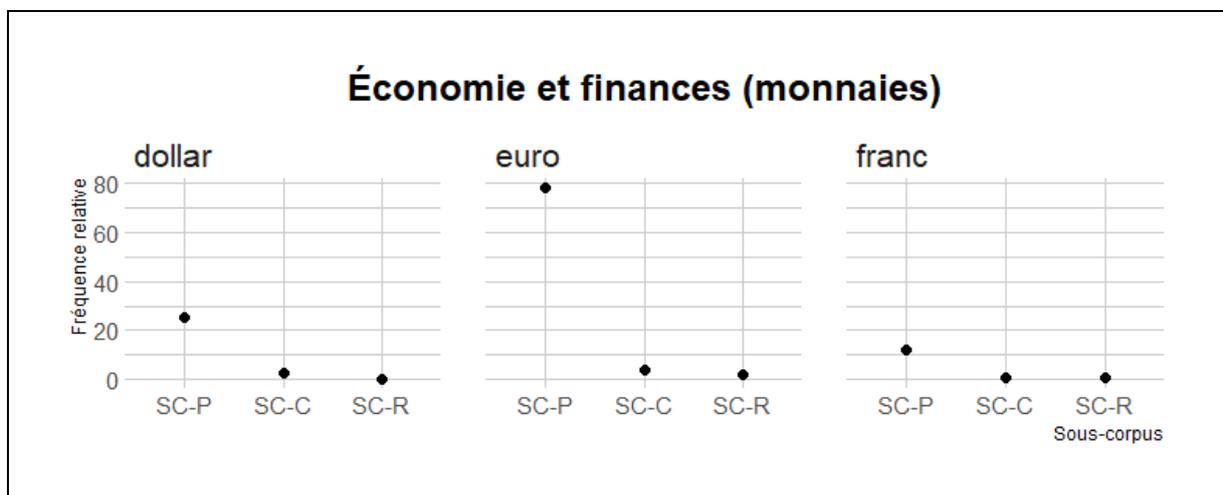


Figure 6.5 : Fréquence relative des unités *dollar*, *euro* et *franc* dans les SC-P, SC-C et SC-R

À ce stade, il semble que la mention des sommes exactes dépensées dans le cadre de la recherche en physique des particules soit un aspect plutôt caractéristique du SC-P¹⁴⁶. Cette observation est par ailleurs cohérente avec l'hypothèse selon laquelle, dans le SC-P, ces aspects financiers ont pour but d'attirer l'attention des lecteurs sur ces articles.

D'autre part, la seconde différence notable concerne l'objet de ces financements, qui se distingue dans le SC-R par rapport aux SC-P et SC-C. En effet, dans le SC-R, l'unité *financement* est plus généralement associée à des projets, à des bourses ou à des thèses, par exemple, comme le montrent les contextes 64 et 65, où cette unité cooccure avec les unités *bourse* et *projet*. En revanche, dans les SC-P et SC-C, il semble que les questions financières concernent plus précisément l'acquisition ou le développement des équipements ainsi que la construction des instruments, notamment le LHC. C'est ce que montrent l'exemple 67 tiré du SC-C, qui explicite la part du LHC qui a été financée par la France, et les exemples 38, 41 et 42 tirés du SC-P (cf. 6.2.3), qui explicitent entre autres les coûts de la construction et de la maintenance du LHC.

6.3.2.2. Interdisciplinarité de la recherche

Les contextes des quatre unités classées dans la catégorie « médical » dont la fréquence relative est plus élevée dans le SC-R que dans le SC-P nous permettent de renforcer notre hypothèse selon laquelle la physique des particules et le domaine médical sont effectivement étroitement liés. Différents indices sont observés dans ces contextes.

Tout d'abord, certains contextes mettent en évidence les collaborations entre ces domaines, voire l'interdisciplinarité de la recherche. Nous reproduisons ci-dessous deux contextes tirés du SC-R illustrant particulièrement bien ces liens entre ces deux domaines.

68. D'autres développements, décrits par ailleurs dans ce rapport d'activité, pourraient donner à l'avenir des actions de valorisation : la *tomographie*, la *hadronthérapie* avec le lancement du projet ETOILE à Lyon et certains développements dans l'énergie nucléaire... (SC-R, LPSC, 2005)
69. Les recherches du **CERN** poussent la technologie à ses limites, au profit de la société. [...] Des détecteurs conçus pour la physique des hautes énergies ont été adaptés pour des applications en imagerie *médicale*, notamment en *tomographie* par émission de **positons** (TEP). (SC-R, CERN, 2006)

Dans ces deux exemples, différents éléments permettent d'interpréter que la recherche en physique des particules s'effectue en collaboration avec le domaine médical, ou profite à ce

¹⁴⁶ Précisons que cet aspect n'est pas totalement absent des SC-C et SC-R, seulement qu'il apparaît de manière beaucoup moins récurrente que dans le SC-P.

domaine, notamment la mention d'« actions de valorisation » dans l'exemple 68 ou d'« applications en imagerie médicale » dans l'exemple 69.

Ensuite, dans d'autres contextes, l'attestation de termes complexes formés à la fois de termes de physique des particules faisant partie de l'échantillon analysé dans notre travail et d'unités identifiées comme *keywords* et classées dans la catégorie « médical » suggère que ces deux domaines entretiennent effectivement des liens privilégiés. Par exemple, le terme *radiothérapie par photons* est formé de *radiothérapie*, qui renvoie à la catégorie « médical », et de *photon* dans l'expansion *par photons*, qui est un terme de notre échantillon.

70. La *radiothérapie* par **photons** est très efficace (participe à 50 % de guérisons) [...] (SC-R, LPSC, 2009)
71. Ainsi, la *tomographie* par émission de **positons** (TEP) fait l'objet depuis six ans de recherches au sein de l'équipe Interface Physique-Médecine. (SC-R, LPSC, 2007)

Par ailleurs, l'exemple 71 comporte un autre indice allant dans le même sens que ces observations : le fait qu'une équipe de recherche se nomme *Interface Physique-Médecine* montre qu'il existe bien des ponts importants entre ces disciplines, que la recherche est véritablement interdisciplinaire.

Enfin, dans le même ordre d'idée, les différents termes ou syntagmes dans lesquels l'adjectif *médical* apparaît nous permettent de consolider nos interprétations. Par exemple, dans le contexte 72, cet adjectif modifie le terme *accélérateur*, formant ainsi le terme *accélérateur médical*. En outre, dans l'exemple 73, l'adjectif *médical* modifie également les noms *application* et *usage* (dans le syntagme *accélérateur à usage médical*), ainsi qu'*imagerie* dans le syntagme *détecteur pour l'imagerie médicale*. Ces emplois suggèrent encore une fois que la recherche en physique des particules profite largement au domaine médical.

72. Depuis près de 25 ans, les équipes des accélérateurs du **CERN** contribuent au développement d'accélérateurs *médicaux*. (SC-R, CERN, 2014)
73. Pour que le **CERN** devienne un important vecteur d'échanges pour la physique *médicale* en Europe, le Bureau du **CERN** pour les applications *médicales* a été créé en janvier. [...] Son programme de travail initial comporte sept axes : le calcul à grande échelle, les détecteurs pour l'imagerie *médicale*, les radio-isotopes, la création d'une nouvelle installation biomédicale, la conception optimisée d'accélérateurs à usage *médical*, la simulation et la dosimétrie, et les applications autres que le traitement du cancer. (SC-R, CERN, 2014)

6.3.2.3. Importance de certains termes dans plusieurs domaines connexes

La prise en compte des contextes des unités *combustible* et *supernova* dans les SC-R et SC-V permet de faire des remarques similaires quant aux liens entre les différents domaines. Dans les exemples 74 et 73, nous notons en effet que l'unité *combustible* est utilisée dans des contextes ayant trait à l'énergie nucléaire, ce qui est clairement mis en évidence par plusieurs indices lexicaux (notamment les unités *uranium* et *réacteur*), de la même manière que nous l'avons constaté dans le SC-P.

74. Une part importante du travail du groupe porte sur les données nucléaires : le groupe développe des outils innovants permettant d'évaluer l'impact des incertitudes des sections efficaces sur l'évolution en cœur des *combustibles* ; il mesure d'une part les sections efficaces de diffusion élastique des **neutrons** sur les noyaux légers tel le lithium et le fluor intervenant dans la composition d'un sel fondu et d'autre part les rendements de production des produits de fission de la filière thorium. (SC-R, LPSC, 2011)
75. Pour obtenir la régénération, il faut changer de technologie, développer de nouvelles filières de réacteurs, utilisant des **neutrons** rapides pour le cycle uranium ou un *combustible* liquide pour le cycle thorium. (SC-V, Élémentaire, 2005)

De même, les contextes de l'unité *supernova*, lorsque celle-ci cooccure avec le terme *neutrino* illustrent les liens entre le domaine spatial et la physique des particules, tout comme nous avons pu attester ces liens à partir des exemples tirés du SC-P. Par exemple, le contexte 77 met particulièrement bien en évidence le fait que, pour permettre l'étude des neutrinos, ceux-ci doivent être produits « dans les centrales nucléaires » ou observés dans des explosions de supernovæ.

76. Le flash de **neutrinos** issus de la *supernova* dure ainsi une dizaine de secondes et, à cause des interactions complexes qui ont eu lieu, on pense que toutes les saveurs de neutrinos et d'antineutrinos émergent de l'étoile. (SC-V, Élémentaire, 2007)
77. Des **neutrinos** sont aussi émis par certains noyaux radioactifs (de manière naturelle ou dans les centrales nucléaires), ou lors de l'explosion de *supernovæ*, et d'autres encore sont les reliques de l'explosion originelle du Big-Bang. (SC-V, Élémentaire, 2006)

6.3.2.4. Aspects extraordinaires, hors norme

Parmi les unités de la catégorie « sensationnalisme », les unités *énorme*, *fantastique* et *magnifique* permettent d'observer, dans les sous-corpus intermédiaires, des contextes similaires aux contextes du SC-P, dans lesquels des aspects extraordinaires ou hors norme liés à la recherche en physique des particules sont mis en avant.

Par exemple, les adjectifs *magnifique* et *fantastique* modifient les mêmes unités dans les sous-corpus intermédiaires et dans le SC-P, notamment les unités *instrument* et *machine*, lesquelles reprennent *LHC*, dans les SC-C et SC-P.

78. Des analyses plus fines et les données supplémentaires que nous fournira cette *magnifique machine* en 2012 nous donneront assurément la réponse. (SC-C, CERN, 2011)
79. "Avec les collisions de noyaux, le **LHC** est devenu une *fantastique machine* à big bang [...] (SC-C, CERN, 2010)
80. L'objectif premier est l'exploitation totale de l'accélérateur **LHC**, un *instrument magnifique*", qui doit être relancé au printemps 2015 pour atteindre des énergies de collision deux fois plus haute (15 TeV) que jusqu'à maintenant. (SC-P, Le Temps, 2014)
81. A des dizaines de mètres dans le sous-sol genevois, le plus *fantastique instrument* scientifique jamais construit est sur le point d'être achevé : le **LHC**. (SC-P, Le Temps, 2007)
82. Voilà des mois que, par milliers, ces chercheurs construisaient, optimisaient, fignolaient ce **LHC**, la plus *fantastique machine* à découverte jamais imaginée, devisée à 10 milliards de francs. (SC-P, Le Temps, 2010)

En particulier, la reprise de *LHC* par *machine* (ou *machine à big bang* ou *machine à découverte*) et *instrument* (ou *instrument scientifique*), qui sont modifiés par les adjectifs *magnifique* et *fantastique*, contribue à déterminer le sens de *LHC* comme une machine ou un instrument magnifique, fantastique, et même extraordinaire. D'après le Petit Robert¹⁴⁷ et le TLFi¹⁴⁸, en effet, ces deux adjectifs peuvent prendre le sens de remarquable, extraordinaire, formidable, incroyable, etc. Cette interprétation est par ailleurs renforcée par la présence d'autres indices, qui dénotent ces mêmes aspects, notamment des superlatifs dans les exemples 81 et 82. Dans cette perspective, il semble donc que notre hypothèse selon laquelle des connotations liées à des aspects extraordinaires et hors norme apparaissent pour certains termes dans le SC-P peut être nuancée. Ces connotations sont en effet véhiculées également dans les sous-corpus intermédiaires, que ce soit dans le SC-C comme dans ce cas précis ou dans les autres intermédiaires, comme dans les exemples suivants, où ce sont particulièrement les superlatifs « le plus puissant » et « le plus grand accélérateur de particules du monde » qui mettent en évidence les aspects extraordinaires et hors norme du LHC.

83. Les physiciens espèrent néanmoins le matérialiser grâce à l'énergie libérée dans les grands accélérateurs, et notamment avec le plus puissant d'entre eux : le **LHC** [...]. (SC-V, La Recherche, 2005)
84. L'installation phare du CERN est le Grand collisionneur de hadrons (**LHC**), le plus grand **accélérateur de particules** du monde. [...] Le 30 novembre 2009, les faisceaux

¹⁴⁷ Le Petit Robert en ligne, <https://petitrobert.lerobert.com/robert.asp> (page consultée le 17 septembre 2020).

¹⁴⁸ Le Trésor de la Langue Française informatisé, <http://stella.atilf.fr/>, page consultée le 17 septembre 2020.

de protons du **LHC** ont circulé à 1,18 TeV, un niveau d'énergie encore jamais atteint dans un accélérateur. (SC-R, CERN, 2009)

6.3.2.5. Aspects mystérieux, imaginaires

Parmi les unités de la catégorie « sensationnalisme » qui évoquent des aspects mystérieux, imaginaires ou fascinants, seul l'adjectif *fantomatique* apparaît avec une fréquence relative plus élevée dans le SC-V que dans le SC-P. L'observation des contextes de cet adjectif dans le SC-V permet d'illustrer certaines similarités, lorsqu'il qualifie des termes tels que *boson de Higgs*, *tachyon*, *particule virtuelle*, *neutrino* ou encore simplement l'hyperonyme *particule*.

Dans le TLFi¹⁴⁹, *fantomatique* est défini au figuré comme « [q]ui est sans réalité » et, dans le Grand Robert en ligne¹⁵⁰, comme « [i]rréel, imaginaire et plus ou moins fantastique ». Dans cette perspective, cet adjectif semble bien contribuer à véhiculer une certaine part de mystère et d'imaginaire autour des termes qu'il modifie, aussi bien dans le SC-V que dans le SC-P, comme l'illustrent les exemples 85 à 90 ci-dessous. Cette part de mystère et d'imaginaire n'est donc pas caractéristique du SC-P uniquement, comme nous en avons fait l'hypothèse *supra*.

85. [...] ou encore des informations supplémentaires obtenues sur les *fantomatiques neutrinos*. (SC-V, Élémentaire, 2016)
86. Du coup, puisque la masse du top est revue à la hausse, la masse du *fantomatique boson de Higgs* croît aussi, passant de 96 GeV à 117 GeV. (SC-V, La Recherche, 2014)
87. En brisant des particules, les physiciens espèrent enfin capter l'empreinte du *fantomatique boson de Higgs*. (SC-P, Le Monde, 2008)
88. La réponse pourrait être le **boson de Higgs**, une *particule fantomatique* [...] (SC-P, Le Monde, 2008)
89. [...] les physiciens espèrent [...] faire surgir des états de la matière inconnus et des *particules fantomatiques* [...] (SC-P, Le Monde, 2008)
90. Par exemple, en remettant au goût du jour les *fantomatiques "tachyons"*, ces particules supraluminiques à l'existence jamais prouvée. (SC-P, Les Échos, 2012)

Cependant, bien que les usages de cet adjectif semblent similaires dans ces deux sous-corpus, certaines nuances peuvent tout de même être observées, selon le sous-corpus et selon le terme pris en compte. Plus précisément, dans le SC-V, *fantomatique* se rapporte le plus souvent à *neutrino* (dans 5 occurrences sur les 8 de cette unité), en modifiant soit ce terme, soit un hyperonyme dans une reprise anaphorique. Par ailleurs, dans ce sous-corpus, seule une occurrence se rapporte au terme *boson de Higgs* (exemple 86), alors que, dans le SC-P, cet adjectif se rapporte essentiellement à *boson de Higgs* (dans 7 occurrences sur les 14 de *fantomatique*), soit directement, soit en modifiant *particule* ou *boson* dans une reprise anaphorique. En outre, même si *fantomatique* est utilisé pour

¹⁴⁹ Le Trésor de la Langue Française informatisé, <http://stella.atilf.fr/> (page consultée le 13 mai 2021).

¹⁵⁰ Le Grand Robert en ligne, <https://grandrobert.lerobert.com/robert.asp> (page consultée le 13 mai 2021).

qualifier d'autres termes (par exemple *particule* ou *tachyon* dans les exemples 89 et 90), cet adjectif ne réfère jamais à *neutrino* dans le SC-P.

Par ailleurs, il est intéressant de constater que les neutrinos sont parfois nommés *particules fantômes* dans les sous-corpus intermédiaires et dans le SC-P, de par leur peu d'interaction avec la matière et la difficulté à les détecter, comme l'illustrent les exemples 91 à 94. De ce point de vue, il est possible que l'adjectif *fantomatique*, qui est dérivé de *fantôme*, renvoie en réalité au syntagme *particule fantôme* dans les occurrences où *fantomatique* modifie *neutrino* (ou un hyperonyme de *neutrino*).

91. Le **neutrino** est une particule élémentaire particulièrement difficile à détecter – à tel point qu'elle s'est vue attribuer le surnom de "*particule fantôme*". (SC-C, IN2P3, 2015)
92. Leur étude est particulièrement difficile car elles interagissent très peu avec la matière. Si peu que les **neutrinos** sont souvent dénommés "*particules fantômes*". (SC-V, La Recherche, 2006)
93. Le **neutrino**, une *particule fantôme* (SC-R, LAL, 2007)
94. Les **neutrinos**, ces insaisissables *particules fantômes*, sont particulièrement délicats à détecter. (SC-P, Le Monde, 2003)

Parallèlement, *particule fantôme* est utilisé dans le SC-P pour qualifier également *boson de Higgs* (exemple 95) et les particules composant la matière noire, encore inconnues des chercheurs (exemple 96).

95. [...] le **boson de Higgs**, "particule de Dieu" pour certains ; *particule fantôme*, surtout, aussi introuvable qu'essentielle. (SC-P, Le Monde, 2014)
96. La quête de la **matière noire** est ainsi devenue un des principaux objectifs du grand collisionneur de hadrons (Large Hadron Collider ou LHC) du CERN, depuis qu'il a repris du service, l'an passé, dans sa configuration surpuissante. Là encore, il ne s'agit pas de "voir" la *particule fantôme*, mais de mettre en évidence le drap qui l'habille. (SC-P, Le Monde, 2015)

Ainsi, lorsque nous comparons les usages de *fantomatique* et de *particule fantôme* dans le SC-P et dans les sous-corpus intermédiaires, il apparaît que, dans le SC-P, ces usages sont plus diversifiés et se rapportent à un plus grand nombre de termes que dans les intermédiaires. À ce stade, il semble que ce soit essentiellement le pouvoir évocateur des unités *fantomatique* et *fantôme* qui expliquerait cette plus grande diversité dans le SC-P. De ce point de vue, même s'il est possible de faire l'hypothèse que les aspects mystérieux mis en évidence par ces unités sont véhiculés aussi bien dans les sous-corpus intermédiaires, et en particulier dans le SC-V, que dans le SC-P, il semble bien, à partir de ces observations, que ces aspects soient plus prononcés et fréquents dans le SC-P.

Enfin, bien que nous nous focalisions ici sur un cas particulier, nous parvenons au même constat à partir d'autres unités de la catégorie « sensationnalisme » du même type. C'est ce

qu'illustrent par exemple les contextes où *matière noire* cooccure avec les unités *mystère*, *mystérieux* et, dans une moindre mesure, *fameux* dans les sous-corpus intermédiaires.

97. L'énergie noire (ou constante cosmologique) et la **matière noire**, aussi *mystérieuses* l'une que l'autre, sont les deux composantes dominantes. (SC-R, LAL, 2010)
98. La nature de la **matière noire** de l'Univers reste l'un des plus grands *mystères* de la physique. (SC-R, CERN, 2013)
99. Cette hypothèse est aujourd'hui infirmée car les neutrinos s'avèrent être trop légers et le *mystère* de la **matière noire** reste entier. (SC-V, Élémentaire, 2007)
100. En effet si cette *fameuse* **matière noire** était constituée de particules exotiques très massives situées dans le halo galactique [...] (SC-V, Élémentaire, 2009)
101. La supersymétrie est l'une des explications possibles du *mystère* de la **matière noire**. (SC-C, CERN, 2014)
102. Il s'agit d'une théorie plus élégante de la matière ordinaire, qui pourrait aussi expliquer la *mystérieuse* **matière noire** qui constitue environ un quart de l'Univers. (SC-C, CERN, 2011)

6.4. Bilan

À la lueur des observations détaillées dans ce chapitre, différents constats peuvent être effectués et plusieurs hypothèses quant à la déterminologisation peuvent être formulées. En effet, l'exploration des *keywords* dans les sous-corpus permet d'obtenir une vue d'ensemble de la diversité des contextes dans lesquels les termes s'insèrent dans le SC-P et qui diffèrent du SC-S, à travers l'identification de thématiques récurrentes mises en évidence dans les contextes des termes par la présence des *keywords*. En outre, la majorité des phénomènes ainsi repérés s'observent également dans au moins un sous-corpus intermédiaire, avec toutefois certaines nuances selon le phénomène et/ou le sous-corpus, par exemple des nuances relatives aux objets des financements entre les SC-C et SC-R et le SC-P, de même que la récurrence de la mention des sommes investies dans le SC-P, qui reste caractéristique de ce sous-corpus. De la même manière, certaines unités apparaissent dans des contextes plus diversifiés dans le SC-P que dans les sous-corpus intermédiaires, notamment *fantomatique* et *fantôme*, ce qui peut dénoter une plus grande créativité dans la presse, que nous mettons en lien avec les recommandations des guides de rédaction journalistique concernant l'adoption d'un langage plus imagé lorsque c'est approprié.

Trois aspects au moins peuvent par ailleurs être retenus de nos observations dans ce chapitre. Premièrement, la mise en évidence des liens que la physique des particules entretient avec d'autres domaines, aussi bien dans le SC-P que dans les sous-corpus intermédiaires, interpelle en particulier l'interdisciplinarité de la recherche et le fait qu'un terme peut occuper une place centrale dans plusieurs domaines. La récurrence des contextes où nous observons ces phénomènes, que ce soit dans le SC-P ou dans les sous-corpus intermédiaires, suggère que ces phénomènes ont de

réelles implications pour le processus de déterminologisation. Par exemple, étant donné que chaque domaine dans lequel un concept est central est susceptible de le conceptualiser d'une manière qui lui est propre, comme l'explique L'Homme (2004 : 43-44), dès lors que nous nous plaçons dans la perspective de la déterminologisation, alors il est possible que toutes ces conceptualisations coexistent dans un contexte non spécialisé et soient reflétées par les différences de distribution mises en évidence dans ce chapitre, contribuant ainsi aux changements sémantiques résultant de la déterminologisation.

Deuxièmement, nous avons montré à plusieurs reprises que les unités de la catégorie « sensationnalisme » ne sont pas des spécificités de la presse. Au contraire, la notion de sensationnalisme est bien présente dans les sous-corpus intermédiaires, particulièrement à travers les unités *énorme*, *fantastique*, *magnifique*, *fantomatique*, *mystère*, *secret* et *mystérieux*, bien que dans une moindre mesure en comparaison avec le SC-P. Ainsi, il semble que, si les unités évoquant cette notion contribuent à l'apparition de nouvelles connotations, alors ces connotations peuvent apparaître dans les sous-corpus intermédiaires au cours du processus de déterminologisation, et non nécessairement dans la presse uniquement. Nous verrons par ailleurs au Chapitre 8 comment ces connotations permettent d'appréhender les mécanismes de changement sémantique résultant de la déterminologisation.

Troisièmement, certains phénomènes restent *a priori* caractéristiques du SC-P, puisqu'ils n'ont pas été observés dans les sous-corpus intermédiaires. Soulignons en particulier les unités *Arlésienne*, *AAA* et *Louvre*, ainsi que les usages métaphoriques des termes. Dans le premier cas, plusieurs hypothèses peuvent être formulées à propos de l'absence de ces unités des sous-corpus intermédiaires. Celle-ci peut en effet être attribuée au fait que les domaines auxquels renvoient les unités sont potentiellement trop éloignés de la physique des particules ou ne collaborent peut-être pas aussi directement (dans le cas de *Louvre*¹⁵¹ et de *AAA*¹⁵²). En ce qui concerne l'unité *Arlésienne*, il semble que son usage dans le SC-P relève d'une volonté stylistique de la part des journalistes, qui n'est probablement pas souhaitable dans les textes des sous-corpus intermédiaires. D'ailleurs, cette hypothèse pourrait également s'appliquer à l'unité *Graal* qui n'apparaît qu'à deux reprises dans le SC-V uniquement¹⁵³. Ce type de référence culturelle serait donc une particularité de la presse. Dans le second cas, il semble bien que l'emploi métaphorique de termes soit typique du SC-P, en

¹⁵¹ Dans le SC-P, cette unité permet d'observer l'utilisation d'accélérateurs de particules dans un but d'analyse d'œuvres d'art au musée du Louvre (cf. exemples 23 et 24, section 6.2.1).

¹⁵² Cette unité est en réalité le sigle d'une entreprise pharmaceutique (Advanced Accelerator Applications) créée par d'anciens collaborateurs du CERN (<https://www.adacap.com/>, page consultée le 19 mai 2021).

¹⁵³ En effet, comme nous le précisons *supra*, bien que cette unité apparaisse *a priori* avec une fréquence relative plus élevée dans le SC-R que dans le SC-P (cf. figure 6.3, section 6.3.1), il s'agit en réalité d'un homographe.

comparaison avec les autres sous-corpus. Ce genre de changement dans l'usage des termes est par ailleurs représentatif de la déterminologisation, comme nous le disons en 6.2.7. Nous reviendrons de manière plus détaillée sur ces usages métaphoriques dans le Chapitre 8, sur lesquels nous nous basons pour interroger de manière plus approfondie le processus de déterminologisation.

Parallèlement, d'autres phénomènes caractéristiques de la presse ont été observés à un autre niveau. À plusieurs reprises, en effet, des contextes identiques dans le SC-P et dans le SC-C ou le SC-V ont été mis en évidence, essentiellement des citations ou des parties d'interviews. Ce constat fait écho à la notion de dialogisme, développée à l'origine par le Cercle de Bakhtine (Bakhtine, 1984) et explorée aujourd'hui notamment en analyse de discours (Moirand, 2011 ; Nowakowska et Sarale, 2011 ; Bres, 2017 entre autres). Cette notion est généralement mise en œuvre pour caractériser le fait que tout discours est orienté vers d'autres discours (Nowakowska et Sarale, 2011 : 9), ce qui se manifeste « sous forme de *traces* analysables par le linguiste » (Bres, 2017 : §8). Si, à la suite par exemple de Bres (*Ibid.* : §15), nous considérons « qu'un genre du discours peut être en interaction dialogique avec un autre genre du discours », alors il semble que les genres de la presse généraliste dont relèvent les textes inclus dans le SC-P sont en « interaction dialogique » avec les genres des textes inclus dans les SC-C et SC-V. Cette remarque fait écho à notre réflexion méthodologique dans l'étape de constitution du corpus, au cours de laquelle nous avons fait le choix d'inclure des communiqués de presse précisément du fait qu'ils sont dédiés aux journalistes (cf. Chapitre 4, 4.2.2.3.1), mais elle permet également de confirmer, du moins en partie, les hypothèses formulées dans cette étape, à propos du rôle d'autres intermédiaires dans le processus de déterminologisation. Ainsi, différents genres de textes, que nous avons qualifiés d'*intermédiaires*, peuvent effectivement contribuer au processus de transfert de termes de textes spécialisés dans des textes non spécialisés, qui semble par ailleurs suivre une progression par étape des experts vers les non-experts. Nous reviendrons plus en détail sur cette idée de progression dans le Chapitre 8.

Enfin, si certains phénomènes sont propres au SC-P, il semble qu'ils soient l'exception et non la règle. Les phénomènes observés dans les sous-corpus intermédiaires démontrent en effet qu'ils ne sont pas nécessairement le fait du genre de la presse. Cet aspect est particulièrement intéressant dans notre travail, car il permet non seulement d'interroger le rôle des textes intermédiaires dans le processus de déterminologisation, mais également de confirmer la pertinence des sous-corpus intermédiaires dans notre étude, lesquels apportent véritablement de nouveaux éléments de réflexion à propos de la déterminologisation. Nous développerons davantage cette réflexion dans le Chapitre 8, en tenant compte des phénomènes mis en évidence par l'exploration du second indice distributionnel déterminé pour notre travail, qui fait l'objet du chapitre suivant.

Chapitre 7 Exploration des contextes syntaxiques en corpus

Dans ce chapitre, nous abordons l'exploration du second indice distributionnel dans le corpus, qui se fonde essentiellement sur l'analyse des contextes syntaxiques des termes. Cet indice nous permet d'apporter un regard complémentaire sur les questions soulevées dans le Chapitre 6. Nous organisons notre propos en trois sections principales. Dans la section 7.1, nous proposons d'exploiter une mesure de la similarité sémantique afin de dresser un premier portrait des comportements distributionnels des termes dans les sous-corpus. Dans ce but, nous mettons en œuvre la similarité de Jaccard, qui rend les comparaisons entre les différentes paires de sous-corpus plus aisées. Dans la section 7.2, nous nous focalisons sur une comparaison entre le SC-P et le SC-S, afin de mettre en évidence les fonctionnements des termes les plus divergents entre ces deux sous-corpus, et nous les interprétons en rapport avec la déterminologisation. Enfin, dans la section 7.3, nous explorons les phénomènes retenus dans les sous-corpus intermédiaires et cherchons particulièrement à interroger le rôle de ces sous-corpus dans le processus de déterminologisation tel que nous pouvons l'observer dans ces données.

7.1. Point de vue quantitatif : mesure objective des différences de distribution

Dans cette section, nous adoptons une perspective essentiellement quantitative afin d'appréhender l'analyse des contextes syntaxiques des termes. Les données récupérées automatiquement sont stockées sous la forme de tableaux, qui regroupent la fréquence des unités composant les contextes syntaxiques¹⁵⁴ dans chaque sous-corpus et pour chaque terme (cf. Chapitre 5, 5.3.2). Puisque ces données sont complexes et comportent plusieurs variables dont nous devons tenir compte pour mettre au jour des phénomènes intéressants en rapport avec la déterminologisation, l'intérêt de l'approche quantitative que nous proposons réside dans la

¹⁵⁴ À ce stade, nous souhaitons clarifier l'emploi que nous faisons dans ce chapitre des termes *contexte* et *unité* pour désigner les données analysées. Ces données sont composées des contextes syntaxiques des 15 termes sélectionnés pour la mise en œuvre de cet indice, c'est-à-dire des unités qui apparaissent dans les contextes distributionnels de ces termes et qui entretiennent avec eux une des trois relations de dépendance syntaxique définies dans le Chapitre 5 (5.3.2). Ces unités sont donc des noms, des adjectifs et des participes passés lorsqu'ils modifient les termes, des verbes lorsque les termes sont en objet et en sujet. Ainsi, dans ce chapitre, nous utilisons les termes *unité* ou *donnée*, ou encore l'expression *unité qui entretient une relation de dépendance syntaxique avec les termes*, pour faire référence à ces données et le terme *contexte* pour désigner les contextes des termes issus du corpus, et non pas les seuls contextes syntaxiques des termes.

possibilité d'obtenir rapidement une première vue d'ensemble de la proximité distributionnelle des termes à l'étude dans les sous-corpus.

Dans ce but, nous mettons en œuvre la mesure de la similarité de Jaccard à partir des unités qui composent les contextes syntaxiques des termes dont la fréquence minimum est de 2. Nous faisons le choix de ne pas considérer les unités qui n'apparaissent qu'à une seule reprise. En effet, il semble peu probable que ces unités mettent en évidence des fonctionnements des termes intéressants du point de vue de la déterminologisation, puisque ce sont des hapax. Ce choix permet en outre de réduire notre échantillon de manière conséquente¹⁵⁵, ce qui a l'avantage de le rendre nettement plus gérable pour l'approche descriptive que nous mettons en œuvre en 7.2 et 7.3. En effet, vu la complexité des données à analyser et des comparaisons à effectuer, ainsi que les difficultés techniques, ergonomiques et cognitives que cela implique (cf. Chapitre 5, 5.2.3), restreindre ces données en vue de cibler le plus précisément possible les phénomènes les plus intéressants *a priori* est indispensable.

Dans cette perspective, le recours à une mesure de similarité sémantique comporte plusieurs avantages : d'abord, la démarche d'analyse globale est soutenue par la prise en compte d'un point de vue objectif sur les données. Ensuite, cette phase permet de dresser un portrait global des comportements des termes dans les sous-corpus, qui rend possibles des comparaisons entre les différents sous-corpus et entre les différents termes. Enfin, la prise en compte de cette mesure vise à guider l'analyse dans la masse de données, par exemple en mettant en évidence les termes pour lesquels la distribution diverge le plus entre deux sous-corpus.

Nous faisons le choix de mettre en œuvre la similarité de Jaccard (Jaccard, 1901)¹⁵⁶. Dans des données textuelles, cette mesure sert le plus souvent à calculer la similarité sémantique de deux unités à partir de la proportion de contextes distributionnels partagés par ces unités et propres à chacune des deux unités dans un corpus (Habert et Zweigenbaum, 2003 : 104 ; Périnet et Hamon, 2014 : 508 ; Levshina, 2015 : 235). Néanmoins, cette mesure peut également être utilisée dans un but légèrement différent, où c'est l'usage d'une même unité qui est interrogé, dans deux corpus distincts. La similarité de Jaccard peut donc permettre de mesurer objectivement les différences de distribution d'une unité dans deux ensembles de données, à partir de la proportion des contextes

¹⁵⁵ En effet, le nombre d'hapax représente, pour chaque terme et pour chaque relation, plus de la moitié des unités extraites. Plus précisément, le pourcentage d'hapax par rapport au nombre total d'unités récupérées pour chaque terme varie entre 55,71 % pour le terme *muon* dans la relation verbe-objet et 84,62 % pour le terme *collisionneur* dans la même relation.

¹⁵⁶ Ce calcul est parfois appelé *coefficient* (Tutin, 2007b ; Levshina, 2015), *indice* (Périnet et Hamon, 2015), *mesure* (Périnet et Hamon, 2014) ou encore *distance* (Habert et Zweigenbaum, 2003) *de Jaccard*. Dans notre travail, nous faisons néanmoins le choix d'utiliser *similarité de Jaccard*, à la suite notamment de Habert *et al.* (2005), étant donné qu'il s'agit bien de mesurer dans quelles proportions deux ensembles sont similaires.

distributionnels de cette unité qui sont communs dans les deux ensembles (Basile et McGillivray, 2018 : 195). C'est précisément dans cette seconde optique que nous nous servons de cette mesure.

Plusieurs raisons sous-tendent le choix de la similarité de Jaccard en particulier. D'abord, cette mesure figure parmi les plus utilisées dans des travaux où il est question de similarité sémantique, d'après Périnet et Hamon (2014 : 508) et Bouhandi (2019 : 478) par exemple. En outre, cette mesure est relativement facile et rapide à mettre en œuvre, et elle s'adapte aisément aux différents besoins de chaque analyse. Ainsi, suivant Habert et Zweigenbaum (2003) et Tutin (2007b), notamment, nous proposons de calculer la similarité de Jaccard uniquement à partir d'unités qui entretiennent certaines relations de dépendance avec les termes à l'étude et qui répondent à une contrainte de fréquence. Enfin, d'après Périnet et Hamon (2015 : 84), cette mesure produit des résultats de qualité satisfaisante sur des corpus spécialisés.

Afin de mettre en œuvre cette mesure, trois informations sont nécessaires (Levshina, 2015 : 223-235) :

- le nombre d'unités partagées dans les contextes des deux unités analysées, que nous notons a à la suite de Levshina (*Ibid.*) ;
- le nombre d'unités attestées uniquement dans les contextes de la première unité, que nous notons b ;
- le nombre d'unités attestées uniquement dans les contextes de la seconde unité, que nous notons c .

À partir de ces informations, la similarité de Jaccard se calcule de la manière suivante : $a/(a+b+c)$ (*Ibid.* : 235).

Dans notre cas, puisque nous utilisons cette mesure pour obtenir un indice de la similarité de la distribution d'un même terme dans deux sous-corpus différents, ces informations doivent être adaptées :

- a devient le nombre d'unités attestées dans les contextes du terme dans deux sous-corpus ;
- b devient le nombre d'unités attestées dans les contextes du terme uniquement dans le premier sous-corpus ;
- c devient le nombre d'unités attestées dans les contextes du terme uniquement dans le deuxième sous-corpus.

Le résultat de ce calcul se situe entre 0 et 1, 1 indiquant que tous les contextes sont identiques et 0 indiquant qu'aucun contexte n'est partagé (Habert *et al.*, 2005 : 288).

Parallèlement, étant donné que la similarité de Jaccard sert à mesurer la similarité pour deux éléments uniquement, nous faisons le choix de mesurer la similarité de la distribution des termes dans le SC-P par rapport aux quatre autres sous-corpus et dans le SC-S par rapport aux quatre autres sous-corpus également. Les résultats obtenus permettent ainsi de repérer dans quelles paires de sous-corpus la distribution des termes est la plus proche et la plus éloignée.

Avec ces différents calculs, nous cherchons également à interroger l'hypothèse d'un changement sémantique progressif entre le SC-S et le SC-P, qui pourrait s'observer dans la diversification de la distribution des termes, du SC-S au SC-P et à travers les sous-corpus intermédiaires. De ce point de vue, nous faisons l'hypothèse que les différences de distribution sont plus importantes entre les SC-S et SC-P qu'entre chacun de ces sous-corpus et les sous-corpus intermédiaires. Le calcul de la similarité de Jaccard représente alors un moyen de vérifier cette hypothèse. Ainsi, pour un terme, un résultat plus faible entre les SC-P et SC-S qu'entre chacun de ces sous-corpus et les sous-corpus intermédiaires reflétera une proportion d'unités partagées entre les SC-S et SC-P plus faible qu'entre ces sous-corpus et les sous-corpus intermédiaires.

Afin de dresser un portrait aussi proche de la réalité que possible, nous calculons la similarité de Jaccard à partir des trois relations simultanément, pour chaque paire de sous-corpus considérée (entre les SC-S et SC-P, entre le SC-S et chacun des sous-corpus intermédiaires, et entre le SC-P et chacun des sous-corpus intermédiaires). La comparaison de ces différents résultats permet ainsi de vérifier si, pour les termes considérés, la similarité est effectivement moindre entre les SC-S et SC-P qu'entre chacun de ces sous-corpus et les sous-corpus intermédiaires. Les résultats sont résumés dans les trois tableaux ci-dessous.

Terme	SC-P et SC-S
accélérateur	0,24
atome	0,05
boson de Higgs	0,24
collisionneur	0,38
détecteur	0,27
électron	0,24
interaction	0,18
ion	0,29
muon	0,05
neutrino	0,31
neutron	0,14
noyau	0,17
particule	0,37
proton	0,3
quark	0,21
Similarité moyenne	0,23

Tableau 7.1 : Similarité de Jaccard entre les SC-P et SC-S

Terme	SC-S et SC-C	SC-S et SC-R	SC-S et SC-V
accélérateur	0,28	0,21	0,3
atome	0,09	0,17	0,11
boson de Higgs	0,27	0,27	0,34
collisionneur	0,34	0,36	0,37
détecteur	0,29	0,34	0,39
électron	0,1	0,27	0,36
interaction	0,1	0,32	0,43
ion	0,26	0,16	0,32
muon	0,07	0,15	0,21
neutrino	0,3	0,3	0,37
neutron	0,1	0,06	0,13
noyau	0,31	0,19	0,19
particule	0,33	0,42	0,47
proton	0,18	0,31	0,41
quark	0,26	0,34	0,43
Similarité moyenne	0,22	0,26	0,32

Tableau 7.2 : Similarité de Jaccard entre le SC-S et les sous-corpus intermédiaires

Terme	SC-P et SC-C	SC-P et SC-R	SC-P et SC-V
accélérateur	0,32	0,39	0,46
atome	0,21	0,18	0,46
boson de Higgs	0,29	0,24	0,43
collisionneur	0,5	0,4	0,51
détecteur	0,31	0,27	0,34
électron	0,17	0,21	0,38
interaction	0,19	0,18	0,26
ion	0,17	0,2	0,45
muon	0,05	0,09	0,18
neutrino	0,34	0,27	0,49
neutron	0,09	0,11	0,35
noyau	0,15	0,36	0,47
particule	0,32	0,36	0,54
proton	0,23	0,25	0,35
quark	0,25	0,19	0,26
Similarité moyenne	0,24	0,25	0,40

Tableau 7.3 : Similarité de Jaccard entre le SC-P et les sous-corpus intermédiaires

Ces résultats permettent d'emblée de nuancer notre hypothèse. Si celle-ci semble se vérifier à partir des similarités moyennes, lorsque chaque terme est pris en compte individuellement, cette hypothèse est, au contraire, infirmée. En effet, si nous tenons compte des similarités moyennes uniquement (dans la dernière ligne de chaque tableau), alors les données indiquent bien que les contextes distributionnels des termes sont moins similaires entre le SC-P et le SC-S qu'entre ces sous-corpus et les sous-corpus intermédiaires, à l'exception de la paire SC-S et SC-C, où la similarité moyenne est légèrement inférieure (0,22 contre 0,23 entre les SC-P et SC-S). Néanmoins, l'observation des similarités de Jaccard pour chaque terme ne permet pas de renforcer cette interprétation, car, pour plusieurs termes, la similarité est plus élevée pour la paire SC-P et SC-S que pour d'autres paires de sous-corpus.

Deux catégories de termes et de fonctionnements peuvent ainsi être distinguées :

- une **première catégorie** regroupe les termes dont la distribution diverge dans une plus large mesure entre les SC-P et SC-S qu'entre ces sous-corpus et les sous-corpus intermédiaires. Plus précisément, deux cas de figure sont observables :
 - o des termes pour lesquels la similarité de Jaccard est plus faible (ou équivalente) entre les SC-P et SC-S qu'entre ces sous-corpus et les sous-corpus intermédiaires (*boson de Higgs, atome, muon et détecteur*) ;

- des termes pour lesquels la similarité de Jaccard est plus faible (ou équivalente) entre les SC-P et SC-S qu'entre ces sous-corpus et les sous-corpus intermédiaires, à l'exception d'une paire de sous-corpus (*noyau, quark, accélérateur et interaction*) ;
- une **seconde catégorie** regroupe les termes dont la distribution diverge dans une plus large mesure entre le SC-P ou le SC-S et les sous-corpus intermédiaires qu'entre les SC-P et SC-S, c'est-à-dire des termes pour lesquels la similarité de Jaccard est plus faible dans plusieurs paires de sous-corpus qu'entre les SC-P et SC-S. Parmi ces cas, nous trouvons :
 - quatre termes pour lesquels la similarité de Jaccard est plus faible dans trois paires de sous-corpus qu'entre les SC-P et SC-S (*proton, particule, électron et collisionneur*) ;
 - deux termes pour lesquels la similarité de Jaccard est plus faible dans quatre paires de sous-corpus qu'entre les SC-P et SC-S (*neutrino et ion*) ;
 - et un terme pour lequel la similarité de Jaccard est plus faible dans cinq paires de sous-corpus qu'entre les SC-P et SC-S (*neutron*).

À la lueur de ces observations, nous pouvons faire l'hypothèse, pour la première catégorie, qu'il s'est produit certains changements sémantiques au cours du processus de déterminologisation, lequel semble par ailleurs cohérent avec notre hypothèse de changement progressif du SC-S au SC-P. Pour la seconde catégorie, nous pouvons faire l'hypothèse que, si ces différences dans les contextes distributionnels des termes renvoient véritablement à des changements sémantiques (ce qui reste à déterminer à partir de l'observation des contextes dans le corpus), alors ces changements ne se produisent pas nécessairement de manière progressive au cours du processus de déterminologisation tel que nous pouvons l'observer dans notre corpus.

Parallèlement, étant donné que la similarité de Jaccard permet d'effectuer des comparaisons non seulement entre les sous-corpus, mais également entre les termes, un autre élément pertinent peut être tiré de ces données. En effet, si nous comparons les valeurs obtenues pour chacun des 15 termes dans le tableau 7.1, nous pouvons aisément identifier les termes pour lesquels ces valeurs sont les plus basses. Ainsi, les cinq termes avec les mesures de similarité les plus basses entre les SC-P et SC-S sont *atome, muon, neutron, noyau et interaction*. Nous pouvons donc faire l'hypothèse que ces termes ont un fonctionnement qui diffère dans une plus large mesure entre ces sous-corpus que les autres termes. Cette hypothèse sera interrogée *infra*, à partir de l'exploration des contextes des termes et de ces unités dans le corpus.

Ces observations permettent ainsi de guider l'analyse dans la suite de la démarche mise en œuvre dans ce chapitre. En effet, avec les différents constats effectués à partir de la similarité de Jaccard, nous pouvons étoffer notre réflexion à propos du processus de déterminologisation et des fonctionnements des termes dans ce processus. Dans les sections suivantes, nous cherchons en particulier à observer ce que révèlent les contextes d'apparition des termes dans le corpus par rapport aux résultats de la similarité de Jaccard et par rapport à la déterminologisation. Nous abordons donc d'abord la comparaison des fonctionnements des termes dans les SC-P et SC-S (en 7.2), puis nous explorons les sous-corpus intermédiaires (en 7.3).

7.2. Observation des différences principales entre les SC-P et SC-S

Nous cherchons dans cette section essentiellement à déterminer ce que les différences de distribution constatées entre les SC-P et SC-S révèlent à propos de la déterminologisation. Afin d'aborder ces questions, nous nous intéressons principalement aux unités qui ne sont attestées que dans l'un de ces deux sous-corpus, indépendamment de leur présence dans les sous-corpus intermédiaires dans un premier temps. Différents fonctionnements sont ainsi repérés, nous les détaillons dans les sections qui suivent.

7.2.1. Différents fonctionnements des termes selon le sous-corpus

7.2.1.1. Modificateurs entrant dans la composition de termes complexes

La première catégorie de fonctionnements concerne les modificateurs qui semblent entrer dans la composition de termes complexes¹⁵⁷. Cette catégorie peut être subdivisée en deux, selon le sous-corpus dans lequel les modificateurs sont repérés. Ainsi, une première sous-catégorie concerne

¹⁵⁷ En réalité, sans une étape de validation par des experts de domaine, nous ne pouvons pas réellement affirmer que ces syntagmes sont effectivement des termes. Néanmoins, plusieurs indices nous permettent tout de même de tendre vers cette interprétation. D'une part, ces syntagmes correspondent aux patrons des termes complexes les plus répandus en français, c'est-à-dire nom + adjectif et nom + nom (L'Homme, 2004), et démontrent un certain degré de figement, ou « *unithood* » (Drouin, 2015 : 420) dans le corpus. D'autre part, l'observation de ces syntagmes dans le corpus montre que, lorsque les modificateurs sont des adjectifs, ceux-ci semblent « entraîne[r] une modification du référent » (Portelance, 1991 : 68), de par leurs propriétés classifiantes, notamment, qui sont caractéristiques des adjectifs entrant dans la composition de termes complexes de type nom + adjectif (Cortès, 2004 : 136-138). Sur ce dernier point, nous pensons que le même raisonnement peut être suivi pour les syntagmes suivant le patron nom + nom. En effet, dans ce patron, le second nom modifie le premier, de la même manière qu'un adjectif modifie un nom. Pour ces raisons, et afin de poursuivre notre réflexion, nous faisons le choix de considérer ces syntagmes comme des termes complexes.

les modifieurs attestés dans le SC-S mais pas dans le SC-P et une seconde catégorie concerne les modifieurs attestés dans le SC-P et non dans le SC-S.

Modifieurs présents dans le SC-S et absents du SC-P. Dans le cas de plusieurs termes, l'observation de leurs modifieurs présents dans le SC-S et absents du SC-P semble indiquer que ceux-ci sont en réalité constitutifs de termes complexes, comme l'illustrent les exemples suivants. Dans ces exemples, le terme soumis à l'analyse est signalé en gras et les modifieurs en italiques.

103. Finalement la sortie de ce réseau de neurones est la nouvelle variable discriminante que nous proposons pour la recherche du **boson de Higgs** *léger standard* dans $D\emptyset$ dans les modes de production associés $W H$ et $Z H$. (SC-S, Université de Strasbourg, 2012)
104. Par conséquent, cette étude, visant à contraindre la normalisation et la probabilité de mauvaise identification, impacte l'ensemble des analyses de l'état final $\mu +$ telle que la recherche de **bosons de Higgs** *supersymétriques* [2, 3], la recherche de particules supersymétriques [...]. (SC-S, Université Clermont 2 - Blaise Pascal, 2011)
105. Le LEP était un **collisionneur** *$e+e-$* avec une énergie dans le centre de masse allant de 91 à 209 GeV. (SC-S, Université de la Méditerranée, Aix-Marseille II, 2010)
106. Le Tevatron est un **collisionneur** *pp* avec une énergie dans le centre de masse de 1.96 TeV. (SC-S, Université de la Méditerranée, Aix-Marseille II, 2010)
107. Un événement charmé avec une désintégration muonique et un **muon** *primaire* non détecté est caractérisé par un **muon** *secondaire* chargé positivement. (SC-S, Université Claude Bernard Lyon-1, 2010)

Cette observation n'est toutefois pas étonnante, dans la mesure où le SC-S est composé de textes de spécialité alors que le SC-P est composé d'articles de presse généraliste. L'attestation de plus nombreux termes complexes dans le SC-S, avec des expansions plus longues, est donc un phénomène tout à fait attendu. En effet, si, à la suite notamment de Condamines *et al.* (2004 : 552), nous partons du principe que des termes avec une plus longue expansion renvoient à des concepts toujours plus précis, alors nos observations indiquent que ces types de termes n'intègrent pas de manière privilégiée la langue générale.

Autrement dit, les termes plus spécifiques apparaissent moins souvent dans le SC-P, tandis que les termes plus génériques – les hyperonymes – le sont de manière plus récurrente. Les différents types de bosons de Higgs, par exemple, sont peu susceptibles d'être évoqués dans la presse. Cependant, l'observation plus fine du terme *boson de Higgs* et de ses modifieurs qui sont attestés dans le SC-S mais pas dans le SC-P met en lumière un phénomène particulier, qui semble même révéler une contradiction dans les données. En effet, l'existence de plusieurs hyponymes de ce terme mise en évidence par différents modifieurs, comme dans les exemples 103 et 104, paraît contradictoire avec les modifieurs *seul* et *unique*, attestés dans le SC-S et non dans le SC-P. Ceux-ci semblent plutôt suggérer que ce terme n'admet pas d'hyponymes.

Dans ce cas, l'observation des contextes dans le SC-S permet de comprendre que, selon la théorie de la physique des particules prise en compte, il peut exister un ou plusieurs types de bosons de Higgs. L'exemple ci-dessous montre ainsi que, dans le Modèle Standard de la physique des particules, un seul boson de Higgs est prédit par la théorie, tandis que, dans des modèles supersymétriques, plusieurs types de bosons de Higgs sont prédits. Autrement dit, il peut exister plusieurs conceptualisations en fonction de la théorie prise en compte. Cet aspect n'est cependant pas observable dans le SC-P et il semble que, selon le sous-corpus, différents aspects de ce concept (ou d'un groupe de concepts) peuvent être mis en évidence.

108. Cette nouvelle particule est-elle l'*unique boson de Higgs* du modèle standard ? Pourrait-il s'agir du premier représentant d'un multiplet de bosons de Higgs tel que le prévoient les modèles supersymétriques ? (SC-S, Reflets de la physique, 2012)

Modificateurs présents dans le SC-P et absents du SC-S. Dans le cas de plusieurs termes, nous observons que certains modificateurs qui entrent dans la composition de termes complexes sont attestés dans le SC-P mais non dans le SC-S. Deux phénomènes peuvent notamment être identifiés à partir de ces termes. Le premier concerne en réalité des variantes dénominatives de termes qui ne sont pas attestées dans le SC-S. Ainsi, dans l'exemple 109, le terme *proton négatif* est une variante d'*antiproton*, comme l'explicitent les parenthèses dans l'exemple. Dans l'exemple 110, le terme *collisionneur électrons-positons* est une variante de *collisionneur e+e-*. Ces observations renvoient au fait que, selon les besoins de la situation de communication, différentes variantes peuvent être utilisées de manière privilégiée. Il est en particulier reconnu que le degré de spécialisation de la communication peut être un facteur influençant le choix d'une variante, par exemple lorsqu'il s'agit de textes de vulgarisation scientifique et même de la presse (Mortureux, 1993, 1995 ; Delavigne, 2001, 2003 ; Moirand, 2007 ; Estopà, 2016 ; Fernández-Silva, 2018 parmi d'autres). Ainsi, dans le cas de *collisionneur électrons-positons*, nous constatons que c'est la variante *collisionneur e+e-*, plus courte, qui est plus largement utilisée par les spécialistes. De la même manière, la variante *proton négatif* est, selon les deux expertes que nous avons interrogées, une ancienne dénomination d'*antiproton*, qui n'est désormais plus employée par les spécialistes.

109. Un atome d'antihydrogène est constitué d'un **proton négatif** (antiproton) et d'un électron positif ou positron. (SC-P, Le Temps, 2010)

110. 1989. Le LEP, grand **collisionneur électrons-positons** installé dans un tunnel circulaire de 27 kilomètres de circonférence entre en service. (SC-P, Le Figaro, 2004)

Le deuxième phénomène fait écho aux observations détaillées dans le Chapitre 6, à propos de la présence des mêmes termes dans des contextes qui renvoient à des domaines différents. Dans

ce cas, nous constatons que certains termes complexes formés à partir des termes à l'étude et des modificateurs repérés sont le plus souvent utilisés dans des contextes qui évoquent d'autres domaines. Ainsi, l'exemple 111 atteste d'un usage des termes *neutron lent* et *neutron rapide* dans le domaine de l'énergie nucléaire et l'exemple 112 atteste d'un usage du terme *ion carbone* dans le domaine de la médecine. Comme dans le Chapitre 6, notre interprétation se fonde en majeure partie sur la cooccurrence de ces termes avec des unités qui renvoient à ces autres domaines, notamment dans ces deux exemples *centrale*, *fission* et *combustible* pour l'énergie nucléaire et *hadronthérapie* pour la médecine.

Par ailleurs, si *neutron lent* et *neutron rapide* sont des hyponymes de *neutron* et puisque *neutron* est un terme du domaine de la physique des particules (selon la procédure de validation détaillée au Chapitre 4, 4.3.2), alors il semble que différents hyponymes peuvent être observés selon le domaine dans lequel l'hyperonyme est utilisé. Ainsi, à partir des données prises en compte, et étant donné qu'ils sont absents du SC-S, nous pouvons supposer que les termes complexes *neutron lent* et *neutron rapide* sont spécifiques au domaine de l'énergie nucléaire. En revanche, d'autres termes complexes formés à partir de la même tête et qui sont attestés dans le SC-S, par exemple *neutron thermique*, sont probablement spécifiques du domaine de la physique des particules.

111. L'objectif principal des six voies de recherche est l'obtention d'un réacteur à **neutrons rapides**, une technologie présentée comme la panacée à tous les maux du nucléaire. Les centrales actuelles fonctionnent avec des **neutrons lents** parce qu'ils génèrent beaucoup de fission, mais ils ont l'inconvénient de produire des noyaux de plus en plus gros dans le combustible. (SC-P, Les Échos, 2004)
112. La hadronthérapie est la prochaine étape. [...] Outre les faisceaux de protons, ils s'intéresseront aussi à l'exploitation des ions lourds, comme les **ions carbone**. (SC-P, Les Échos, 2012)

7.2.1.2. Adjectifs axiologiques

La deuxième catégorie de fonctionnements concerne les modificateurs des termes qui sont en réalité des adjectifs axiologiques. Ces adjectifs sont attestés dans le SC-P mais non dans le SC-S et renvoient à la notion de sensationnalisme, de la même manière que nous l'observons au Chapitre 6 (6.1.1.1). En particulier, les exemples 113 à 115 mettent en évidence le mystère qui entoure la recherche en physique des particules, notamment avec les termes *boson de Higgs* et *neutrino*, et les exemples 116 et 117 la taille hors norme de différents instruments, avec les termes *accélérateur* et *détecteur*.

113. Après la lumière toute faite sur la nouvelle particule découverte au CERN le 4 juillet - qui ressemble furieusement au *mythique boson de Higgs* -, place à une traque plus obscure (SC-P, Le Temps, 2012)
114. En 2009, elles ont resserré le filet autour de cette "**particule divine**", encore hypothétique, qui donnerait leur masse à toutes les autres particules. (SC-P, Le Monde, 2010)
115. Comprendre les *mystérieux neutrinos* (SC-P, Les Échos, 2006)
116. Les physiciens pensent déjà à l'ère post-LHC : un *gigantesque accélérateur* de 100 km de circonférence, qui pourrait se faire soit au CERN soit en Chine dès 2020 (SC-P, Le Temps, 2014)
117. Elles auront lieu quatre fois par tour, au sein d'*énormes détecteurs* capables de reconnaître chacune des particules élémentaires ainsi libérées. (SC-P, Le Monde, 2007)

En comparaison avec nos observations détaillées au Chapitre 6, ces exemples permettent de repérer d'autres unités évoquant cette notion de sensationnalisme. Ainsi, les unités *gigantesque* et *divin*, notamment, qui n'avaient pas été repérées auparavant, complètent nos observations et montrent en particulier la diversité des manières d'évoquer le sensationnalisme dans ce sous-corpus.

7.2.1.3. Adjectifs dérivés de pays ou de régions

Un autre type d'adjectif est repéré dans le SC-P, en comparaison avec le SC-S, il s'agit d'adjectifs dérivés de pays ou de régions, notamment *américain*, *français* et *européen* dans les exemples suivants.

118. Les premières découvertes ne sont pas attendues avant 2009, soit un an avant le début de la fermeture du Fermilab, un **accélérateur américain** cinq fois moins puissant que le LHC mais qui traque lui aussi le "Higgs". (SC-P, Le Figaro, 2008)
119. Le nouvel **accélérateur français** Spiral2 (Système de production d'ions radioactifs accélérés en ligne de 2e génération) prépare activement sa première expérience, prévue pour le second semestre 2017. (SC-P, Les Échos, 2016)
120. Il permettra notamment d'installer le futur **détecteur européen** Eureka dont la capacité sera très largement supérieure à celle d'Edelweiss 2. (SC-P, Le Figaro, 2009)

Il ressort de ces exemples que lorsque des instruments précis sont évoqués dans le SC-P (ici par le biais des termes *accélérateur* et *détecteur*), il semble que les pays ou les régions dans lesquels ces instruments ont été conçus ou qui ont pris part au projet sont précisés. En comparaison avec le SC-S, cet aspect est typique du SC-P. Il s'agit par ailleurs d'unités qui sont ressorties de l'extraction de *keywords* et que nous avons classées comme caractéristiques de la presse dans le Chapitre 6 (cf. 6.1.1.1). Ce qui est intéressant dans ce cas, c'est que, bien que nous ne les ayons pas retenues au chapitre précédent pour cette raison, il apparaît que, lorsque ces adjectifs modifient les termes à l'étude, ils permettent de relever un élément intéressant à mettre en relation avec la

déterminologisation. En effet, ces modificateurs mettent en évidence un aspect particulier des instruments auxquels font référence les termes *accélérateur* et *détecteur* dans le SC-P, qui est totalement absent du SC-S. Il semble donc que, lorsque certains termes sont utilisés dans des textes non spécialisés tels que ceux constituant le SC-P, leur distribution permet d'observer qu'un certain point de vue est porté sur certains aspects des concepts impliqués. Dans le cas de ces termes, notamment, nous pouvons déduire des exemples ci-dessus que la mention des pays abritant ces instruments revêt une importance particulière pour la presse.

7.2.1.4. Verbes faisant partie de collocations spécialisées

L'observation des verbes dont les termes sont sujets et/ou objets ainsi que de leurs contextes dans les sous-corpus met en évidence différents fonctionnements, que l'on peut rapprocher de collocations spécialisées. De manière générale, la notion de collocation peut se définir comme la « cooccurrence lexicale privilégiée de deux éléments linguistiques entretenant une relation syntaxique » (Tutin et Grossmann, 2002 : 9). Cependant, cette notion étant large et pouvant recouvrir différents types d'associations, Tutin propose d'aborder sa définition à l'aide de différents critères, notamment la fréquence, l'arbitrarité ou non-prédictibilité des associations et le fait que la base et le collocatif entretiennent une relation syntaxique (Tutin, 2010 : 34-40).

Parallèlement, la notion de collocation spécialisée¹⁵⁸ se distingue par le fait que, dans ce cas, la collocation est propre à la langue de spécialité d'un domaine en particulier. La base est généralement un terme de ce domaine (L'Homme, 2017 : 218) et le collocatif peut avoir un sens spécialisé dans ce domaine (L'Homme, 1998 : 516). Dans nos données, deux phénomènes au moins peuvent être repérés, selon le sous-corpus dans lequel les collocations sont observées.

Collocations présentes dans le SC-P et absentes du SC-S. Pour les termes *atome* et *neutron*, l'une des différences principales dans les verbes dans la relation verbe-objet concerne les verbes *fusionner* pour le premier terme et *produire* et *émettre* pour le second, qui apparaissent dans le SC-P mais non dans le SC-S. À partir de l'observation des contextes dans le SC-P, il semble que ces verbes forment avec ces termes des collocations qui sont en lien avec le domaine de l'énergie nucléaire, comme dans les exemples ci-dessous. Dans ces exemples, la présence de l'unité *Iter* et des syntagmes *fusion nucléaire*, *combustible nucléaire* et *réacteur à fusion*, qui renvoient à l'énergie nucléaire (cf. Chapitre 6, 6.1.1.2), permettent de renforcer cette interprétation.

¹⁵⁸ Pour un panorama détaillé de la notion de collocation en langue de spécialité, voir notamment Clas (1994), Larivière (1998), L'Homme (1998), Williams (1998) et Maniez (2002).

121. Le principe de la fusion nucléaire consiste à *fusionner* un **atome** de deutérium et un atome de tritium dans une enceinte fermée et maintenue dans un vide très poussé. [...] Iter devrait recevoir une puissance d'entrée de 50 MW et produire une énergie de sortie de l'ordre de 500 MW pendant quatre cents secondes pour confirmer la validité technologique du concept. (SC-P, Les Échos, 2003)
122. L'hélium 3 fait aussi rêver nombre d'ingénieurs nucléaires : n'*émettant* aucun **neutron**, il serait le combustible idéal d'un futur réacteur à fusion. (SC-P, Le Monde, 2004)
123. Cette technologie très innovante consiste à utiliser un accélérateur de particules pour *produire* les **neutrons** qui déclenchent les réactions de fission dans le combustible nucléaire. (SC-P, Les Échos, 2009)

D'autres collocations peuvent être distinguées, qui renvoient à d'autres domaines encore. Notamment, pour les verbes lorsque le terme *atome* est objet, nous repérons *piéger*. Les contextes de cette paire verbe-objet dans le SC-P pointent par ailleurs vers deux usages distincts : d'une part, un usage relatif à l'horloge atomique Pharao¹⁵⁹ et, d'autre part, un usage relatif à l'informatique quantique¹⁶⁰, comme le montrent les deux exemples ci-dessous.

124. L'horloge Pharao consiste à *piéger* les **atomes** de césium, à les refroidir, puis à les injecter dans un faisceau micro-onde. [...] La performance espérée donne le vertige : elle ne perdra qu'une seconde tous les 300 millions d'années - à comparer aux 50 millions d'années des horloges terrestres. (SC-P, Les Échos, 2008)
125. Il permet naturellement une puissance de calcul considérable", explique Philippe Grangier de l'Institut d'optique de Palaiseau, où des qubits sont fabriqués à partir d'**atomes** de rubidium *piégés* à l'aide de "pinces optiques" et conservés sous vide. (SC-P, Les Échos, 2013)

La prise en compte des verbes dans cette étape d'analyse rend alors possibles des distinctions fines selon les domaines dans lesquels les termes à l'étude semblent utilisés. Nous constatons en particulier une certaine diversité dans les domaines auxquels renvoient les contextes des termes, plus grande que ce que nos observations ont permis de dégager au Chapitre 6 (6.1). Il semble ainsi que la présence des mêmes termes dans des domaines différents ait une réelle pertinence pour l'étude du processus de déterminologisation (nous y reviendrons au Chapitre 8).

L'observation de ces collocations permet toutefois d'aller plus loin dans la réflexion. En effet, si, comme le suggère Williams (1998 : 156-157), différentes langues de spécialité sont caractérisées par différents « réseaux de collocations » (*collocational networks*), alors il semble que lorsque le même terme relève de différents domaines, celui-ci peut s'intégrer dans différents réseaux de collocations selon chaque domaine pris en considération, et donc dans différents réseaux

¹⁵⁹ Centre national d'études spatiales, projet Pharao, <https://pharao.cnes.fr/fr> (page consultée le 17 novembre 2020).

¹⁶⁰ Le terme *qubit* désigne l'« unité de base de l'informatique quantique », selon le Grand dictionnaire terminologique (<http://gdt.oqlf.gouv.qc.ca/>, page consultée le 17 novembre 2020).

conceptuels également. C'est ce que montrent les exemples ci-dessus : étant donné que le SC-P n'est pas composé de textes spécialisés dans un certain domaine, alors les différents réseaux de collocations d'un même terme peuvent y être observés.

Collocations présentes dans le SC-S et absentes du SC-P. Dans d'autres cas, nous remarquons certaines collocations spécialisées dans le SC-S et non dans le SC-P. Cependant, dans les exemples suivants, il est intéressant de constater que, bien que la collocation *prédire + particule* ne soit pas attestée dans le SC-P, deux formulations alternatives sont relevées, à savoir *imaginer + particule* et *inventer + particule*. Ainsi, dans les exemples 126 et 127 tirés du SC-P, nous observons des contextes similaires aux exemples 128 et 129, qui traitent tous de la prédiction d'un type de particule dans le cadre d'un modèle théorique particulier¹⁶¹.

126. François Englert et Peter Higgs n'ont obtenu leur Nobel pour le boson éponyme que l'an dernier, soit 49 ans après avoir *imaginé* cette **particule**. (SC-P, Le Figaro, 2014)

127. Mais cela ne satisfait pas nombre de physiciens. Ils ont donc d'ores et déjà *inventé* une **particule** responsable de la gravitation et l'ont baptisé [sic] graviton. (SC-P, Le Figaro, 2008)

128. Enfin, les extensions supersymétriques du Modèle Standard *prédisent* de nouvelles **particules** pouvant se désintégrer en lepton faisant des états finals tauiques une sonde de nouvelle physique. (SC-S, Université Paris-Sud, 2011)

129. L'élégance de la supersymétrie ne doit cependant pas faire oublier qu'elle n'a, jusqu'à présent aucune validité expérimentale, puisque bien qu'activement recherchée au LEP et au TEVATRON, aucune des nouvelles **particules** qu'elle *prédit* n'a encore été observée. (SC-S, Université Claude Bernard Lyon-1, 2004)

Ces exemples amènent également une autre remarque, qui renvoie plus précisément aux caractéristiques stylistiques de la presse. En effet, l'emploi des verbes *imaginer* et *inventer* peut contribuer aux stratégies qui visent à susciter l'intérêt des lecteurs et qui « jouent » sur la fascination que la recherche en physique des particules peut exercer, particulièrement lorsque l'existence de certaines particules est prédite par plusieurs modèles théoriques mais non nécessairement prouvée. De la même manière, bien que les exemples tirés du SC-S et du SC-P expriment des idées similaires, les arguments sujets des verbes sont de natures différentes entre ces deux sous-corpus. Ainsi, dans le SC-S, les sujets de *prédire* sont inanimés. Il s'agit en réalité des différents modèles théoriques de la physique des particules. Au contraire, dans le SC-P, les sujets sont humains (« ils », qui reprend « physiciens » ainsi que « François Englert et Peter Higgs »). Dans ce cas, nous pouvons faire l'hypothèse que l'emploi de verbes plus imagés et le rapprochement effectué avec les chercheurs

¹⁶¹ D'ailleurs, selon les expertes que nous avons interrogées, la recherche expérimentale en physique des particules consiste précisément à observer les particules prédites par les modèles théoriques, afin de prouver la validité de ces modèles. Ainsi, certaines particules, comme le boson de Higgs, sont prédites à l'intérieur d'un modèle théorique avant d'être observées expérimentalement.

peuvent s'expliquer par les spécificités stylistiques et rédactionnelles de la presse, qui est contrainte d'attirer l'attention des lecteurs et de susciter l'intérêt, comme nous le détaillons au Chapitre 6 (6.1.1.1).

Ces observations mettent ainsi en lumière des fonctionnements des termes qui diffèrent entre le SC-P et le SC-S et permettent de nuancer l'importance des différences de distribution repérées entre le SC-P et le SC-S. En effet, celles-ci révèlent le plus souvent non pas des changements sémantiques importants, mais plutôt des fonctionnements propres aux deux sphères représentées dans ces sous-corpus. Il semble ainsi que certains termes sont susceptibles de passer de manière privilégiée dans la presse, selon s'ils sont plutôt génériques ou plutôt spécifiques (comme dans le cas des hyponymes de *boson de Higgs* qui ne sont pas attestés dans le SC-P) ou encore selon la forme du terme (comme dans le cas de certaines variantes dénominatives, notamment *collisionneur électrons-positons*, qui est privilégié par rapport à *collisionneur e+e-*)¹⁶². De même, certaines différences de distribution semblent en réalité refléter des différences de perception du référent ou la mise en évidence de certains aspects des concepts en jeu.

7.2.2. Apparition de thématiques différentes selon le sous-corpus

En plus des fonctionnements détaillés dans la section précédente, nos observations montrent que certaines différences de distribution dans le SC-P par rapport au SC-S permettent de repérer plusieurs thématiques en lien avec les termes pris en compte. Nous distinguons plusieurs phénomènes selon si les termes renvoient à des types d'instruments dédiés à la recherche ou à des types de particules¹⁶³.

7.2.2.1. Thématiques autour des instruments de recherche

Les observations détaillées dans cette section s'articulent autour des termes *accélérateur*, *détecteur* et *collisionneur*. Quatre thématiques sont notamment repérées, dans le SC-P.

Pays hôtes des instruments. Comme nous le disons *supra*, l'emploi de certains modificateurs qui mettent l'accent sur les pays dans lesquels les instruments dédiés à la recherche en physique des

¹⁶² C'est par ailleurs l'une des hypothèses de Meyer et Mackintosh (2000b : 127-129), en particulier à propos des termes métaphoriques et des variantes « moins spécialisées » de termes en comparaison avec des variantes perçues comme « plus spécialisées » (par exemple *cancer* par rapport à *carcinoma*).

¹⁶³ Précisons que les interprétations proposées dans cette section se fondent non seulement sur les données du corpus, mais également sur les différents entretiens menés avec les expertes, qui ont été étalés sur toute la durée de notre travail de thèse.

particules ont été construits ou sur les pays et les régions qui ont largement contribué à leur construction reflète une des caractéristiques de la presse (cf. 7.2.1.3 ; Chapitre 6, 6.1.1.1). Le verbe *accueillir* permet de compléter ces premières observations, notamment lorsque les termes *collisionneur* et *accélérateur* sont objets. Ce cas met particulièrement en évidence les discussions ou négociations autour du lieu de la construction de futurs instruments de grande ampleur. Le plus souvent, ces instruments sont le fruit de collaborations internationales. Par exemple, la construction du Grand collisionneur de hadrons (LHC) au CERN a impliqué les États membres de l'organisation ainsi que plusieurs États non membres, tels que le Japon, les États-Unis ou l'Inde¹⁶⁴. Il semble alors que la détermination du pays dans lequel les futurs instruments seront construits constitue un enjeu important dans ces collaborations internationales. C'est ce qu'illustre l'exemple ci-dessous.

130. Pékin propose de financer et d'*accueillir* le successeur du LHC, un **collisionneur** géant de plus de 50 km de circonférence. (SC-P, Le Figaro, 2014)

Emplacement des instruments. Les verbes et les modificateurs des termes *accélérateur*, *collisionneur* et *détecteur* mettent en évidence une thématique à propos des emplacements des instruments, qui diffère dans une certaine mesure selon le sous-corpus considéré. Ainsi, dans le SC-P, l'accent est porté sur le caractère souterrain des gros instruments dédiés à la recherche, avec notamment les unités *enfoui* et *enterré* qui modifient ces termes et le verbe *abriter* lorsque les termes sont objets (exemples 131 et 132). En revanche, dans le SC-S, c'est l'emplacement des instruments les uns par rapport aux autres qui est mis en évidence par des unités telles que *proche*, *lointain*, *placer* ou *situer* (exemples 133 et 134). Soulignons par ailleurs que l'emplacement des instruments d'un point de vue géographique, en précisant le pays, la ville ou l'université ou le laboratoire de rattachement, est commun à ces deux sous-corpus (exemples 135 et 136).

131. D'énormes cavernes ont été creusées à 100 mètres sous terre pour *abriter* des **détecteurs** de plusieurs milliers de tonnes qui tenteront de "voir" le résultat de collisions entre deux faisceaux de protons lancés à près de 300 000 km/s. (SC-P, Le Figaro, 2004)

132. En effet, à partir du 20 mai plus personne n'aura accès aux tunnels qui *abritent* l'**accélérateur**, le démarrage du LHC étant prévu pour cet été. (SC-P, Le Temps, 2008)

133. Pour accroître [sic] l'énergie des particules, on les fait circuler dans des accélérateurs linéaires ou circulaires et on *place* des **détecteurs** de particules aux points de collision. (SC-S, Université de la Méditerranée, Aix-Marseille II, 2008)

134. Ce **détecteur**, *situé* entre le solénoïde et le calorimètre, couvre la région |d| ... (SC-S, Université Paris-Sud, 2011)

135. La violation de la vitesse de la lumière a été observée sur un faisceau de neutrinos [...] produits par l'**accélérateur** du Cern, près de Genève... (SC-P, Le Figaro, 2011)

¹⁶⁴ Organisation européenne pour la recherche nucléaire, Chiffres clés du LHC, <https://home.cern/fr/resources/faqs/facts-and-figures-about-lhc> (page consultée le 25 novembre 2020).

136. Le complexe d'**accélérateurs** du Tevatron. Le laboratoire Fermilab, situé à proximité de Chicago (Illinois) aux Etats Unis accueille diverses expériences de physique des hautes énergies. (SC-SUniversité de Strasbourg, 2012)

Coûts de construction et de maintenance des instruments. Dans le SC-P, nous repérons un autre aspect à l'aide des verbes *financer* et *justifier* lorsque les termes sont objets et *coûter* lorsque les termes sont sujets. Ces verbes permettent non seulement de constater que les questions de financement liées à la construction de tels instruments sont particulièrement présentes dans le SC-P en comparaison avec le SC-S, comme nous l'avons montré dans le Chapitre 6 (6.2.3 et 6.3.2.1), mais également d'observer un certain discours autour de la justification de la construction de tels instruments (exemples 137 et 138). Il semble ainsi que, dans le SC-P, les acteurs de la recherche en physique des particules participent à justifier ces activités de recherche qui ont un certain coût pour la société, en particulier les importantes sommes nécessaires à la réalisation des gros instruments.

137. Maintenant que le boson manquant a été découvert, comment convaincre de poursuivre ces recherches ? Si on ne trouve aucun indice qu'il existe quelque chose au-delà du modèle standard, *justifier* de nouveaux **accélérateurs** sera difficile. (SC-P, Le Monde, 2012)

138. "Pour que la construction de la future machine soit soutenue par les décideurs, dit Rolf Heuer, elle doit être motivée par un défi scientifique clair", tout comme la traque du boson de Higgs était la raison d'être du LHC. Et tous espèrent que ce dernier permettra sous peu d'entrouvrir des portes vers cette physique inédite. "D'un autre côté, on pourrait ne pas avoir besoin de ces résultats pour *justifier* un nouvel **accélérateur**, ose Yifang Wang. Le savoir actuel, et les lacunes qu'il contient, devrait suffire comme raison." (SC-P, Le Temps, 2014)

Possibilités de visite et d'inauguration des instruments. Une dernière thématique est repérée de cette manière qui met également clairement en évidence l'existence d'un point de vue particulier sur les instruments. En effet, les verbes *inaugurer* et *visiter*, qui sont attestés dans le SC-P et non dans le SC-S, suggèrent que les instruments désignés par *détecteur* et *accélérateur* s'apparentent à des objets que l'on peut inaugurer et visiter, de la même manière que l'on peut inaugurer et visiter un monument ou un bâtiment public par exemple. Plus précisément, il semble que l'utilisation du verbe *inaugurer* dénote un caractère officiel, comme l'illustre l'exemple 139 ci-dessous. Cet aspect est particulièrement intéressant, dans la mesure où la construction et la mise en marche de ces instruments sont le plus souvent le fruit de collaborations internationales. La tenue d'un événement visant à inaugurer ces instruments, ou à en « [c]élébrer l'achèvement », d'après la définition du TLFi¹⁶⁵, semble ainsi ajouter une dimension solennelle ou formelle à une collaboration impliquant différents pays. Enfin, pour ce qui est de l'unité *visiter*, le contexte reproduit dans l'exemple 140

¹⁶⁵ Le Trésor de la Langue Française informatisé, <http://stella.atilf.fr/>, page consultée le 25 novembre 2020.

renvoie en réalité à des actions de valorisation et/ou de vulgarisation de la part du CERN, qui prennent la forme d'une « journée portes ouvertes ».

139. CLAUDIE HAIGNERÉ et Guido Possa, les ministres français et italien délégués à la recherche, ont *inauguré*, mardi 18 novembre à La Seyne-sur-Mer (Var), le **détecteur** de neutrinos Antares. (SC-P, Le Monde, 2003)
140. Dimanche, la journée portes ouvertes du CERN a attiré les foules : 50000 personnes ont profité du riche programme tout public mis en place à cette occasion et 21000 d'entre elles ont pu descendre 100 mètres sous terre pour *visiter* le grand **collisionneur** de hadrons (LHC). (SC-P, Le Temps, 2008)

7.2.2.2. Thématique autour des particules

Dans cette section, nous montrons comment l'observation fine des différences dans les verbes utilisés avec les termes désignant des types de particules (ou l'hyperonyme *particule*) lorsqu'ils sont objets ou sujets permet de dégager plusieurs thématiques, qui s'actualisent de différentes manières selon le sous-corpus considéré. Trois thématiques ont ainsi été repérées.

Recherche vs étude des particules. Ce premier cas concerne les termes *particule*, *boson de Higgs* et *neutrino*, pour lesquels nous dégageons deux thématiques à partir des verbes attestés dans le SC-P mais non dans le SC-S et des verbes attestés dans le SC-S mais non dans le SC-P. Dans le SC-P, la présence des verbes *trouver*, *débusquer*, *dénicher*, *attraper*, *traquer*, *chasser*, *cerner*, *piéger* ou encore *capturer* lorsque les termes sont objets indique qu'une thématique autour de la recherche de certaines particules (qui n'ont pas encore été observées par les chercheurs ou qui sont difficiles à observer) est particulièrement présente dans ce sous-corpus. En revanche, dans le SC-S, les verbes *reconstruire*, *caractériser*, *comprendre*, *classer*, *définir* ou encore *étudier* mettent plutôt en évidence les efforts fournis par les chercheurs pour mieux comprendre les propriétés de ces particules, pour améliorer leur connaissance de ces particules.

Rôle des particules dans les accélérateurs vs interactions entre les particules. Pour les termes *proton* et *électron*, les différences entre les deux sous-corpus considérés montrent que, dans le SC-P, les contextes insistent sur le rôle de ces particules dans le fonctionnement des accélérateurs de particules, notamment sur le fait qu'elles doivent être guidées dans les accélérateurs (exemple 141) et qu'elles entrent en collision et se désintègrent (exemples 142 et 143), ce qui produit les données dont ont besoin les chercheurs. En revanche, dans le SC-S, les verbes utilisés avec ces termes mettent plutôt en évidence les interactions des particules entre elles et avec d'autres particules (exemples 144 à 146).

141. De puissants aimants *guident* les **électrons** [...] pour les faire tourner dans ce grand anneau à raison de 350 000 tours par seconde, vingt-quatre heures sur vingt-quatre, sept jours sur sept... (SC-P, Le Figaro, 2015)
142. D'où la construction d'un microscope géant, le Large Hadron Collider (LHC), qui, depuis fin 2009, *fracasse* à grande vitesse des **protons** les uns contre les autres, dans un manège souterrain de 27 kilomètres de circonférence. (SC-P, Le Monde, 2011)
143. ... cet accélérateur fait, dès 1989, *s'entrechoquer* des **électrons** et leurs alter ego positifs, des positrons (SC-P, Le Temps, 2004)
144. Au delà de t 150 s (T 90 keV), les **protons** et les neutrons injectés *interagissent* avec les baryons du plasma avant de thermaliser. (SC-S, Université Montpellier 2, 2008)
145. Pendant la collision, les deux **protons** *interagissent* et forment de nouvelles particules, quarks ou leptons... (SC-S, Université Clermont 2 - Blaise Pascal, 2011)
146. Ce gaz est injecté dans un instrument appelé le Duoplasmatron, un cylindre de métal qui à l'aide un champ électrique va *séparer* les **protons** des **électrons**. (SC-S, Université Claude Bernard Lyon-1, 2014)

Différentes utilités pour la recherche. Ce dernier cas concerne plus particulièrement le terme *quark*, pour lequel nous observons deux aspects distincts de la recherche autour de cette particule dans les SC-S et SC-P. Ainsi, les contextes de ce terme avec des verbes tels que *libérer*, *agglutiner* ou *agglomérer* dans le SC-P montrent un aspect de la recherche faisant intervenir les quarks et les gluons. Comme l'illustrent les exemples 147 et 148, l'accent est mis sur le passage de « l'état premier de la matière », c'est-à-dire le moment où les quarks et les gluons n'étaient pas liés entre eux, à l'état actuel, en quelque sorte, où les quarks et les gluons constituent les autres particules. La thématique qui ressort de ces contextes inclut également la reconstitution expérimentale de cet « état premier de la matière », pour mieux comprendre les origines de la matière. Dans le SC-S, au contraire, et c'est ce qui est particulièrement intéressant, les verbes et leurs contextes mettent plutôt en évidence l'étude des propriétés des différents types de quarks et de leurs interactions avec d'autres particules que les gluons, comme le montrent les exemples 149 et 150. Deux aspects du concept ressortent donc de ces différences : dans le SC-P, son étude permet de mieux comprendre les origines de la formation de la matière, tandis que dans le SC-S, l'étude de cette particule permet de mieux comprendre ses propriétés, ce qui contribue notamment aux recherches qui portent sur le boson de Higgs (exemples 151 et 152 en particulier).

147. Juste après la déflagration originelle du Big Bang, notre Univers était un plasma - une "soupe" incommensurablement dense et chaude - de particules fondamentales. Quelques fractions de seconde plus tard sont apparus les **quarks**, qui se sont *agglutinés* en protons et neutrons, et ont formé les noyaux des atomes. (SC-P, Le Monde, 2008)
148. Dans les entrailles de son grand accélérateur de particules, cette énergie prodigieuse doit permettre de *libérer* les **quarks** de l'emprise des gluons, pour retrouver l'état premier de la matière. (SC-P, Le Temps, 2010)
149. Ceci implique que les **quarks** observables peuvent *se coupler* par interaction faible à des quarks d'une autre famille, ce qui serait impossible sinon. (SC-S, Université de la Méditerranée, Aix-Marseille II, 2007)

150. En fait, expérimentalement, on n'*observe* pas directement les **quarks** ou les gluons, mais les jets de hadrons (tels le pion ou le proton) qui se forment immédiatement par interaction forte dans la direction du quark ou du gluon produit. (SC-S, Reflets de la physique, 2009)
151. Notre idée principale est que le boson de Higgs, objet lourd, *produit* deux **quarks** b avec une bonne séparation angulaire (SC-S, Université de Strasbourg, 2012)
152. Le boson de Higgs se désintègre en deux **quarks** b et le boson Z en deux neutrinos. (SC-S, Université de la Méditerranée, Aix-Marseille II, 2010)

Tous ces exemples permettent ainsi d'illustrer le fait que les divergences de distribution repérées entre les SC-P et SC-S mettent essentiellement en évidence certains aspects particuliers des concepts ou certaines thématiques particulières de recherche en physique des particules où ces concepts interviennent. À ce stade, nous pensons que ces aspects reflètent l'existence de différents points de vue, lesquels sont plus ou moins représentés dans les sous-corpus, en fonction des caractéristiques des genres textuels considérés et de la prise en compte de l'intérêt supposé des lecteurs. En outre, à partir de ces éléments, nous pouvons faire l'hypothèse que, lorsqu'un terme intègre la langue générale, tous les aspects relatifs au concept ne passent pas nécessairement de manière homogène ou uniforme au cours du processus. En effet, nos observations ne rendent pas seulement compte d'une répartition différente de ces aspects selon le sous-corpus considéré, mais elles indiquent que certains de ces aspects sont caractéristiques de l'un ou l'autre des sous-corpus. Notamment, il semble que certains aspects repérés dans le SC-S ne « passent » pas dans le SC-P et que d'autres aspects, absents du SC-S, « apparaissent » dans le SC-P, probablement en fonction de ce qui est jugé comme étant le plus intéressant pour le lectorat ciblé. Dans la section suivante, nous proposons d'explorer davantage les données dans cette direction en nous focalisant sur les sous-corpus intermédiaires.

Parallèlement, la confrontation entre ces observations et les résultats du calcul de la similarité de Jaccard montre que, le plus souvent, les faibles scores obtenus notamment par les termes *atome*, *muon*, *neutron*, *noyau* et *interaction* révèlent en réalité soit une différence importante dans le nombre d'unités qui ont été récupérées et qui ont servi au calcul, soit des usages très diversifiés dans les deux sous-corpus, par exemple lorsque les termes apparaissent dans des contextes qui évoquent des domaines différents de la physique des particules. De même, les importantes différences de distribution mises en évidence par la similarité de Jaccard pour les termes *boson de Higgs*, *quark*, *accélérateur* ou encore *détecteur* entre le SC-P et le SC-S renvoient plutôt au constat que, selon le sous-corpus pris en considération, différents aspects liés à ces concepts ou différentes thématiques sur lesquelles le sous-corpus insiste peuvent être dégagées, ce qui se traduit par une distribution divergente.

Au vu des observations détaillées dans cette section, il est désormais clair que l'analyse contrastive de la distribution des termes dans le SC-P et dans le SC-S, à partir des contextes syntaxiques, ne permet pas d'observer une véritable « dilution du sens terminologique d'origine », pour reprendre la distinction proposée par Meyer et Mackintosh (2000a : 205). Néanmoins, différents phénomènes sont identifiés, qui soutiennent les premières conclusions auxquelles nous parvenons dans le Chapitre 6 et qui les complètent. Parallèlement, nos observations permettent de montrer que ces différents fonctionnements des termes en situation de déterminologisation peuvent renvoyer à la mise en évidence de différents aspects des concepts, selon les besoins de la situation de communication ou selon le point de vue adopté. Une question cruciale doit toutefois être abordée. Ces aspects ou ces points de vue sont-ils réellement caractéristiques du SC-P ? Avec cette question, nous cherchons en réalité à déterminer si ces aspects « apparaissent » dans le SC-P ou s'ils sont également présents dans les sous-corpus intermédiaires, c'est-à-dire dans différentes étapes du processus de déterminologisation telles que nous pouvons les observer dans le corpus. Nous discutons cette question dans la section suivante.

7.3. Exploration dans les sous-corpus intermédiaires

Nous nous focalisons ici sur l'observation, dans les sous-corpus intermédiaires, des phénomènes repérés *supra* et cherchons en particulier à comprendre ce que leur répartition dans ces sous-corpus peut révéler à propos de la déterminologisation.

7.3.1. Modificateurs entrant dans la composition de termes complexes

Dans la section 7.2.1.1, nous constatons que plusieurs modificateurs servent à former des termes complexes, lesquels sont attestés soit dans le SC-S soit dans le SC-P. Dans cette section, nous observons dans quelle mesure ces termes complexes sont également attestés dans les sous-corpus intermédiaires. Nous articulons notre propos en suivant la même distinction que précédemment, à savoir entre les modificateurs présents dans le SC-S et absents du SC-P et les modificateurs présents dans le SC-P et absents du SC-S.

Modificateurs présents dans le SC-S et absents du SC-P. Parmi les modificateurs attestés dans le SC-S et non dans le SC-P, certains sont également attestés dans au moins un sous-corpus intermédiaire, tandis que d'autres ne sont présents que dans le SC-S. Afin d'illustrer notre propos, nous reproduisons certains modificateurs des termes *boson de Higgs*, *muon* et *collisionneur* dans les trois tableaux ci-dessous, ainsi que leur fréquence absolue dans les cinq sous-corpus.

Modifieurs de <i>boson de Higgs</i>	SC-S	SC-C	SC-R	SC-V	SC-P
chargé	42	-	27	-	-
neutre	27	-	2	1	-
léger	26	1	2	3	-
additionnel	15	-	-	-	-
lourd	15	-	1	-	-
supplémentaire	10	-	-	-	-
supersymétrique	3	-	-	-	-
pseudoscalaire	2	-	-	-	-

Tableau 7.4 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de *boson de Higgs*

Modifieurs de <i>collisionneur</i>	SC-S	SC-C	SC-R	SC-V	SC-P
hadronique	11	-	2	4	-
proton-antiproton	6	1	6	12	-
pp ⁻	2	-	-	-	-
e ⁺ e ⁻	1	-	5	3	-

Tableau 7.5 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de *collisionneur*

Modifieurs de <i>muon</i>	SC-S	SC-C	SC-R	SC-V	SC-P
isolé	41	-	-	-	-
primaire	34	-	-	-	-
cosmique	33	1	28	9	-
secondaire	21	-	-	-	-
central	4	-	-	-	-
charger	4	-	-	-	-

Tableau 7.6 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de *muon*

Dans ces tableaux, nous observons ainsi que certains modifieurs :

- sont attestés dans le SC-S et dans tous les sous-corpus intermédiaires (par exemple, *léger* pour *boson de Higgs*, *proton-antiproton* pour *collisionneur* ou encore *cosmique* pour *muon*),
- sont attestés dans le SC-S et dans deux sous-corpus intermédiaires (par exemple, *neutre* pour *boson de Higgs*, qui apparaît dans les SC-R et SC-V ou *e⁺e⁻* pour *collisionneur* dans les SC-R et SC-V),
- sont attestés dans le SC-S et dans un sous-corpus intermédiaire (par exemple, *lourd* et *chargé* pour *boson de Higgs*, qui apparaissent dans le SC-R),

- sont attestés dans le SC-S uniquement (par exemple, *additionnel*, *supplémentaire* ou *supersymétrique* pour *boson de Higgs*, pp^- pour *collisionneur*, ou encore *isolé*, *primaire* ou *secondaire* pour *muon*).

De manière générale, soulignons que les modifieurs attestés dans le SC-S et dans tous les sous-corpus intermédiaires sont moins nombreux que les autres. Par exemple, pour le terme *boson de Higgs*, 54 modifieurs sont attestés dans le SC-S, contre seulement 11 dans le SC-C, 17 dans le SC-R et 19 dans le SC-V. À ce stade, il semble donc qu'une plus grande proportion de modifieurs dans le SC-S peut refléter une plus grande proportion de termes complexes dans le SC-S en comparaison avec les quatre autres sous-corpus, ou alors la présence de termes complexes comportant de plus longues expansions (comme le terme *boson de Higgs léger standard*, illustré dans l'exemple 103 dans la section 7.2.1.1). Cette affirmation est à mettre en relation avec le fait que les termes complexes avec des expansions plus longues (donc un nombre de modifieurs plus élevé) renvoient à des concepts plus précis. Le SC-S étant le sous-corpus composé des textes les plus spécialisés, cette répartition des modifieurs dans les sous-corpus semble confirmer cette affirmation (cf. 7.2.1.1).

Modifieurs présents dans le SC-P et absents du SC-S. Parmi les modifieurs présents dans le SC-P et absents du SC-S qui entrent dans la composition de termes complexes, certains sont attestés dans au moins un sous-corpus intermédiaire et d'autres uniquement dans le SC-P. Ainsi, la présence de *neutron lent*, *neutron rapide* ou *ion carbone* dans les SC-R et SC-V, par exemple, dans des contextes qui se rapportent à d'autres domaines que la physique des particules (notamment l'énergie nucléaire dans l'exemple 153 et la médecine dans l'exemple 154 ci-dessous), permet de renforcer les hypothèses émises *supra* et évoque encore une fois l'interdisciplinarité de la recherche et le fait qu'un terme peut être central dans différents domaines.

153. La production de ^{239}Pu a déjà lieu dans les réacteurs actuels, mais ses propriétés neutroniques rendent impossible la régénération [...] Pour obtenir la régénération, il faut changer de technologie, développer de nouvelles filières de réacteurs, utilisant des **neutrons rapides** pour le cycle uranium ou un combustible liquide pour le cycle thorium. (SC-V, Élémentaire, 2005)

154. Le Centre national d'hadronthérapie oncologique (CNAO) de Pavie (Italie) a utilisé pour la première fois des **ions carbone** pour traiter un patient. (SC-R, CERN, 2012)

En revanche, l'absence de *négatif* dans les modifieurs du terme *proton* dans tous les sous-corpus intermédiaires suggère que *proton négatif* est effectivement une variante d'*antiproton* qui est plutôt caractéristique du SC-P. Enfin, soulignons le cas d'*électron-positon*, qui modifie le terme *collisionneur* dans le SC-P et non dans le SC-S, et de sa variante $e+e-$, qui modifie *collisionneur* dans le

SC-S et non dans le SC-P. Ces deux modifieurs sont attestés dans au moins un sous-corpus intermédiaire, comme le montre le tableau 7.7 ci-dessous.

Modifieurs de <i>collisionneur</i>	SC-S	SC-C	SC-R	SC-V	SC-P
e+e-	1	-	5	3	-
électron-positon	-	5	37	12	3

Tableau 7.7 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de *collisionneur*

L'absence de *e+e-* du SC-P (et du SC-C) et la présence d'*électron-positon* dans le SC-P et dans les trois sous-corpus intermédiaires contribue selon nous non seulement à justifier la pertinence des intermédiaires dans notre analyse, mais suggère également que ces intermédiaires ont un rôle important dans le processus de déterminologisation. En effet, il semble d'une part que la prise en compte du SC-S seul comporte le risque de fournir une image imparfaite (ou incomplète) de la diversité des usages à l'intérieur d'une langue de spécialité. D'autre part, cette observation permet de nourrir l'hypothèse selon laquelle les termes passent dans la langue générale (du moins dans la presse dans ce cas) par le biais de différents intermédiaires. Nous pouvons même aller plus loin : contrairement au terme *proton négatif*, qui est une variante d'*antiproton* caractéristique du SC-P, *collisionneur électron-positon* apparaît bien dans tous les intermédiaires et dans le SC-P. Ainsi, bien que nous ne puissions pas prouver que ce terme est attesté dans le SC-P parce qu'il y a été transféré par les sous-corpus intermédiaires, nos observations suggèrent tout de même que le processus d'intégration de termes dans la langue générale ne s'effectuerait pas systématiquement à partir des textes les plus spécialisés (tels que ceux inclus dans le SC-S). Cette question sera davantage détaillée dans le Chapitre 8.

7.3.2. Adjectifs axiologiques

De manière générale, l'observation dans les sous-corpus intermédiaires des adjectifs axiologiques, que, dans ce cas, nous rapprochons de la notion de sensationnalisme, montre qu'ils ne sont pas réellement caractéristiques du SC-P. Néanmoins, la répartition de ces unités dans les sous-corpus intermédiaires permet de dégager certaines tendances, que nous illustrons avec les termes *boson de Higgs*, *particule* et *neutrino* et des unités suscitant la part de mystère et la fascination entourant la recherche autour de ces termes, d'une part, et *accélérateur*, *collisionneur* et *détecteur* d'autre part, avec des adjectifs évoquant la taille hors norme des instruments.

Unités suscitant le mystère et la fascination. À partir des données reproduites dans les trois tableaux ci-dessous, nous constatons que ces unités ne sont jamais attestées dans le SC-S,

lorsqu'elles modifient les termes *boson de Higgs*, *particule* ou *neutrino*, à une exception près. En effet, une seule occurrence de *mystérieux* apparaît dans le SC-S, avec le terme *particule* (exemple 155). Dans cet exemple, « particule mystérieuse » est en réalité une reprise anaphorique de *boson de Higgs*. De la même manière, seule cette unité est attestée dans le SC-R, lorsqu'elle modifie le terme *neutrino*. Précisons que cette occurrence de *mystérieux* apparaît dans le titre d'un séminaire cité dans ce sous-corpus, son importance est donc toute relative (exemple 156).

155. Avec les données actuellement analysées, le Tevatron exclut que la masse du boson de Higgs du Modèle Standard soit entre 160 et 170 GeV/c² (1). L'analyse des prochaines données couvrira la zone entre 115 et 160 GeV/c², favorisée par la théorie et par les indications expérimentales indirectes. Nous concluons sur les perspectives de découverte de cette **particule** *mystérieuse* dans les deux ou trois prochaines années. (SC-S, Reflets de la physique, 2009)

156. Jean-Eric Campagne *Mystérieux Neutrino* 12 février 2003 - SFP, Lyon (SC-R, LAL, 2004)

Modifieurs de <i>boson de Higgs</i>	SC-S	SC-C	SC-R	SC-V	SC-P
fameux	-	-	-	8	47
insaisissable	-	-	-	1	6
mythique	-	-	-	-	3
fantomatique	-	-	-	1	2

Tableau 7.8 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de *boson de Higgs*

Modifieurs de <i>particule</i>	SC-S	SC-C	SC-R	SC-V	SC-P
insaisissable	-	4	-	3	16
mystérieux	1	1	-	3	14
fugace	-	1	-	-	6
énigmatique	-	1	-	1	5
divin	-	-	-	-	4
mythique	-	-	-	-	3
fantomatique	-	-	-	2	3

Tableau 7.9 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de *particule*

Modifieurs de <i>neutrino</i>	SC-S	SC-C	SC-R	SC-V	SC-P
fameux	-	-	-	1	2
mystérieux	-	-	1	2	1

Tableau 7.10 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de *neutrino*

En revanche, un plus grand nombre de ces unités apparaissent dans le SC-C, en particulier lorsqu'elles modifient le terme *particule*, dans des contextes où ce terme reprend *boson de Higgs* ou *neutrino*, comme l'illustrent les exemples 157 et 158. De plus, presque toutes les unités repérées dans le SC-P sont également attestées dans le SC-V et lorsqu'elles modifient le terme *particule*, il s'agit le plus souvent de reprises de *boson de Higgs* ou de *neutrino* (exemples 159 et 160).

157. Cette accumulation de données permet de marquer un progrès sensible dans la quête du boson de Higgs, mais ne suffit pas pour trancher sur l'existence ou la non-existence de cette *insaisissable* **particule**. (SC-C, CERN, 2011)
158. L'expérience Opera, se déroulant au laboratoire souterrain du Gran Sasso en Italie, étudie le phénomène d'oscillations des neutrinos, ces **particules énigmatiques** interagissant rarement avec la matière et remplissant l'Univers pratiquement au même titre que les photons. (SC-C, IN2P3, 2013)
159. Les neutrinos sont des **particules** bien *mystérieuses*, puisqu'elles interagissent très peu avec la matière et peuvent donc la traverser sans difficulté. (SC-V, Élémentaire, 2007)
160. Pas question, toutefois, d'observer directement la *fameuse* **particule** : le boson de Higgs est trop instable. (SC-V, La Recherche, 2012)

Ces observations concordent donc avec celles du Chapitre 6 (6.3.2.5). Par ailleurs, si ces adjectifs modifient les termes dans tous les sous-corpus, l'échantillon de données présenté ici semble indiquer que leur répartition dans les sous-corpus suit un certain schéma. Ainsi, les unités que nous prenons en compte ici sont les plus répandues dans le SC-P, puis dans le SC-V. Elles apparaissent dans une moindre mesure dans le SC-C et sont plus rares dans le SC-R et même très rares dans le SC-S. Il semblerait alors que plus les textes se rapprochent de la langue générale, plus ces unités sont fréquentes.

Un constat similaire peut être effectué pour d'autres groupes d'unités, notamment les verbes *prédire*, *imaginer* et *inventer* lorsque le terme *particule* est objet. Dans la section 7.2.1.4, nous soulignons le fait que les verbes *imaginer* et *inventer* sont utilisés dans des contextes similaires aux contextes du verbe *prédire* dans le SC-P et le SC-S respectivement. Dans ces contextes, ces trois verbes semblent renvoyer à la même action, mais en suscitant différemment l'imagination des lecteurs : les verbes *inventer* et *imaginer* interpelleraient plus particulièrement l'imaginaire supposé autour de la recherche en physique des particules, tandis que *prédire* apparaît comme une collocation spécialisée de ce domaine.

Le tableau 7.11 illustre la répartition de ces verbes dans les sous-corpus intermédiaires. Nous constatons ainsi que les unités *imaginer* et *inventer* sont attestées dans le SC-P et dans le SC-V uniquement, ce qui suggère bien que, dans des textes se rapprochant de la langue générale, il est plus probable d'observer des unités plus évocatrices que les collocations propres au domaine.

Verbes lorsque <i>particule</i> est objet	SC-S	SC-C	SC-R	SC-V	SC-P
prédire	10	1	7	5	-
imaginer	-	-	-	1	3
inventer	-	-	-	3	3

Tableau 7.11 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de verbes lorsque *particule* est objet

Soulignons également que l'attestation des verbes *imaginer* et *inventer* dans le SC-V met encore une fois en évidence la pertinence de ce sous-corpus dans notre étude de la déterminologisation. En effet, avec une comparaison SC-S/SC-P seule, il aurait été aisé de supposer que ce type de verbes « apparaît » dans la presse et que les changements dans les contextes distributionnels des termes seraient à imputer uniquement à la presse généraliste. Néanmoins, le cas de ces deux verbes permet une nouvelle fois de nuancer cette interprétation et montre que certaines différences dans la distribution d'un terme s'observent en réalité dans une étape intermédiaire du processus de déterminologisation. Nous reviendrons par ailleurs plus en détail dans le Chapitre 8 sur la proximité apparente entre le SC-P et le SC-V.

Adjectifs évoquant la taille hors norme des instruments. À partir des modificateurs *gigantesque* et *énorme* repérés dans le SC-P, nous nous sommes intéressée aux autres modificateurs qui caractérisent la taille des instruments désignés par les termes *détecteur*, *accélérateur* et *collisionneur*. Les tableaux 7.12 à 7.14 ci-dessous illustrent la fréquence absolue de ces modificateurs, pour ces trois termes.

Modificateurs de <i>détecteur</i>	SC-S	SC-C	SC-R	SC-V	SC-P
immense	-	1	-	-	19
grand	6	8	33	16	15
énorme	-	1	1	2	11
géant	1	2	1	8	11
gigantesque	-	4	2	4	9
haut	-	-	2	-	2
volumineux	-	2	6	1	1
gros	1	1	3	1	1

Tableau 7.12 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modificateurs de *détecteur*

Modifieurs d'<i>accélérateur</i>	SC-S	SC-C	SC-R	SC-V	SC-P
grand	3	22	26	49	103
géant	3	-	-	6	36
gigantesque	-	-	-	1	8
immense	-	-	-	-	8
gros	-	-	-	-	6

Tableau 7.13 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs d'*accélérateur*

Modifieurs de <i>collisionneur</i>	SC-S	SC-C	SC-R	SC-V	SC-P
grand	23	157	88	136	121
géant	-	-	-	2	3

Tableau 7.14 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de modifieurs de *collisionneur*

Deux constats peuvent être faits. Premièrement, aucune unité n'est relevée dans le SC-S uniquement. Toutes les unités qui sont attestées dans ce sous-corpus le sont également au moins dans le SC-P et un sous-corpus intermédiaire (*grand*, *gros* et *géant*), l'unité *grand* étant d'ailleurs la seule à être attestée dans les cinq sous-corpus avec les trois termes. La question de la taille des instruments désignés par ces termes est donc exprimée dans tous les sous-corpus. Cela dit, l'observation des contextes montre que, le plus souvent, pour le terme *accélérateur*, il est en réalité fait référence à certains accélérateurs en particulier, notamment le Grand accélérateur national d'ions lourds (GANIL) (exemple 161) ou le LHC lorsqu'il est qualifié de « plus grand accélérateur du monde » (exemple 162). De la même manière, le terme *collisionneur* est généralement modifié par l'adjectif *grand* dans la dénomination française du LHC, c'est-à-dire *Grand collisionneur de hadrons* (exemple 163), ou dans la dénomination française du LEP, c'est-à-dire *Grand collisionneur électron-positron* (exemple 164).

161. Le projet SPIRAL2 (Système de Production d'Ions Radioactifs Accélérés en Ligne 2) est une extension du *Grand Accélérateur* National d'Ions Lourds (GANIL) de Caen. (SC-R, LPSC, 2009)
162. ... les représentants des états membres du CERN décident, en 1982, de construire le plus *grand accélérateur* du monde pour explorer un domaine d'énergie allant de 70 à 200 GeV. (SC-V, Élémentaire, 2016)
163. Aujourd'hui, les faisceaux de protons sont entrés en collision dans le *Grand collisionneur* de hadrons (LHC), au Cern à Genève, à une énergie record de 13 téraélectronvolts (TeV). (SC-C, IN2P3, 2015)
164. À ce moment, le CERN envisage de construire un *grand collisionneur* électron-positron, le LEP. (SC-S, Reflets de la physique, 2009)

Deuxièmement, nous observons une nouvelle fois la même tendance déjà repérée pour les unités *imaginer, inventer, prédire* et les unités *fameux, insaisissable, mythique, fantomatique*, etc., c'est-à-dire que les unités les plus évocatrices et qui peuvent interpeller la notion de sensationnalisme ne sont pas nécessairement caractéristiques du SC-P. Nous remarquons ainsi que les unités *immense, énorme* et *gigantesque* modifient également les termes *détecteur* et *accélérateur* dans au moins un sous-corpus intermédiaire.

En outre, il est intéressant d'observer les différences qui s'illustrent dans les contextes plus larges de ces unités, selon le sous-corpus, et notamment pour l'unité *haut*. Ainsi, si cette unité prise hors contexte ne semble pas particulièrement renvoyer à la hauteur hors norme des instruments, les contextes de cette unité lorsqu'elle modifie le terme *détecteur* indiquent pourtant, dans le SC-P, que leur hauteur est tout de même considérable, puisqu'elle est comparée à la hauteur de cathédrales, comme dans l'exemple 165. En revanche, les contextes issus du SC-R sont beaucoup plus neutres (exemple 166).

165. L'instrument - le plus grand et plus complexe du monde, devisé à 4,7 milliards de francs - aura pour but principal de révéler, avec ses quatre **détecteurs** *hauts* comme des cathédrales (ALICE, ATLAS, CMS et LHCb), l'élusif "boson de Higgs". (SC-P, Le Temps, 2007)

166. Le **détecteur** ALICE (*haut* de 16 m, long de 26 m et pesant environ 10 000 tonnes) est constitué d'une partie centrale qui sert à mesurer les hadrons, les électrons et les photons, et un spectromètre de mesure des muons. (SC-R, CERN, 2005)

Au vu de ces différentes observations, il semble que la notion de sensationnalisme soit bien évoquée dans tous les sous-corpus intermédiaires. Ce constat permet de soutenir l'hypothèse d'un rôle particulier des sous-corpus intermédiaires dans le processus de déterminologisation, de la même manière que dans le Chapitre 6, notamment pour ce qui est du traitement sensationnaliste de sujets liés à la recherche en physique des particules. Ainsi, contrairement à ce qui pourrait être conclu à partir de la littérature, le sensationnalisme n'est pas une caractéristique de la presse, ou des médias généralistes, uniquement, du moins pas à travers les unités repérées dans notre travail, étant donné que ces unités sont bien attestées dans les sous-corpus intermédiaires. Dans le Chapitre 8, nous reviendrons plus en détail sur le rôle des intermédiaires dans le processus de déterminologisation, tel que nous pouvons l'observer dans nos données.

7.3.3. Répartition des différents aspects des concepts dans les sous-corpus intermédiaires

Dans la section 7.2.2, nous rapprochons certaines différences de distribution dans le SC-P par rapport au SC-S à la mise en évidence de différents aspects des concepts. L'exploration des unités retenues dans les sous-corpus intermédiaires nous permet ici de distinguer deux tendances, selon si ces aspects sont aussi présents dans les sous-corpus intermédiaires ou non.

Aspects observés dans les sous-corpus intermédiaires. Dans un premier temps, nous remarquons que la majorité des aspects qui touchent globalement aux instruments désignés par les termes *accélérateur*, *collisionneur* et *détecteur* sont présents dans les sous-corpus intermédiaires. Ainsi, sont également attestés dans ces sous-corpus :

- les modifieurs indiquant l'appartenance à un pays ou à une région, à savoir *américain*, *européen* et *français* (dans les SC-R et SC-V) ;
- le verbe *abriter* lorsque les termes sont objets ;
- et les verbes *inaugurer* et *visiter*, dans les SC-C et SC-R uniquement.

Dans le cas d'*inaugurer*, les contextes démontrent également un caractère officiel (exemple 167), comme nous le remarquons dans le SC-P. Notons que le terme concerné dans cet exemple est *Grand collisionneur de hadrons*. Dans le cas de *visiter*, les contextes semblent renvoyer à plusieurs actions de valorisation ou de vulgarisation impliquant le grand public (exemples 168 et 169), de la même manière que dans le SC-P.

167. Le Président de la Confédération suisse, Pascal Couchepin, et le Premier ministre français, François Fillon, seront présents demain au CERN, entourés de ministres de la recherche des États membres du CERN et du monde entier, pour *inaugurer* le **Grand collisionneur** de hadrons (LHC), l'instrument scientifique le plus grand et le plus complexe du monde, qui a fait circuler son premier faisceau le 10 septembre 2008. (SC-C, CERN, 2008)
168. Plus de 70 000 personnes sont venues aux journées portes ouvertes du CERN, les 28 et 29 septembre. Parmi elles, 20 000 ont pu *visiter* sous terre les **accélérateurs** et les **détecteurs**. (SC-R, CERN, 2013)
169. Par ailleurs, en février, plus de 600 élèves d'établissements scolaires voisins ont *visité* le **détecteur** CMS. (SC-R, CERN, 2015)

Dans un second temps, nous remarquons que les différentes thématiques mises en évidence par l'attestation de certains verbes dans le SC-S et non dans le SC-P, et vice versa, s'observent également dans les trois sous-corpus intermédiaires. Dans la section 7.2.2.2, nous dégageons trois tendances, à propos des termes *particule*, *neutrino* et *boson de Higgs*, des termes *électron* et *proton*, et du

terme *quark*. Cependant, afin d'illustrer notre propos, nous ne prenons ici que le cas du terme *particule* et des verbes lorsque ce terme est objet.

Le tableau 7.15 regroupe la fréquence brute et la répartition de ces verbes dans les cinq sous-corpus, ce qui permet de constater qu'en réalité, seule une minorité d'unités ne sont attestées que dans le SC-P ou que dans le SC-S.

Verbes lorsque <i>particule</i> est objet	SC-S	SC-C	SC-R	SC-V	SC-P
trouver	-	2	2	10	8
traquer	-	2	3	1	6
débusquer	-	-	-	-	3
dénicher	-	-	-	-	2
attraper	-	-	-	1	2
chasser	-	-	-	1	2
piéger	-	-	2	2	2
reconstruire	4	1	1	2	-
caractériser	2	-	1	1	-
comprendre	2	2	1	3	-
définir	2	-	-	-	-
classer	1	-	-	3	-

Tableau 7.15 : Répartition et fréquence absolue dans les sous-corpus d'une sélection de verbes lorsque *particule* est objet

Si nous considérons que les verbes attestés dans le SC-P, et non dans le SC-S, lorsque *particule* est objet représentent une thématique particulière liée à ce concept, alors la répartition de ces verbes dans les sous-corpus intermédiaires indique que cette thématique n'est pas réellement caractéristique du SC-P. Le même constat peut être effectué pour la thématique mise en évidence par les verbes présents dans le SC-S mais absents du SC-P. Comme nous le remarquons dans le tableau 7.15, ces verbes sont de manière générale également attestés dans les sous-corpus intermédiaires. Ces éléments permettent ainsi de réaffirmer la pertinence de ces sous-corpus pour l'observation du processus de déterminologisation (nous y revenons au Chapitre 8).

Ces remarques s'appliquent également au cas de *boson de Higgs*, pour lequel il semble que la conceptualisation change en fonction du modèle théorique pris en considération. Ainsi, comme nous le montrons *supra* (dans le tableau 7.4 particulièrement), les hyponymes possibles de *boson de Higgs* sont attestés dans les sous-corpus intermédiaires, ce qui suggère que les conceptualisations propres à certains modèles théoriques que nous observons dans le SC-S sont aussi présentes dans

les sous-corpus intermédiaires. C'est ce que les exemples 170 à 172 ci-dessous permettent de constater, à partir des unités *seul* et *unique*.

170. S'agit-il du *seul* et *unique* **boson de Higgs** du Modèle standard, du premier élément d'une famille d'objets de type Higgs, ou d'une tout autre particule ? (SC-R, CERN, 2013)
171. Toutefois, la supersymétrie ne prévoit pas un *seul* **boson de Higgs**, mais au moins cinq, dont le plus léger ressemble au boson du modèle standard. (SC-V, Pour la Science, 2012)
172. Dans la version la plus simple de la supersymétrie, appelée Modèle Standard Supersymétrique Minimal ou MSSM, on devrait observer cinq bosons de Higgs [...] au lieu de l'*unique* **boson de Higgs** neutre du Modèle Standard (qui manque toujours à l'appel). (SC-V, Élémentaire, 2010)

Aspects spécifiques au SC-P. L'exploration des sous-corpus intermédiaires dans cette section montre que seule une minorité d'aspects sont réellement spécifiques au SC-P. En réalité, seuls les verbes *justifier* et *financer* lorsque le terme *accélérateur* est objet et le verbe *coûter* lorsqu'il est sujet sont attestés dans le SC-P uniquement. Ces verbes uniquement renverraient à un aspect qui apparaîtrait dans le SC-P. Cette observation nous permet en outre de revenir sur les constats effectués au Chapitre 6 (6.3.2.1), à partir de l'analyse des unités classées dans la catégorie « économie et finances », et de les affiner. En effet, dans cette section nous avons constaté que les unités de cette catégorie sont réparties dans tous les sous-corpus intermédiaires, et notamment dans le SC-R. En revanche, nous avons remarqué que les objets des financements évoqués ne sont pas nécessairement les mêmes entre le SC-R et le SC-P. L'observation des contextes des verbes *justifier*, *financer* et *coûter* avec le terme *accélérateur* (en réalité le plus souvent *accélérateur de particules*) apporte toutefois un élément supplémentaire. Il semble en effet que, bien que la thématique financière soit présente dans les sous-corpus intermédiaires de la même manière que dans le SC-P, comme nous le montrons au Chapitre 6, l'aspect concernant le besoin de justification des financements, et en particulier des importantes sommes investies, que nous relevons dans le SC-P est réellement caractéristique du SC-P.

7.4. Bilan

Les observations détaillées dans ce chapitre révèlent principalement la diversité des fonctionnements des termes dans le corpus et la diversité des phénomènes associés à la déterminologisation. Différents points méritent d'être soulevés. Tout d'abord, la richesse des fonctionnements des termes dans le SC-P, et notamment les précisions apportées dans ce chapitre par rapport au précédent, permet de confirmer la complémentarité des deux indices distributionnels

mis en œuvre. En effet, dans le Chapitre 6, différents phénomènes de nature thématique sont particulièrement mis en évidence. Or, dans ce chapitre-ci, le fonctionnement des termes dans le corpus est appréhendé de manière plus fine, à l'aide des unités qui entretiennent avec eux une relation syntaxique. Ce point permet ainsi de repérer différents termes complexes et différentes collocations spécialisées, dont la répartition dans les sous-corpus varie, selon les caractéristiques des genres textuels dont relèvent les textes des sous-corpus et, dans certains cas, selon le domaine auquel ils font référence. Mais il nous permet surtout de montrer que certaines différences de distribution reflètent en réalité différents aspects des concepts, sur lequel l'un ou l'autre sous-corpus semble insister, voire différentes manières de conceptualiser le même concept, selon l'orientation théorique (comme dans le cas de *boson de Higgs*, cf. 7.2.1.1 et 7.3.3) ou selon le domaine.

Ensuite, parmi les phénomènes repérés, nos observations permettent surtout de confirmer plusieurs hypothèses formulées dans le Chapitre 6, notamment à propos des liens étroits entre la physique des particules et d'autres domaines. Ainsi, il semble bien que le passage d'un terme dans la langue générale puisse être influencé par les différents domaines dans lesquels ce terme est utilisé. D'une manière similaire, certains intermédiaires semblent avoir une plus grande importance que d'autres. Nous remarquons (en 7.1 et 7.3.2) que les fonctionnements des termes dans le SC-P sont particulièrement proches de leurs fonctionnements dans le SC-V, ce qui interpelle sur les rapports qu'entretiennent la presse et la vulgarisation, notamment leur rôle dans la transmission des connaissances spécialisées vers des locuteurs non spécialistes.

Enfin, ce chapitre permet encore une fois de confirmer la pertinence de la notion de sensationnalisme pour appréhender les changements sémantiques se produisant au cours du processus de déterminologisation. Celle-ci s'observe dans le SC-P et dans les sous-corpus intermédiaires par le biais de différentes unités, qui sont essentiellement des adjectifs axiologiques selon les modificateurs identifiés comme tels dans ce chapitre. La récurrence avec laquelle ces adjectifs modifient les termes, le plus souvent dans le SC-P mais également dans les sous-corpus intermédiaires, peut, selon nous, être responsable de (ou du moins influencer dans une large mesure) l'apparition de certaines connotations dans le sens des termes, en particulier pour ce qui est des aspects liés au mystère et à l'extraordinaire à propos de la recherche de particules et des performances des instruments de recherche (cf. 7.2.1.2 et 7.3.2).

Tous ces éléments mettent en lumière différentes manières d'approfondir notre réflexion sur la déterminologisation. Dans le chapitre suivant, nous montrons comment nos observations contribuent à mieux saisir la notion de déterminologisation, à travers trois axes de description.

Chapitre 8 Vers une meilleure compréhension du processus de déterminologisation

Dans ce chapitre, nous tirons profit des observations détaillées dans les deux chapitres précédents pour développer davantage la réflexion sur la déterminologisation, en tant que processus et résultat. Nous montrons en particulier comment la diversité des phénomènes relevés dans le sous-corpus *Presse* et dans les sous-corpus intermédiaires permet d'enrichir les descriptions existantes de la déterminologisation et d'apporter, dans le même temps, un éclairage nouveau et approfondi sur cette question.

Nous articulons notre réflexion autour de trois axes : le premier concerne principalement les fonctionnements des termes dans le SC-P. Nous expliquons d'abord comment les notions de métaphore et de multidimensionnalité permettent d'appréhender ces fonctionnements, puis nous montrons comment la richesse de ces descriptions contribue à redéfinir la place de la déterminologisation dans les mécanismes de création néologique. Le deuxième axe se focalise sur la diversité des intermédiaires intervenant potentiellement dans le processus de déterminologisation, qu'il s'agisse des intermédiaires identifiés dans notre travail et représentés par le biais des sous-corpus intermédiaires ou d'autres types d'intermédiaires mis en évidence dans les données. Enfin, le troisième axe concerne plus particulièrement la définition de la déterminologisation comme un processus non linéaire de passage de termes vers la langue générale et la mise au jour de facteurs de déterminologisation.

8.1. Saisir la richesse des phénomènes sémantiques en jeu

L'exploration des deux indices de type distributionnel définis pour notre étude a permis de mettre au jour la diversité des fonctionnements des termes de physique des particules dans le SC-P en comparaison avec le SC-S. Cette diversité peut être mise en relation avec les travaux existants sur la déterminologisation, notamment lorsqu'il est question des usages métaphoriques des termes. En revanche, les phénomènes plus subtils mis en évidence particulièrement dans le Chapitre 7 sont moins décrits et amènent de nouvelles interrogations.

Dans cette section, nous cherchons donc à approfondir notre réflexion en questionnant la place de la métaphore dans le processus de déterminologisation (dans la section 8.1.1) et en appréhendant la diversité des fonctionnements dans le SC-P à travers la notion de multidimensionnalité (dans la section 8.1.2). Enfin, nous montrons dans la section 8.1.3 que la richesse des phénomènes sémantiques se produisant au cours du processus interroge les rapports

que la déterminologisation entretient avec la néologie. À travers l'exemple du terme *accélérateur de particules*, nous insistons sur la pertinence d'étudier ces rapports et notamment la place de la déterminologisation dans la question globale de la néologie.

8.1.1. Préciser la place de la métaphore dans la déterminologisation

Dans le Chapitre 6, nous avons repéré différents usages métaphoriques pour les termes *matière noire* et *accélérateur de particules* (et dans une moindre mesure *anti-matière*) dans le SC-P, qui renvoient au deuxième type de changement sémantique identifié par Meyer et Mackintosh (2000a : 205^{sqq.} ; cf. Chapitre 2, 2.1.1.3). L'exploration fine des contextes d'apparition de ces termes dans le SC-P permet de mettre au jour les fonctionnements sous-tendant la création d'un potentiel métaphorique des termes dans ce sous-corpus. Nos observations nous amènent également à interroger un autre aspect, qui concerne plus particulièrement les difficultés relatives à l'analyse des emplois métaphoriques des termes dans ce sous-corpus.

8.1.1.1. Création du potentiel métaphorique

L'observation des différents contextes mis en évidence par le premier indice distributionnel dans le Chapitre 6 permet d'interroger le fonctionnement de la métaphore. Il semble en particulier qu'un lien puisse être établi entre les unités de la catégorie « sensationnalisme » et les usages métaphoriques des termes *matière noire* et *accélérateur de particules*. Ce lien nous amène à définir le potentiel métaphorique des termes.

Par exemple, lorsque le terme *matière noire* est employé de manière métaphorique, celui-ci semble désigner quelque chose de mystérieux, d'inconnu, voire d'invisible ou de quasi inexistant. Ces aspects sont par ailleurs mis en évidence par la cooccurrence récurrente de ce terme (lorsqu'il n'est pas employé métaphoriquement) avec des unités telles que *mystère*, *Arlésienne*, *secret*, *énigmatique* ou encore *mystérieux*. Ainsi, si nous tenons compte d'une définition de la métaphore comme un transfert par analogie à partir de traits sémantiques communs (Neveu, 2004), alors nous pouvons déduire que, lorsque *matière noire* est métaphorique, la métaphore est créée à partir des traits /mystère/ et /inconnu/, qui sont actualisés dans les exemples 173 et 174 ci-dessous et dans les contextes où *matière noire* est métaphorique (exemples 56, 57 et 59 à 61 au Chapitre 6, 6.2.7). Dans cette perspective, il semble que la cooccurrence récurrente de ce terme avec certaines unités de la catégorie « sensationnalisme » crée effectivement la possibilité d'un fonctionnement métaphorique.

173. En point de mire : la quête de réponses aux mystères de l'Univers concernant l'existence de particules inédites, la *fantomatique* "**matière noire**" [...] (SC-P, Le Temps, 2015)
174. Dès lors, il ne serait plus nécessaire d'invoquer la présence d'une *mystérieuse* et introuvable **matière noire** pour équilibrer le système. (SC-P, Le Figaro, 2006)

Un raisonnement similaire peut être suivi pour le terme *accélérateur de particules*. Celui-ci est en effet utilisé de manière métaphorique pour désigner quelque chose, ou quelqu'un, de particulièrement rapide et/ou performant. Lorsque le terme désigne quelque chose ou quelqu'un de performant, il semble que la métaphore s'appuie sur les différents contextes où *accélérateur de particules* cooccure avec des unités évoquant un caractère extraordinaire ou hors norme. Par exemple, dans le contexte 175 ci-dessous, nous observons qu'*accélérateur de particules* cooccure avec l'unité *puissant*. D'autres contextes montrent que ce terme cooccure également avec les unités *performant* et *record* notamment, comme dans les exemples 176 et 177. Par ailleurs, l'attestation du superlatif dans *le plus grand* | *puissant* | *performant* *accélérateur de particules* ajoute un aspect extraordinaire et hors norme supplémentaire. Dans ces contextes, les traits /*puissant*/, /*performant*/ et /*record*/ sont actualisés et nous pouvons donc supposer que les usages métaphoriques où ce terme désigne quelque chose ou quelqu'un de performant s'appuient précisément sur ces traits.

175. Voilà le plus grand, complexe et puissant **accélérateur de particules** jamais construit. Un outil qui servira à décrypter les codes secrets de la matière [...] (SC-P, Le Temps, 2008)
176. Le plus grand **accélérateur de particules** du monde (LHC) a établi vendredi un nouveau record du monde [...] (SC-P, Le Temps, 2011)
177. [...] la mise en service du LHC, appelé à devenir l'**accélérateur de particules** le plus performant du monde. (SC-P, Le Temps, 2006)

En revanche, une autre hypothèse peut être formulée pour les contextes d'*accélérateur de particules* où la métaphore désigne quelque chose ou quelqu'un de particulièrement rapide. En effet, Oliveira (2009 : 93sq.) remarque que, lorsque des termes métaphoriques s'intègrent dans la langue générale, les unités qui les composent et qui existent indépendamment dans la langue générale ont une certaine influence sur les changements sémantiques survenant dans le processus. Il semble en particulier que le sens de ces unités dans la langue générale interfère avec l'usage terminologique de ces composés et que leur « sens primitif se maintient sous forme de trace connotative » (*Ibid.* : 94).

Oliveira donne notamment l'exemple du terme *souffle au cœur* :

« [n]ous remarquons grâce à cette enquête que le souffle au cœur est ancré dans la mémoire de beaucoup de personnes comme le signe d'une maladie grave du cœur. Or, un cœur malade, d'après eux, ne doit pas se fatiguer. Partant de ce

présupposé, lorsqu'on est atteint d'un souffle au cœur il faut faire le moins d'efforts possible et donc surtout éviter le sport. » (*Ibid.* : 101)

L'autrice poursuit et explique que « l'utilisateur commun s'arrête au premier sens de "souffle" sans se soucier de l'analogie qui se cache derrière ce mot » (*Ibid.* : 102). Perdure ainsi l'idée que « lorsqu'on souffre d'un problème de "souffle au cœur" on ne peut pas pratiquer de sport dans la mesure où on s'essouffle très rapidement » (*Ibid.*). Dans cet exemple, nous constatons que ce sont bien, d'une part, le sens de *souffle* dans la langue générale et, d'autre part, les liens morphologiques et sémantiques entre *souffle* et *s'essouffler* qui interfèrent dans le sens de *souffle au cœur* dans un contexte non spécialisé.

Meyer et Mackintosh font un constat similaire. Les autrices expliquent que, pour les termes qui ont été créés à partir d'unités de la langue générale, le sens « général » de ces unités peut intervenir dans les changements sémantiques de ces termes dans le processus de déterminologisation, car ce sens « général » coexiste nécessairement avec le sens « terminologique » dans un contexte non spécialisé (Meyer et Mackintosh, 2000b : 130-131).

Ainsi, bien qu'*accélérateur de particules* ne soit pas un terme métaphorique, il a en commun avec le terme cité par Oliveira d'être composé d'unités existant indépendamment dans la langue générale. Nous pensons alors que les « traces connotatives » mentionnées par l'autrice pourraient contribuer à créer la possibilité d'un fonctionnement métaphorique (avec par exemple *vitesse* ou *grande vitesse*), de même que la coexistence de ce terme avec les unités le composant dans la langue générale et les éventuelles connotations qui lui sont associées, comme le soulignent Meyer et Mackintosh.

De même, nous pouvons faire l'hypothèse que la coexistence non seulement de l'unité *accélérateur* dans la langue générale, mais également d'unités appartenant à la même famille morphologique, telles qu'*accélération* et *accélérer*, peuvent avoir une influence sur le sens d'*accélérateur de particules* qui est véhiculé dans le SC-P (sur le même modèle que *souffle au cœur* et *s'essouffler* dans l'exemple d'Oliveira) et contribuent probablement aux métaphores observées. En effet, lorsque le terme est utilisé de manière métaphorique pour désigner quelque chose ou quelqu'un de particulièrement rapides et/ou qui rend quelque chose plus rapide, les métaphores semblent avoir été créées à partir du trait /vitesse/ (ou /grande vitesse/), que nous retrouvons dans *accélérateur*, *accélération* et *accélérer* d'après leurs définitions dans le Petit Robert en ligne¹⁶⁶, reproduites ci-dessous :

¹⁶⁶ Le Petit Robert 2020, en ligne, <https://petitrobert.lerobert.com/robert.asp> (page consultée le 11 septembre 2020).

- *accélérateur* : « [o]rgane qui commande l'admission du mélange gazeux au moteur (l'admission accrue augmente la vitesse). »
- *accélération* : « [a]ugmentation de vitesse. »
- *accélérer* : « [r]endre plus rapide. »

Ces deux cas de figure, *accélérateur de particules* et *matière noire*, permettent donc de montrer que, lorsqu'un terme est utilisé de manière métaphorique dans un contexte non spécialisé, la métaphore peut être créée aussi bien à partir de la coexistence des composants du terme et du terme dans la langue générale que sur la cooccurrence récurrente du terme avec certaines unités, qui véhiculent des connotations particulières.

À partir de ces éléments, nous proposons de définir le potentiel métaphorique des termes comme la possibilité pour un terme d'être employé métaphoriquement. Selon nous, cette possibilité dépend de la morphologie du terme et de ses usages dans des textes non spécialisés (notamment médiatiques), en particulier des connotations qui peuvent être ajoutées.

8.1.1.2. Difficultés spécifiques à l'analyse des métaphores

Les réflexions détaillées *supra* s'appuient essentiellement sur l'analyse poussée de deux termes, qui sont les seuls pour lesquels plusieurs occurrences d'usages métaphoriques ont été repérées avec les indices mis en œuvre. Cependant, d'autres usages métaphoriques pour d'autres termes sont attestés dans le SC-P, notamment *anti-matière*, dont nous montrons un exemple dans le Chapitre 6 (exemple 57 ; 6.2.7), ou *particule élémentaire* dans l'exemple ci-dessous.

178. Marielle de Sarnez la **particule élémentaire** des centristes (SC-P, Le Monde, 2007)

La difficulté principale réside toutefois dans le repérage de ces emplois métaphoriques. En effet, les indices distributionnels mis en œuvre dans notre travail reposent sur la récurrence des décalages de distribution observables dans le SC-P par rapport au SC-S, alors que les décalages permettant de repérer les métaphores ne sont le plus souvent pas récurrents. Seule l'observation de chaque contexte peut confirmer qu'une occurrence est métaphorique. En outre, certains indices lexicaux n'indiquent pas systématiquement la présence d'une métaphore. Par exemple, l'unité *livre* permet de repérer des usages métaphoriques pour les termes *matière noire* et *anti-matière*, dans l'exemple 179. Cependant, il serait incorrect de supposer que, lorsque des termes de physique des particules coexistent avec cette unité dans un contexte littéraire (ou avec d'autres unités évoquant un contexte littéraire), alors l'usage de ces termes est métaphorique. C'est ce qu'illustre l'exemple 180 ci-dessous, où *livre* cooccure avec le terme *Cern*. Sans tenir compte d'une observation

minutieuse des contextes du corpus, il serait aisé de déduire que l'exemple 179 est un autre cas où des termes de physique des particules s'intègrent dans un contexte littéraire, à l'instar de l'exemple 180. Corrélativement, il serait également aisé de déduire que ces deux exemples relèvent d'usages métaphoriques de ces termes. Pourtant, les usages qui sont faits de *matière noire*, *anti-matière* et *Cern* sont de natures tout à fait différentes.

179. Quand on écrit des nouvelles, il y a cet avant et cet après que l'on dit pas mais que l'on suppose. Ce n'est pas écrit, et pourtant cela compte : c'est la **matière noire** de la nouvelle. Son **anti-matière**. Sa narration invisible. Dans mon *livre*, je voulais que l'enchaînement des chapitres rende visibles cet avant et cet après. (SC-P, Le Monde, 2009)
180. Comment croire à l'avion hypersonique du **Cern**, fusée qui amène en 64 minutes le héros du *livre* de Boston à Genève ? (SC-P)

Ces observations renvoient à une difficulté connue des chercheurs s'intéressant aux métaphores et se situant dans une approche en corpus. En effet, l'identification des métaphores (ou la validation des métaphores potentielles identifiées automatiquement) repose essentiellement sur l'interprétation des contextes par l'analyste (Deignan, 2005 : 92-93 ; Stefanowitsch, 2006 : 64, 2020 : 397*sqq.* ; Philip, 2010 : 191 ; Mpouli, 2019 : 98*sqq.*). Cette difficulté est illustrée notamment par Deignan, qui montre que la même unité apparaissant dans la même collocation peut renvoyer aussi bien à un usage littéral qu'à un usage métaphorique (Deignan, 2005 : 83). De la même manière, Stefanowitsch précise que « there is nothing in the word itself that distinguishes its literal and metaphorical uses » (Stefanowitsch, 2020 : 397). Dans ce cas, l'identification repose d'abord sur les décalages de distribution observés entre plusieurs occurrences d'une unité, que ce soit à partir d'unités de champs lexicaux différents de celui de l'unité sur laquelle porte l'analyse (Philip, 2010 : 191*sqq.* ; Mpouli, 2019 : 92-94 ; Stefanowitsch, 2020 : 397) ou à partir d'une analyse distributionnelle automatique (à condition que les écarts de distribution soient suffisamment récurrents et stables, comme l'expliquent Del Tredici *et al.* (2019)), puis sur une vérification en contexte.

Dans nos données, même lorsque la présence de certaines unités peut parfois constituer un indice d'usages métaphoriques, seule l'observation des contextes permet de confirmer qu'il s'agit bien d'une métaphore. De ce point de vue, l'analyse des usages métaphoriques des termes dans le SC-P reste un défi, de par la difficulté à les repérer systématiquement de manière outillée, avec la méthode définie dans notre travail. Cette méthode devrait donc être améliorée, afin de repérer un plus grand nombre d'occurrences de termes utilisés métaphoriquement et afin de mieux caractériser les fonctionnements liés à ces usages métaphoriques.

8.1.2. Aborder la déterminologisation à travers la multidimensionnalité

Dans les deux chapitres précédents, nous avons montré que les fonctionnements des termes dans le SC-P révèlent l'importance de certains termes pour plusieurs domaines (médecine, énergie nucléaire, spatial, arts et culture), dans des contextes explicitant l'interdisciplinarité de la recherche ou non. Nous avons également constaté certaines nuances dans l'usage des termes selon le domaine considéré, voire des points de vue différents, qui sont susceptibles de coexister dans le SC-P, ou même d'être neutralisés, pour ne conserver qu'un sens superficiel. Cette remarque fait écho à ce que Meyer et Mackintosh (2000b : 134) décrivent comme « the growing *multi-disciplinarity* of knowledge development, and *convergence* of disciplines ». Ainsi, d'un point de vue sémantique, « the meaning of a term in one domain may be coloured by its uses in others » (*Ibid.*) et tous ces usages, avec toutes les nuances qu'ils comportent, sont susceptibles de participer à la construction du sens dans le SC-P. De la même manière, nous avons montré que différents aspects sont mis en évidence par la distribution des termes dans le SC-P, en comparaison avec le SC-S. Selon nous, ces aspects peuvent également révéler l'existence de différents points de vue, en fonction du sous-corpus.

Ces descriptions font écho à la notion de multidimensionnalité qui, de manière générale, renvoie à la possibilité pour un même concept d'être classé de différentes manières selon les caractéristiques prises en compte. Dans cette section, nous montrons que la notion de multidimensionnalité permet à la fois d'enrichir la description du fonctionnement des termes dans le SC-P et d'alimenter la réflexion sur la déterminologisation.

8.1.2.1. Définition de la multidimensionnalité et intérêt pour notre étude

Comme le dit Péraldi, « il est possible d'aborder les choses sous différents angles, plusieurs perspectives étant possibles selon les personnes et les points de vue adoptés » (Péraldi, 2011 : 54). L'autrice précise que cette possibilité est d'ailleurs très bien connue en terminologie. Différentes notions permettent d'en rendre compte : par exemple, Condamines et Rebeyrolle (1997) proposent d'appréhender la question de la polysémie en langue de spécialité à travers la notion de point de vue. Afin de montrer la pertinence de cette proposition, les autrices procèdent à l'analyse d'un corpus composé de textes provenant de deux divisions du Cnes¹⁶⁷ (c'est-à-dire des textes rédigés par et pour deux groupes d'experts distincts). Leur analyse est appuyée par une collaboration avec des experts de domaine et permet de mettre en lumière l'existence de différents points de vue à propos de *satellite* (*Ibid.* : 181-182). Ainsi, un « satellite peut être considéré comme un *corps artificiel* »

¹⁶⁷ Centre national d'études spatiales, <https://cnes.fr/fr> (page consultée le 27 novembre 2020).

ou « comme un *mobile* en ce sens qu'il constitue un corps qui peut être mû, dont on peut changer la position », ou encore comme une « *plate-forme* [...] par métonymie puisque la plate-forme est une partie du satellite » (*Ibid.*). En outre, si certains de ces points de vue sont uniformément répartis dans les textes provenant de chacune des deux divisions, et peuvent donc être considérés comme communs à tous les groupes d'experts du Cnes, d'autres semblent plus spécifiques à l'un ou l'autre des deux groupes considérés (*Ibid.* : 182-183).

Une deuxième notion, la variation conceptuelle, est développée notamment par Freixa et Fernández-Silva (2017 : 160*sqq.*). Selon les autrices, la variation conceptuelle en terminologie se caractérise par un changement de point de vue menant à des conceptualisations différentes pour le même référent. Elles expliquent que « the process through which reality is perceived, structured in the form of concepts and conceptual structures, and finally named by linguistic units, is marked by diversity » (*Ibid.* : 166). Un tel changement peut se refléter dans la dénomination, c'est-à-dire qu'une certaine variation au niveau du concept peut s'observer par la coexistence de plusieurs variantes dénominatives. Dans ce cas, les autrices renvoient aux notions d'instabilité terminologique ou d'imprécision conceptuelle, qui s'illustrent par exemple par un foisonnement terminologique lorsqu'un nouveau concept doit être nommé (comme dans les études de cas de Dury et Lervad (2008)). Cependant, d'après Freixa et Fernández-Silva, la variation conceptuelle ne mène pas systématiquement à la coexistence de plusieurs variantes dénominatives et, dans ce cas, la variation serait repérée à partir du contexte d'utilisation du terme, au sens large, c'est-à-dire dans des textes relevant de différents domaines ou de différentes orientations théoriques, ou encore dans des textes appartenant à des genres différents ou relevant de degrés de spécialisation différents. Comme l'expliquent Freixa et Fernández-Silva (2017 : 166), « the variety of situations in which specialized knowledge is activated and exchanged explains differences in the way concepts are structured in the mind of individuals ».

Soulignons que cette notion de variation conceptuelle figure également dans les travaux de Pelletier (2012) et de Kim (2017). La première explique que « dans la variation conceptuelle, la modification de sens s'effectue à partir de la perception, de la conception en fonction des objectifs des sujets ciblés contextuellement ». C'est donc le « signifié qui varie en fonction de la perception du locuteur, sans qu'il y ait changement de référent » (Pelletier, 2012 : 43). De même, Kim (2017 : 42) affirme que « [l]a variation conceptuelle correspond à l'existence de plusieurs signifiés liés à un même signifiant et à un même référent », qui peut « résulte[r] d'une différence de perception du locuteur dans une communauté linguistique donnée ».

Une troisième notion appréhende ces phénomènes d'une manière similaire. Il s'agit de la multidimensionnalité. À l'origine, cette notion est proposée par Bowker (1993) pour rendre compte

du fait qu'il n'existe pas une seule manière claire et univoque de classer les concepts d'un domaine. Au contraire, selon le type de caractéristiques qui détermine la classification, différentes manières d'organiser les concepts les uns par rapport aux autres sont possibles. Pour Bowker, une dimension « represents one way of classifying a group of concepts » (*Ibid.* : 41-42). Aussi, lorsqu'un groupe de concepts peut être classé de plusieurs manières, ces concepts sont considérés comme multidimensionnels. À ce propos, l'auteur donne l'exemple du concept WINE (en anglais), dont les différents types peuvent être classés selon la couleur (RED WINE, WHITE WINE), selon le taux de sucre (DRY WINE, SWEET WINE) ou encore selon le pays d'origine (FRENCH WINE, SPANISH WINE). Cette conception de la multidimensionnalité est notamment partagée par Kageura (1997) et Rogers (2004). Par ailleurs, Bowker s'intéresse à la multidimensionnalité des concepts dans une perspective visant majoritairement à améliorer les ressources terminographiques, pour mieux tenir compte de ce phénomène, en proposant des moyens concrets de le représenter dans les ressources (Bowker, 1993, 1997a).

D'autres auteurs explorent cette notion, à partir des propositions de Bowker notamment, et dans plusieurs domaines, par exemple en chimie organique (Péraldi, 2011), en sciences de la terre, en océanographie, en psychologie, en histoire (Fernández-Silva, 2016), en écologie (Tercedor, 2011) ou encore en psychiatrie (León-Araúz et Reimerink, 2014). Ces travaux montrent que l'existence de différents types de caractéristiques propres à un concept ou à un groupe de concepts a non seulement des conséquences sur l'organisation de ces concepts dans un système conceptuel, mais également sur le choix des termes utilisés. Les dénominations choisies peuvent ainsi refléter la valorisation d'une dimension par rapport à d'autres et plusieurs dénominations peuvent coexister selon les dimensions représentées (Bowker, 1997b : 285). Partant de ce constat, la question de la multidimensionnalité permet de nourrir la réflexion sur la variation dénominative en terminologie et plusieurs travaux se penchent en particulier sur les liens entre multidimensionnalité et variation dénominative (Fernández-Silva *et al.*, 2011 ; Fernández-Silva, 2013, 2018 ; León-Araúz, 2015, 2017 ; Rogers, 2018). De manière générale, il ressort de ces travaux que chaque variante, en insistant sur une dimension précise du concept qu'elle dénomme, dénote un point de vue particulier ou une certaine nuance dans la manière d'utiliser le concept (Péraldi, 2011 : 59 ; Fernández-Silva *et al.*, 2011 : 52).

Parallèlement, L'Homme propose de distinguer deux phénomènes relatifs à la multidimensionnalité. Selon l'auteur, la multidimensionnalité « can first be viewed as a classification challenge whereby concepts are organized differently according to the characteristic that is taken in consideration » (L'Homme, 2020 : 88), reprenant ainsi la notion telle qu'elle a d'abord été développée par Bowker. Le second phénomène renvoie au fait que « the same concept can be

conceptualized from different perspectives » (*Ibid.*). L'Homme inclut ici aussi bien les cas où les différents points de vue sont associés à différents domaines que les cas où un concept peut posséder plusieurs dimensions dans le même domaine. Selon elle, ce second cas est délicat à appréhender, particulièrement dans l'optique de constituer des ressources terminographiques, étant donné qu'il est très difficile de représenter formellement un point de vue ou une perspective (*Ibid.* : 90).

Ainsi, bien que ces trois notions – multidimensionnalité, point de vue, variation conceptuelle – permettent d'appréhender des phénomènes similaires, nous faisons le choix de poursuivre avec la multidimensionnalité, notamment à partir des propositions de L'Homme (2020). En effet, ce qui rend ses propositions particulièrement intéressantes pour notre propos réside dans la possibilité de considérer la centralité d'un terme dans des domaines différents par le prisme de la multidimensionnalité (cf. 8.1.2.2). Nous retenons en outre des travaux cités dans cette section que la multidimensionnalité sert à appréhender des changements qui peuvent dépendre des points de vue caractéristiques à différents groupes d'experts dans un même domaine, des spécificités des différents genres textuels, ou encore des degrés de spécialisation de situations de communication données. Dans ce contexte, la multidimensionnalité permet alors de caractériser les différents aspects mis en évidence par des différences de distribution dans le SC-S et dans le SC-P ainsi que les différents points de vue susceptibles d'être portés sur un terme selon le domaine par lequel il est appréhendé.

L'exploitation de la multidimensionnalité pour appréhender les différents phénomènes qui se produisent au cours du processus de déterminologisation amène cependant une remarque : cette notion est le plus souvent mise en œuvre dans un contexte de terminologie et de terminographie, où les termes sont analysés dans des textes spécialisés et où l'accent est porté sur les concepts et leur classification à l'intérieur d'un domaine précis, en vue de leur description dans des ressources. Or la déterminologisation est un processus qui s'effectue dans le continuum entre langue de spécialité et langue générale et dont les conséquences sur le sens des termes s'observent dans des textes non spécialisés tels que les articles de presse généraliste inclus dans le SC-P. Dans cette perspective, il peut sembler paradoxal de mettre en œuvre cette notion pour discuter des phénomènes constatés dans des textes non spécialisés, qui ne relèvent donc pas d'un domaine en particulier. Cependant, même si nous reconnaissons le caractère peu commun de cette proposition, nous pensons qu'elle se justifie pour au moins trois raisons.

Premièrement, l'une des causes de la multidimensionnalité dépend du degré de spécialisation des textes servant à l'analyse¹⁶⁸. De manière générale, cette affirmation renvoie à des textes oscillant entre une situation de communication spécialisée telle que le représente le SC-S et une situation de communication moins spécialisée, représentée par des textes dédiés à des semi-experts ou à des non-experts, par exemple dans un contexte de vulgarisation scientifique. Néanmoins, si nous considérons que les degrés de spécialisation peuvent s'appréhender à travers un continuum entre une situation très spécialisée et une situation non spécialisée, alors il semble légitime de considérer que la multidimensionnalité peut être observée dans ce continuum et peut donc être invoquée pour caractériser des phénomènes dans un contexte non spécialisé également.

Deuxièmement, comme nous l'expliquons au Chapitre 1 (1.1.3.2), nous continuons à considérer les unités sur lesquelles nous travaillons dans le SC-P comme des termes. Aussi, il semble légitime de conserver ce point de vue terminologique sur notre objet d'étude, même dans un contexte non spécialisé, et par conséquent de continuer à considérer les concepts dans ce même contexte, et donc leur caractère multidimensionnel.

Dernièrement, en lien avec cette seconde raison, nous faisons le choix de nous aligner sur les propositions de Meyer et Mackintosh, qui expliquent que le premier type de changement sémantique relatif à la déterminologisation se caractérise par le fait que le concept sous-jacent reste perceptible lorsque le terme est employé par des non-spécialistes (Meyer et Mackintosh, 2000a : 202-203). Il semble alors qu'il soit possible, selon les autrices, de considérer les concepts même dans un contexte non spécialisé et, par extension, de mettre en œuvre des notions habituellement réservées à l'analyse de corpus spécialisés pour commenter nos observations de la déterminologisation.

Dans ce contexte, nous exploitons la multidimensionnalité pour approfondir notre réflexion sur les phénomènes observés dans nos données de deux manières : à travers les différents domaines dans lesquels un terme est utilisé et à travers les genres textuels et les degrés de spécialisation.

8.1.2.2. Multidimensionnalité à travers les domaines

Suivant les propositions de L'Homme (2020 : 88*sqq.*), la multidimensionnalité peut renvoyer à la possibilité d'appréhender un concept de la perspective de plusieurs domaines, lorsque ce concept a une certaine importance dans différents domaines. Il s'agit par ailleurs d'une des causes

¹⁶⁸ Péraldi mentionne à ce propos des causes liées à « [l']objectif du texte [...] et/ou [au] type d'utilisateur » ainsi qu'au « niveau d'expertise » (Péraldi, 2011 : 56).

de multidimensionnalité selon Péraldi, particulièrement lorsqu'il est question de recherche interdisciplinaire (Péraldi, 2011 : 56).

L'autrice précise en outre que la multidimensionnalité implique qu'un

« même objet peut être désigné dans un texte selon plusieurs perspectives, si diverses caractéristiques du concept représentant cet objet sont activées ou si différentes relations envers d'autres concepts sont établies. » (Péraldi, 2012 : §40)

Ainsi, cela signifierait que l'activation de certaines caractéristiques du concept dépendrait du domaine par le biais duquel le concept est considéré. Dans des textes spécialisés d'un domaine en particulier, il est peu probable que les caractéristiques associées à d'autres domaines soient observables. Cependant, dès lors que les textes relèvent d'une situation de communication non spécialisée (comme c'est le cas du SC-P), alors il est possible que les caractéristiques associées à plusieurs domaines coexistent, autrement dit que différentes dimensions soient actualisées. En effet, étant donné que les textes pris en compte dans le SC-P sont des articles de presse généraliste, qui ne relèvent pas strictement du domaine de la physique des particules, il est possible que des dimensions spécifiques de domaines différents apparaissent simultanément dans le SC-P, précisément en raison de la non-spécialisation des textes considérés.

Ainsi, comme nous le constatons dans les données, certaines occurrences d'un terme dans le SC-P permettent d'observer différentes dimensions, selon le domaine dans lequel ce terme est central en langue de spécialité. Du point de vue de la déterminologisation, cela signifie que les usages des termes en langue générale sont « teintés » des spécificités de chaque domaine, et non uniquement d'un seul domaine, qui serait pris comme « point de départ » du processus. Il semble ainsi que le sens des termes en langue générale ne se construit pas seulement à partir d'un domaine, mais de l'ensemble des domaines dans lesquels ces termes sont utilisés.

Les différences de distribution observées entre le SC-P et le SC-S mettent particulièrement en évidence l'existence de différentes dimensions, à travers les unités reconnues comme *key* et classées dans des catégories renvoyant notamment aux domaines de la médecine, du spatial ou encore du nucléaire. À des fins d'illustration, reprenons les termes *photon* et *proton*, discutés dans le Chapitre 6 (6.2.1). Pour ce qui est du terme *photon*, nous repérons des contextes du SC-P qui évoquent plusieurs autres domaines, comme la médecine (exemple 181), le spatial (exemple 182) et la cryptographie quantique (exemple 183)¹⁶⁹. Ces domaines sont identifiés à partir de la cooccurrence du terme avec *radiothérapie* (181), *télescope* et *astronome* (182), *cryptographique*, *cryptage*,

¹⁶⁹ Les domaines illustrés dans ces exemples ne sont pas exhaustifs.

réseau quantique notamment (183). L'exemple 184 illustre un contexte d'utilisation du terme *photon* qui relève de la physique des particules.

181. La radiothérapie classique utilise des **photons** [...] (SC-P, Les Échos, 2006)
182. Pour scruter le monde avec leurs télescopes, les astronomes se servaient jusqu'à présent de la lumière, c'est-à-dire, en termes scientifiques, des **photons** [...] (SC-P, Le Monde, 2008)
183. Le principal intérêt d'un réseau quantique est la mise en œuvre de protocoles cryptographiques incassables. Ces systèmes de cryptage reposent sur un principe fondamental de la mécanique quantique : il est impossible de détecter un **photon**, c'est-à-dire de lire l'information quantique qu'il transporte, sans le perturber - sans " changer son état quantique ", disent les physiciens. Toute écoute d'une ligne quantique peut ainsi être repérée. (SC-P, Le Monde, 2005)
184. Comment expliquer que certaines [particules élémentaires], comme les **photons**, ont une masse nulle, tandis que d'autres, comme les bosons Z, ont une masse élevée ? (SC-P, Le Temps, 2011)

Si nous acceptons que différentes dimensions sont observables dans les contextes d'un terme qui renvoient à des domaines différents, alors il est possible que ces dimensions participent à la construction du sens de ce terme dans un contexte non spécialisé et, par extension, au processus de déterminologisation. Ainsi, dans l'exemple 181, la multidimensionnalité est mise en évidence par l'utilisation des photons en radiothérapie, qui illustre une dimension liée au domaine médical. De la même manière, dans l'exemple 184, la relation de co-hyponymie entre *photon* et *boson Z* (l'hyperonyme étant *particule élémentaire*, repris anaphoriquement par le pronom *certain* dans cet exemple) dénote le point de vue de la physique des particules.

Nous suivons le même raisonnement pour le terme *proton*. Dans les exemples suivants, ce terme s'intègre en effet dans des contextes qui explicitent l'utilité des protons en médecine (soit seul, comme dans l'exemple 185, soit par le biais du terme complexe *faisceau de protons*, dans les exemples 186 et 187).

185. [...] la protonthérapie fait appel à des **protons**, autrement dit des noyaux d'hydrogène. (SC-P, Les Échos, 2011)
186. Le PSI, seul centre suisse à pratiquer l'hadronthérapie, traite actuellement une centaine de patients par an grâce à des **faisceaux de protons**. (SC-P, Le Temps, 2014)
187. Ce mode de traitement utilise un accélérateur de particules (synchrotron) pour irradier avec une grande précision les tumeurs grâce à des **faisceaux de protons** et d'ions carbone (photo). (SC-P, Les Échos, 2008)

L'utilisation des protons dans le domaine médical, en particulier dans des méthodes de traitement du cancer, peut être interprétée comme une dimension de ce concept qui est caractéristique de ce domaine. Dans cette perspective, il est peu probable que cette dimension soit

également observable dans des textes qui relèvent du domaine de la physique des particules. Cependant, au cours du processus de déterminologisation, toutes ces dimensions peuvent être transférées et il semblerait que la multidimensionnalité de *proton* participe à véhiculer le sens de *proton* à la fois comme une particule intervenant dans une méthode de traitement du cancer par l'irradiation des tumeurs et comme une particule constituant le noyau des atomes (cf. Chapitre 4, 4.1.2).

Ces réflexions mettent ainsi en lumière le fait que la multidimensionnalité peut se caractériser par l'utilisation d'un concept dans des domaines différents et qu'à l'intérieur de chaque domaine, différentes dimensions sont susceptibles d'être observées. Par ailleurs, il semble que toutes les dimensions peuvent potentiellement être actualisées dans la langue générale, du moins un plus grand nombre de dimensions que celles observées à l'intérieur de chaque domaine pris isolément. Nous revenons plus en détail en 8.2.1.2 sur le rôle de la multidimensionnalité dans le processus de passage des termes dans la langue générale.

8.1.2.3. Multidimensionnalité à travers les genres et les degrés de spécialisation

Le second élément qui peut être abordé en rapport avec la multidimensionnalité concerne les différents aspects des concepts repérés dans le SC-P par rapport au SC-S, détaillés au Chapitre 7 (7.2.2 et 7.3.3). Dans ce cas, il ne s'agirait pas de dimensions caractéristiques des domaines dans lesquels un concept peut être utilisé, mais de dimensions qui varient *a priori* dans le même domaine en fonction de la situation de communication, notamment en fonction des genres textuels et des degrés de spécialisation. Si nous reprenons les « causes potentielles de multidimensionnalité » énumérées par Péraldi (2011 : 56), plusieurs d'entre elles peuvent effectivement être mises en relation avec les genres et les degrés de spécialisation, notamment le « niveau d'expertise » ou « l'objectif du texte ». De la même manière, la multidimensionnalité peut dépendre de « divergences majeures selon les théories scientifiques/écoles de pensée » (*Ibid.*).

Dans cette perspective, il semble que certaines différences de distribution mises en évidence dans le SC-P ou dans le SC-S peuvent être associées à des dimensions des concepts impliqués. Certaines dimensions seraient donc caractéristiques du SC-P ou du SC-S, comme nous l'avons vu avec les termes *quark*, *boson de Higgs*, *particule*, *collisionneur* ou *détecteur* par exemple. De même, toutes les dimensions ne sont pas nécessairement représentées dans les textes de spécialité, du moins pas ceux qui constituent le SC-S. L'intérêt d'aborder les différences de fonctionnement des termes dans ces deux sous-corpus à travers le prisme de la multidimensionnalité réside selon nous dans la possibilité de décrire plus finement les phénomènes liés à la déterminologisation. En effet, les différences de distribution ne doivent pas nécessairement être mises en relation avec des différences

de sens, mais bien avec des différences de points de vue ou de conceptualisation, ou encore des nuances selon les besoins de la situation de communication. Par exemple, les aspects liés au financement de nouvelles infrastructures de recherche et/ou aux moyens de justification des sommes investies dans ce but représentent une dimension particulière qui apparaît largement plus régulièrement dans le SC-P que dans le SC-S, probablement du fait qu'il s'agit d'un aspect de la recherche qui est généralement traité par les médias (nous y revenons en 8.3.2).

Ainsi, la multidimensionnalité permet de rendre compte de phénomènes qui se produisent au cours du processus de déterminologisation qui sont plus subtils que les changements sémantiques habituellement décrits dans la littérature (cf. Chapitre 2, 2.2.2.1). C'est parce que les concepts sont multidimensionnels et que leurs dimensions peuvent s'actualiser dans des situations de communication très différentes (selon le domaine, le genre, le degré de spécialisation notamment) qu'une telle diversité dans les fonctionnements des termes correspondants peut être observée dans le SC-P. Parallèlement, si la multidimensionnalité est une caractéristique des concepts, c'est bien à travers des aspects linguistiques qu'elle peut être appréhendée, par l'analyse fine de la distribution des termes.

Avec cette dernière remarque, nous souhaitons insister sur le fait que la multidimensionnalité permet de rendre compte de la complexité des phénomènes en jeu dans la déterminologisation, complexité qui se caractérise par le fait que les « dimensions sont susceptibles de s'entrecroiser, de s'additionner, mais également de se contredire, générant des réseaux notionnels complexes et parfois extrêmement difficiles à représenter » (Péraldi, 2011 : 70).

8.1.3. Affirmer la place de la déterminologisation dans la néologie

Dans les deux sections précédentes, nous montrons comment nos observations permettent de caractériser finement les changements sémantiques qui se produisent au cours du processus de déterminologisation. Si les usages métaphoriques renvoient à des changements tout à fait attendus, les modifications plus subtiles que nous mettons en lien avec la multidimensionnalité sont plus rarement décrites dans la littérature. Les descriptions des fonctionnements des termes dans le SC-P rendent alors possible de mieux saisir la richesse des changements en jeu. L'ensemble des phénomènes décrits permet en outre d'interroger le rôle de la déterminologisation dans les mécanismes sémantiques de création néologique et, plus largement, d'approfondir la réflexion sur les liens entre déterminologisation et néologie.

En effet, nous avons vu au Chapitre 1 (1.4) que l'apparition de termes dans la langue générale est un phénomène parfois abordé du point de vue de la néologie, de même que les

changements sémantiques par métaphore sont considérés comme des mécanismes de néologie sémantique. Parallèlement, nous avons montré que le fonctionnement des termes dans le SC-P diffère toujours dans une certaine mesure de leur fonctionnement dans le SC-S. Aussi, si nous tenons compte d'une définition large de la néologie sémantique, suivant les propositions de Reutenauer (2012b), alors il semble que nos observations démontrent bien que le processus de déterminologisation contribue de manière générale à la néologie. Dans cette perspective, nous pensons que les phénomènes mis au jour dans nos données confirment la pertinence de la déterminologisation comme facteur participant aux mécanismes de création néologique.

Nous illustrons notre propos à l'aide des termes *matière noire* et *accélérateur de particules*. À partir de nos observations dans le SC-P, un parallèle peut être dressé entre les exemples de *matière noire* et d'*accélérateur de particules* cités *supra* et les exemples de *to download* et *to recycle* discutés par Meyer et Mackintosh (2000b), qui attestent de l'émergence de nouveaux sens par métaphore. Cependant, les occurrences des usages métaphoriques des deux termes en question sont peu nombreuses dans nos données, en comparaison avec l'ensemble de leurs occurrences dans le SC-P, et ne semblent pas renvoyer à un sens tout à fait stabilisé en langue générale. Au contraire, à l'instar de Meyer et Mackintosh à propos du terme *anorexic* (*Ibid.* : 120), nos observations suggèrent plutôt que les termes possèdent le potentiel d'être employés métaphoriquement et que les différents usages métaphoriques attestés dans nos données comportent certaines nuances selon les besoins de la situation de communication, selon l'effet recherché, selon la proximité voulue avec le domaine d'origine, et peut-être même selon les connaissances des journalistes du concept d'origine.

Par exemple, le terme *accélérateur de particules* peut être utilisé pour qualifier une personne ou un groupe de personnes (exemples 188 et 189) ou un objet, un concept (exemples 190 et 191) ou encore un pays (exemple 192). Nous pouvons par ailleurs distinguer plusieurs nuances dans ces exemples. Dans l'exemple 188, le terme semble renvoyer à une personne qui contribue au succès d'autres personnes ou à lancer leur carrière. Dans l'exemple 189, l'usage d'*accélérateur de particules* met en avant la rapidité d'un rugbyman. Dans l'exemple suivant, le terme réfère à la capacité d'un titre de musique de faire connaître un genre musical ou l'auteur du titre. Un emploi similaire s'observe également dans l'exemple 191, où *accélérateur de particules* semble synonyme de *tremplin*. Enfin, dans l'exemple 192, le terme démontre un usage synonyme à celui de l'unité *révélateur*. Dans tous ces exemples, il semble que l'emploi de ce terme dénote une volonté d'insister sur la capacité d'accélération de la personne ou de l'objet désigné, qu'il s'agisse d'accélérer une carrière, le jeu ou l'issue d'un match, la diffusion d'un genre musical ou encore certaines actions sur le plan politique.

188. Collectionneur compulsif, amateur de cirque et de rock, endetté chronique, cet **accélérateur de particules** aime "renifler" l'air du temps et la truffe des gens à qui il décide de donner leur chance. (SC-P, Les Échos, 2012)
189. Barrett, l'**accélérateur de particules** All Blacks ; Successeur désigné de Dan Carter, l'ouvreur néo-zélandais impressionne par sa vitesse d'exécution. (SC-P, Le Figaro, 2016)
190. Quand le trio de jazz s'enclenche sur "Countdown", **accélérateur de particules** bop des débuts de John Coltrane, il fait mine de trouver le tempo avec les phalanges. (SC-P, Le Temps, 2006)
191. " Les Arts déco sont un **accélérateur de particules** ", se félicite cet ancien élève des Beaux-Arts d'Angoulême, qui est resté à Strasbourg à la fin de ses études, en 2015. (SC-P, Le Monde, 2016)
192. ELLE AVAIT un peu disparu des écrans. [...] Et puis, la revoici Martine. En guerre. Colère. La Chine a des effets étonnants sur les prétendants à l'Élysée. Elle agit comme un révélateur. Un **accélérateur de particules**. Elle avait inspiré "la bravitude" à Ségolène Royal. Elle requinque la première secrétaire du PS. (SC-P, Le Figaro, 2010)

En outre, si nous considérons ces exemples du point de vue de la néologie sémantique, alors il semble que cette diversité et ce manque apparent de stabilité dans les usages du terme *accélérateur de particules* peuvent également renvoyer à un processus de stabilisation dans l'usage d'un néologisme sémantique. Ce processus correspond à un « état transitoire », ou à une « évolution en cours » en discours, selon Reutenauer (2012b : 61), qui précède la reconnaissance d'un nouveau sens stable en langue. En effet, l'auteur démontre que l'émergence, dans la presse, d'un nouveau sens plus général pour un terme passe entre autres par l'attestation de ce terme dans des contextes qui renvoient à plusieurs domaines. Il semble bien que ce soit ce que nous observons ici avec le terme *accélérateur de particules*, qui s'insère, dans ces exemples, dans des contextes renvoyant à la télévision, au sport, à la musique, aux arts et à la politique.

Ainsi, puisque les usages métaphoriques d'un terme peuvent être interprétés comme des conséquences de la déterminologisation de ce terme et puisque ces usages métaphoriques peuvent conduire à l'émergence et à la stabilisation d'un nouveau sens pour ce terme, alors la déterminologisation doit effectivement être considérée comme contribuant aux mécanismes de création néologique.

8.2. Comprendre la diversité des intermédiaires

Dans les Chapitres 6 et 7, nous avons montré que la plupart des différences constatées dans la distribution des termes dans le SC-P par rapport au SC-S ne renvoient pas à des fonctionnements tout à fait caractéristiques du SC-P, car ceux-ci sont en réalité également attestés dans un ou plusieurs sous-corpus intermédiaires. Ce constat nous permet de développer notre réflexion à propos du rôle des textes se situant à l'intermédiaire entre langue de spécialité et langue générale

dans les changements sémantiques qui se produisent au cours du processus de déterminologisation tels que nous pouvons les observer dans nos données. Nous pensons ainsi que l’attestation de ces fonctionnements dans les sous-corpus intermédiaires et dans le SC-P permet de nuancer le rôle généralement attribué à la presse et aux médias généralistes, non seulement dans la diffusion des termes, mais également dans les changements sémantiques observés.

Parallèlement, l’observation de fonctionnements dans le SC-P uniquement nous amène à considérer la possibilité d’une certaine progression dans la diversification des fonctionnements des termes au cours du processus, dans le continuum langue de spécialité-langue générale. C’est ce que nous proposons d’aborder dans la section suivante, à partir des deux aspects discutés *supra*, la métaphore et la multidimensionnalité.

En outre, nous avons relevé en particulier au Chapitre 7 d’importantes similarités dans la distribution des termes dans le SC-P et le SC-V. Cette caractéristique nous amène à rediscuter la place de la vulgarisation scientifique dans le processus (en 8.2.2) et le statut de la presse (en 8.2.3). Enfin, nous montrons dans la section 8.2.4 que nos observations permettent de repérer d’autres types d’intermédiaires potentiels, dont il faut tenir compte dans l’étude de la déterminologisation.

8.2.1. Interroger la continuité du processus par la prise en compte des sous-corpus intermédiaires

8.2.1.1. Potentiel métaphorique

Nous expliquons *supra* comment le potentiel métaphorique d’un terme peut se définir à partir des différents usages de ce terme dans le SC-P, notamment lorsqu’il cooccure de manière récurrente avec certaines des unités classées dans la catégorie « sensationnalisme ». Parallèlement, l’analyse systématique des contextes d’apparition de ces unités dans le corpus montre qu’elles ne sont pas caractéristiques du SC-P uniquement, comme nous en avons fait l’hypothèse, car elles sont également attestées dans les sous-corpus intermédiaires (cf. Chapitre 6, 6.3). Ces éléments permettent ainsi de rendre compte des dynamiques de changement sémantique en jeu dans le processus de déterminologisation et particulièrement dans la création du potentiel métaphorique des termes. Afin de discuter cet aspect, nous reprenons les exemples de *matière noire* et d’*accélérateur de particules*.

Dans les exemples tirés des sous-corpus intermédiaires, le terme *matière noire* est modifié par l’adjectif *mystérieux* et apparaît dans des syntagmes du type « le mystère de la matière noire », de manière similaire aux contextes observés dans le SC-P. Il semble alors que des connotations liées aux aspects mystérieux entourant la thématique de la recherche de la matière noire soient

observables également dans les contextes du terme *matière noire* dans les sous-corpus intermédiaires, de la même manière que dans le SC-P. Dans cette perspective, si les usages métaphoriques du terme *matière noire* s'appuient effectivement sur ces connotations, du moins en partie, comme nous l'expliquons en 8.1.1.1, alors il semble que les fonctionnements créant la possibilité d'un usage métaphorique dans le SC-P soient aussi observables dans les sous-corpus intermédiaires.

Les mêmes conclusions peuvent être tirées à partir des usages métaphoriques du terme *accélérateur de particules*. En effet, nous avons vu que ces usages dans le SC-P se fondent en partie sur les cooccurrents du terme qui mettent en évidence les aspects extraordinaires et particulièrement puissants ou performants des accélérateurs de particules, notamment du LHC au CERN. Or ce terme est attesté dans des contextes tout à fait similaires dans les sous-corpus intermédiaires (cf. 6.3.2.4). À partir de ces constats, il apparaît que les usages repérés dans les sous-corpus intermédiaires contribuent effectivement aux phénomènes décrits, notamment par l'apparition de cooccurrents différents des cooccurrents attestés dans le SC-S et par le transfert d'une partie de ces cooccurrents dans la presse.

Par ailleurs, étant donné que les usages métaphoriques de ces deux termes sont attestés dans le SC-P uniquement, il semble qu'à chaque étape du processus de déterminologisation tel que nous pouvons l'étudier dans nos données (du SC-S aux sous-corpus intermédiaires et des sous-corpus intermédiaires au SC-P), un changement peut être constaté. D'abord, dans les sous-corpus intermédiaires en comparaison avec le SC-S, nous observons une modification de la distribution de *matière noire* et d'*accélérateur de particules* ; en effet, ces termes cooccurrent désormais avec des unités évoquant plusieurs aspects mystérieux, hors normes ou extraordinaires. Puis, dans le SC-P en comparaison avec les sous-corpus intermédiaires, nous observons l'apparition de nouveaux usages, où *matière noire* est utilisé métaphoriquement pour désigner quelque chose de mystérieux, d'inconnu ou d'invisible, et où *accélérateur de particules* est utilisé pour désigner quelque chose ou quelqu'un de particulièrement rapide ou performant, ou qui contribue à la rapidité d'autre chose.

À la lueur de ces éléments, il semble donc bien qu'une certaine progression dans les changements sémantiques relevant de la déterminologisation puisse être constatée dans le corpus. Cette idée de progression est en particulier mise en évidence par l'attestation dans les sous-corpus intermédiaires des fonctionnements des termes qui, d'après nos observations dans la section 8.1.1.1, contribuent à créer le potentiel métaphorique des termes dans le SC-P. Autrement dit, la création de ce potentiel n'est pas le fait de la presse uniquement.

8.2.1.2. Multidimensionnalité

Cette idée de progression semble se refléter dans nos observations relatives à la multidimensionnalité, dans les sous-corpus intermédiaires, car la majorité des dimensions observées dans le SC-P sont également présentes dans les sous-corpus intermédiaires. En effet, dans les Chapitres 6 et 7, nous avons remarqué des contextes mettant en évidence l'interdisciplinarité de la recherche en physique des particules, en particulier dans le SC-R et dans le SC-V.

Dans cette perspective, étant donné que chaque domaine dans lequel un terme est utilisé est susceptible de porter ses propres points de vue sur ce terme et puisque nous associons ces points de vue à différentes dimensions, alors il semble que certaines dimensions sont également observables dans les sous-corpus intermédiaires. C'est le cas notamment des dimensions relatives à l'utilisation de photons ou de positons dans le domaine médical ou encore des dimensions relatives à l'utilisation de neutrons dans les réacteurs nucléaires, qui apparaissent bien dans au moins un sous-corpus intermédiaire, mais que nous ne constatons pas dans le SC-S. Le même constat peut être effectué pour les dimensions repérées dans le SC-P à l'aide des contextes syntaxiques, que nous observons dans au moins un sous-corpus intermédiaire (par exemple la dimension permettant de considérer les accélérateurs de particules et les détecteurs comme des monuments ou des bâtiments que l'on peut visiter et inaugurer ; cf. Chapitre 7, 7.3.3).

D'autres dimensions, en revanche, ne sont pas observables dans les sous-corpus intermédiaires, notamment celles associées aux liens que la physique des particules entretient avec certains domaines. Par exemple, l'utilisation d'accélérateurs de particules dans le but d'analyser des œuvres d'art est mise en évidence dans plusieurs contextes dans le SC-P uniquement (cf. Chapitre 6, 6.2.1) et n'apparaît jamais dans les sous-corpus intermédiaires.

Ces différents éléments pointent vers la même idée de progression entre le SC-S et les sous-corpus intermédiaires et entre les sous-corpus intermédiaires et le SC-P que celle développée dans la section 8.2.1.1. Ce constat nous permet d'aller plus loin dans notre réflexion à propos du processus de déterminologisation. En effet, à partir de nos observations, il semble bien que, si les concepts sont multidimensionnels, toutes leurs dimensions ne passent pas de manière uniforme au cours du processus. Au contraire, certaines dimensions sont présentes uniquement dans le SC-S et dans au moins un sous-corpus intermédiaire, ce qui renvoie à des dimensions qui ne passeraient pas dans le SC-P. D'autres dimensions, présentes dans le SC-P et dans au moins un sous-corpus

intermédiaire, ne sont pas caractéristiques du SC-S sans pour autant « apparaître » dans le SC-P¹⁷⁰. Enfin, certaines dimensions sont présentes uniquement dans le SC-P et renvoient donc à des aspects caractéristiques de ce sous-corpus. Il semble ainsi que le nombre de dimensions a tendance à augmenter lorsque le degré de spécialisation des textes diminue et lorsque les locuteurs auxquels sont destinés ces textes augmentent en nombre et se diversifient.

Par ailleurs, puisque ces dimensions peuvent également dépendre des domaines par lesquels les concepts sont appréhendés, alors il paraît évident qu'un plus grand nombre de dimensions est observé lorsque le degré de spécialisation est plus faible. En effet, il semble logique que le SC-S ne permette d'observer que les dimensions caractéristiques de la physique des particules et du point de vue de la recherche, étant donné les genres textuels pris en considération dans ce sous-corpus. De la même manière, le fait que les textes inclus dans les sous-corpus intermédiaires sont destinés à un lectorat plus diversifié et moins spécialisé que les textes du SC-S pourrait expliquer la diversification des dimensions. Par exemple, dans la mesure où les textes du SC-V cherchent à transmettre certaines notions de la physique des particules ou certaines découvertes ou événements de ce domaine, alors il semble pertinent de considérer que ces explications passent par la clarification des liens que la physique des particules entretient avec d'autres domaines ou par l'explicitation du développement d'applications qui sont utiles pour la société. Le même raisonnement peut être suivi pour le SC-P, dont les textes sont non spécialisés et dont le lectorat est encore plus large.

Il semble donc bien que, tant dans le cas de la multidimensionnalité que dans le cas des usages métaphoriques des termes, une certaine continuité dans le changement soit observable à travers les sous-corpus. Cependant, bien que, dans nos propos, nous traitions les sous-corpus intermédiaires comme un seul ensemble, il n'est pas sûr que les genres dont relèvent les textes inclus dans chacun de ces sous-corpus aient la même importance dans les changements observés. Au contraire, certaines nuances existent bien selon le sous-corpus, et donc selon les caractéristiques des textes les composant.

Plus précisément, le sous-corpus *Vulgarisation* se démarque des autres, de par les similarités importantes qu'il partage avec le sous-corpus *Presse*. Ces similarités sont notamment mises en évidence dans le Chapitre 6 (6.3.1), avec la répartition des unités de la catégorie « sensationnalisme » et, dans une moindre mesure, leur fréquence, ainsi que dans le Chapitre 7 (7.1), où le score de

¹⁷⁰ Notons que, pour clarifier notre propos, nous ne tenons pas compte des différences observées entre les sous-corpus intermédiaires, mais il serait erroné de déduire que ces trois sous-corpus permettent toujours d'observer les mêmes dimensions d'un concept. Au contraire, chaque sous-corpus possède ses spécificités (nous y revenons notamment dans la section 8.2.2).

similarité de Jaccard est plus élevée pour chaque terme entre le SC-P et le SC-V qu'entre le SC-P et les autres sous-corpus. Aussi, étant donné que les rapports entre vulgarisation scientifique et déterminologisation apparaissent à plusieurs reprises dans notre travail, il nous semble nécessaire de revenir sur ces liens, en fonction des travaux pris en compte dans les Chapitres 1 et 2 et de nos observations dans le corpus.

8.2.2. Clarifier la place de la vulgarisation

Au vu de ce qui précède, et bien que nous ayons pu confirmer la pertinence de la prise en compte de textes intermédiaires pour interroger le processus de déterminologisation, il semble que le SC-V ait un statut un peu différent des SC-C et SC-R. Nos observations dans le corpus ainsi que les constats effectués à partir des travaux existants sur la vulgarisation permettent de mieux saisir les rapports que la vulgarisation entretient avec la déterminologisation.

Par exemple, rappelons que le passage des termes dans la langue générale est parfois associé à la vulgarisation et que ces deux phénomènes sont parfois regroupés dans la question plus large de sociodiffusion des sciences (cf. Chapitre 2, 2.1). Cependant, nous avons également vu que les discours de transmission des connaissances, et donc les textes de vulgarisation, sont en réalité des vecteurs par lesquels les termes atteignent les non-spécialistes, et peuvent par conséquent être considérés comme des intermédiaires dans le processus de déterminologisation. Plus précisément, nous avons vu au Chapitre 1 que la vulgarisation peut être définie comme la diffusion de connaissances spécialisées auprès du grand public, qui s'effectue par le biais de diverses stratégies (de reformulation, notamment) assurant leur appropriation par le grand public. La vulgarisation est donc considérée comme un phénomène qui contribue à la circulation des termes en dehors de la sphère spécialisée, dans lequel s'inscrit également la déterminologisation. Cependant, les textes de vulgarisation se caractérisent par la présence d'importantes marques de didacticité, qui dénotent la volonté sous-jacente de diffuser des connaissances, alors que la déterminologisation, entendue comme le passage de termes dans la langue générale et les changements sémantiques inhérents à ce passage, est un processus qui s'effectue de manière progressive, voire inconsciente. Par exemple, Condamines et Picton précisent

« qu'il s'agit le plus souvent d'un phénomène non conscient à la fois des locuteurs et des interlocuteurs. Il se différencie alors de la vulgarisation par le fait qu'il ne relève pas d'un processus délibéré de mise à portée de connaissances scientifiques auprès du grand public » (Condamines et Picton, 2014a : 168)

C'est dans ce sens que l'on peut considérer, à l'instar notamment de Halskov (2005 : 54), que la déterminologisation relèverait plutôt de l'évolution de la langue de manière générale. Ainsi,

une différence cruciale entre ces deux phénomènes est reflétée par le caractère conscient de la vulgarisation et inconscient de la déterminologisation.

Parallèlement, puisqu'il semble que le SC-P partage plus de similarités avec le SC-V qu'avec les autres sous-corpus intermédiaires, du moins pour ce qui est des fonctionnements des termes que nous avons observés, alors il semble que la vulgarisation contribue dans une plus large mesure au processus de déterminologisation que les textes relevant des genres pris en compte dans les deux autres sous-corpus intermédiaires. De ce point de vue, nous pensons que la vulgarisation doit être considérée comme un intermédiaire privilégié dans ce processus¹⁷¹.

Ces remarques permettent également de mettre en évidence le statut particulier de la presse dans la déterminologisation, qui ne doit pas être considérée comme le « point d'arrivée » de ce processus (qui serait plutôt la langue générale), mais comme un genre relevant d'une situation de communication non spécialisée qui permet de repérer et de caractériser certaines manifestations de la déterminologisation et qui participe au passage des termes dans la langue générale¹⁷². Il semble alors que les similarités observées entre le SC-P et le SC-V indiquent plutôt que les articles de presse et les textes de vulgarisation ont un objectif en commun, celui de susciter la curiosité des lecteurs et de capter leur attention. Nous l'expliquons dans le Chapitre 6, pour la presse, cet objectif peut inciter les journalistes à rédiger avec un style plus imagé, à insister sur les aspects extraordinaires ou spectaculaires de certaines informations ou encore à « jouer » sur la fascination que certains thèmes peuvent exercer sur le grand public. Mais susciter l'intérêt et maintenir l'attention des lecteurs est un objectif également poursuivi par la vulgarisation scientifique. Il semble donc tout à fait probable que certaines stratégies comparables soient employées dans ce but, dans les textes rassemblés dans le SC-P et dans le SC-V.

De la même manière, nous identifions également des marques de didacticité dans le SC-P. Celles-ci s'illustrent par ailleurs dans des contextes tout à fait similaires aux contextes du SC-V, comme le montrent les exemples 193 à 196 ci-dessous, tirés de ces deux sous-corpus. Dans les deux premiers exemples, le terme *antimatière* est reformulé, ou expliqué (passage en italique), et l'explication est signalée par un marqueur (souligné dans les exemples). Il s'agit ici d'exemples du paradigme définitionnel d'*antimatière*. Dans les deux autres exemples, un terme et une variante dénominative sont juxtaposés. Dans ce cas, il s'agit d'exemples du paradigme désignationnel d'*anti-électron* (cf. Chapitre 1, 1.2.1).

¹⁷¹ Pour autant que la vulgarisation soit « réussie », c'est-à-dire que les connaissances soient effectivement transmises, voire que les locuteurs se les soient appropriées.

¹⁷² Rappelons que notre choix de tenir compte d'articles de presse généraliste repose précisément sur ces caractéristiques (cf. Chapitre 4, 4.2.2.1.1).

193. L'annonce était très attendue puisque cette **antimatière** (*des particules "miroirs", qui ont des charges opposées aux particules qui composent la matière classique*) ne peut avoir été produite que par "une physique exotique", d'une nature nouvelle jamais observée sur Terre. (SC-P, Le Figaro, 2013)
194. L'**antimatière est faite de** particules qui ont des propriétés égales mais opposées à celles qui composent la matière de tous les jours. (SC-V, LHC-France, 2016)
195. Pourtant, il est possible de produire des quantités significatives d'antihydrogène au CERN en mélangeant des **antiélectrons** (*positons*) et des antiprotons de basse énergie. (SC-P, Le Temps, 2014)
196. Schématiquement, un neutron peut s'associer à un **positron** (*anti-électron*) pour former un proton et un anti-neutrino. (SC-V, Élémentaire, 2005)

La presse garde toutefois sa spécificité par rapport aux textes de vulgarisation, dans la mesure où il est peu probable que les journalistes généralistes possèdent réellement toutes les connaissances nécessaires pour vulgariser (Peynaud, 2014). Au contraire, Peynaud explique que les journalistes cherchent à atteindre un équilibre « entre la tentation de trop simplifier et la volonté de produire un discours accessible » (*Ibid.* : §9) et limitent donc la part de reformulation ou d'explication. Cette stratégie vise ainsi à tenir compte des profils variés et difficiles à cerner précisément des locuteurs auxquels s'adressent les articles. C'est en ce sens que Peynaud qualifie le lectorat de la presse généraliste d'indéfinissable (*Ibid.* : §8). La presse généraliste se caractérise donc par une certaine part de reformulation, jugée nécessaire à la compréhensibilité des articles par les lecteurs, mais qui ne vise pas à « leur faire acquérir des connaissances sur les sujets traités au-delà de l'espace de l'article » (*Ibid.*)¹⁷³.

Ces observations permettent ainsi de mettre en lumière le rôle particulier de la vulgarisation dans la déterminologisation. De fait, la vulgarisation, en tant que discours de transmission des connaissances, peut s'illustrer dans une diversité de situations de communication, par exemple dans la presse, dans une certaine mesure et en tenant compte de ses particularités, comme nous venons de le voir. Toutefois, si la vulgarisation peut être considérée comme un vecteur privilégié du passage des termes dans la langue générale, il semble que la presse ait également une importance particulière dans ce processus.

8.2.3. Définir le double statut de la presse

Nous l'avons dit, la presse constitue une alternative opérationnelle à la langue générale pour une approche en corpus, en tant que situation de communication non spécialisée, et elle joue, dans le même temps, un rôle reconnu dans la diffusion des termes auprès de non-spécialistes.

¹⁷³ Ces observations rejoignent par ailleurs celles des chercheurs travaillant sur les médias de manière générale, sous l'angle des discours de transmission des connaissances, et qui constatent également certaines marques de didacticité dans les médias (Beacco *et al.*, 2002 ; Moirand, 2007, 2014).

Parallèlement, nous avons vu *supra* que le SC-P est le sous-corpus qui regroupe la plus grande diversité de fonctionnements des termes, dont certains s'apparentent à ceux que l'on observe dans le SC-V, c'est-à-dire qui s'inscrivent dans des contextes qui démontrent certaines marques de didacticité. D'autres en revanche relèvent de changements sémantiques plus importants, notamment des emplois métaphoriques. Cette diversité peut alors être mise en relation avec le double statut de la presse, à la fois comme média contribuant à la circulation des termes et comme lieu privilégié de l'observation des manifestations de la déterminologisation, en tant que situation de communication non spécialisée. Autrement dit, la presse permet d'observer la déterminologisation à la fois comme processus et comme résultat. Ce double statut se répercute ainsi inévitablement sur l'analyse, qui doit tenir compte du fait que les observations effectuées dans le SC-P ne permettent pas véritablement de saisir les fonctionnements des termes dans la langue générale, mais permettent uniquement de s'en approcher et de formuler des hypothèses à leur propos (cf. Chapitre 4, 4.2.2.1).

Cette remarque renvoie également à la difficulté d'approcher la notion de langue générale en corpus, qui reste difficile à définir de manière opérationnelle. Pourtant, c'est bien dans la langue générale que les manifestations de la déterminologisation devraient être observées, afin de compléter la caractérisation des phénomènes liés. Une autre alternative consisterait par exemple à tenir compte d'autres données qui relèveraient également d'un contexte non spécialisé.

Enfin, dans le même ordre d'idée, nos réflexions nous mènent à considérer que l'analyse des intermédiaires représentés dans les SC-C, SC-R et SC-V pourrait être complétée par d'autres types d'intermédiaires.

8.2.4. Identifier d'autres types d'intermédiaires

Si notre hypothèse du rôle de textes intermédiaires dans le processus de déterminologisation semble être confirmée par l'exploration des SC-C, SC-R et SC-V, d'autres intermédiaires sont toutefois susceptibles d'intervenir. Par exemple, si nous tenons compte de la complexité des modèles de transmission des connaissances spécialisées vers le grand public (cf. Chapitre 1, 1.2.2), dans lesquels les médias jouent un rôle crucial (Beacco *et al.*, 2002 ; Moirand, 2014), alors il est possible que la multiplicité des supports médiatiques existants entre également en jeu dans la déterminologisation. Nous pensons plus particulièrement aux réseaux sociaux, qui représentent désormais un enjeu important de la communication vers le grand public. Par ailleurs, les réseaux sociaux pourraient intervenir à plusieurs niveaux : d'abord entre les spécialistes et le grand public, par le biais des organismes de recherche scientifique. Ensuite, ils pourraient jouer un

rôle d'intermédiaire similaire à celui attribué à la presse, par le biais des médias d'information qui communiquent également *via* les réseaux sociaux par exemple, et enfin un rôle de représentation des échanges entre non-spécialistes, par le biais de comptes d'utilisateurs *lambda*.

Cependant, la prise en compte des réseaux sociaux ajoute une certaine complexité à l'analyse dans des données réelles. Nous avons vu au Chapitre 4 la difficulté à sélectionner des données textuelles pertinentes dans ce but, tout en limitant les biais associés à notre propre subjectivité et à un manque de reproductibilité de la démarche, mais le travail sur des données issues des réseaux sociaux implique d'autres contraintes, liées à l'instabilité des réseaux, à la limite de la taille des textes selon les réseaux, aux possibilités de sélection des textes selon certains paramètres, etc.

De la même manière, bien que nous nous soyons focalisée sur des données écrites pour des raisons pratiques flagrantes pour une approche en corpus, il semble évident que la déterminologisation ne se produise pas uniquement par le biais de la communication écrite, mais aussi par le biais de la communication orale, par exemple à travers différents supports de vulgarisation, comme des émissions de télévision ou des podcasts. Ce constat nous amène à interroger le rôle d'autres types d'intermédiaires dans le processus de déterminologisation. Nos données permettent en particulier de faire l'hypothèse d'au moins deux autres types d'intermédiaires possibles : les œuvres de fiction et les textes relevant de différents domaines dans lesquels un même terme est utilisé.

Dans le premier cas, nous avons repéré différents contextes dans le SC-P qui évoquent certaines œuvres de fiction, notamment des romans, des séries télévisées ou des films. Dans ces œuvres, la physique des particules semble faire partie du scénario. Ainsi, l'exemple 197 ci-dessous traite d'un roman de Dan Brown (*Anges et Démons*), l'exemple 198 de son adaptation cinématographique, dont une partie a été filmée au CERN d'après le contexte reproduit ici, et l'exemple 199 d'une série télévisée (*The Flash*).

197. Le méchant du roman veut voler de l'**antimatière** pour fabriquer une bombe. (SC-P, Le Temps, 2013)
198. A découvrir les premières scènes, précisément tournées au **CERN** avec la bénédiction de sa direction, on ne donne pourtant pas cher du film. (SC-P, Le Temps, 2009)
199. [...] tout en étant le héros de la série dérivée *The Flash*, diffusée en deuxième partie de soirée. Dans cette première saison, l'explosion d'un **accélérateur de particules** transforme le jeune expert médico-légal en surhomme, capable de se déplacer à toute vitesse. (SC-P, Le Figaro, 2015)

La présence de la physique des particules dans des œuvres de fiction peut alors contribuer à diffuser des termes auprès du grand public. Ces œuvres de fiction constituent, en ce sens, un autre type d'intermédiaires intervenant potentiellement dans le processus de déterminologisation, du moins pour ce domaine. C'est également ce que suggère l'exemple suivant, en particulier par le biais du CERN et de son importance pour ce domaine (voir notamment le passage indiqué en gras).

200. Science, religion, catastrophe ? Evidemment le LHC ne pouvait pas échapper au spécialiste de ce mélange, Dan Brown. L'auteur américain a décrit un CERN fantaisiste dans la première partie d'*Anges et Démons* (J.-C. Lattès, 2005), publié aux Etats-Unis avant le *Da Vinci Code* (J.-C. Lattès, 2004), mais adapté au cinéma seulement en 2009. Cette fois, le service communication [du CERN] a anticipé le coup et préparé un site complet pour redresser les torsions imposées à la réalité par l'écrivain. Non, les physiciens ne portent pas de blouses blanches, et ne passent pas leur temps à circuler dans les couloirs avec des dossiers sous le bras. Non, au rythme où il fabrique des quantités infinitésimales d'antimatière, cet envers de notre matière qui s'annihile avec elle quand elle la rencontre, le CERN ne pourrait pas fabriquer une bombe avec ces particules étranges avant plusieurs millions d'années, si tant est qu'il parvienne un jour à les stocker. Les physiciens ne se sont pas formalisés de toutes ces libertés prises avec la science. **Anges et démons a contribué à faire connaître le CERN dans le monde entier**, autrement que par la peur du trou noir. Et il a capté un peu du charme de ce pays des physiciens, ni secret ni dangereux, mais voué à demeurer, tant qu'y tourneront ces expériences extrêmes, enveloppé de mystère. (SC-P, *Le Monde*, 2010)

Dans le second cas, nous avons montré aux Chapitres 6 et 7 que les contextes d'apparition des termes dans les sous-corpus intermédiaires permettent d'observer l'interdisciplinarité qui caractérise la recherche en physique des particules. Ce constat renvoie au fait que les mêmes termes peuvent être utilisés dans différents domaines et peuvent donc, comme le souligne L'Homme (2004 : 43-44), occuper une place centrale dans ces différents domaines. Dans cette perspective, il est possible que ces termes intègrent la langue générale non seulement par le biais de textes relevant de la physique des particules, mais également par le biais de textes relevant des autres domaines dans lesquels ils sont utilisés, ces textes constituant un autre type d'intermédiaires. Le processus de déterminologisation ne se produirait donc pas nécessairement à travers des textes relevant de la langue de spécialité d'un seul domaine. Au contraire, des textes relevant des langues de spécialité des différents domaines dans lesquels un terme est employé pourraient intervenir.

Dans la section 8.1.2.2, nous montrons plusieurs exemples où les termes *photon* et *proton* s'insèrent dans des contextes renvoyant à d'autres domaines que la physique des particules, dans le SC-P. De ce point de vue, nous pouvons faire l'hypothèse que ces deux termes ne sont pas transférés dans la presse par le biais de textes relevant du domaine de la physique des particules uniquement, mais également par le biais de textes relevant des différents domaines que nous

identifications dans ces exemples, c'est-à-dire la médecine, le spatial et la cryptographie quantique pour *photon* et la médecine pour *proton*. C'est ce que représente la figure 8.1 ci-dessous.

Dans cette figure, nous illustrons schématiquement le rôle de textes de différents domaines dans le processus de déterminologisation, à partir des exemples cités. Sur la partie gauche de la figure sont représentés des textes spécialisés qui relèvent de différents domaines dans lesquels les termes *proton* et *photon* sont présents et sur la partie droite est représentée la presse généraliste. Le passage des termes *proton* et *photon* dans la presse, à partir de textes de ces différents domaines, est représenté par des flèches (trait plein pour *proton* et pointillé pour *photon*). Cette figure illustre également le fait que, puisque la presse ne traite pas de sujets limités strictement à un domaine en particulier (contrairement aux textes spécialisés), alors elle est susceptible de contenir des usages associés aux différents domaines dans lesquels les termes sont utilisés.

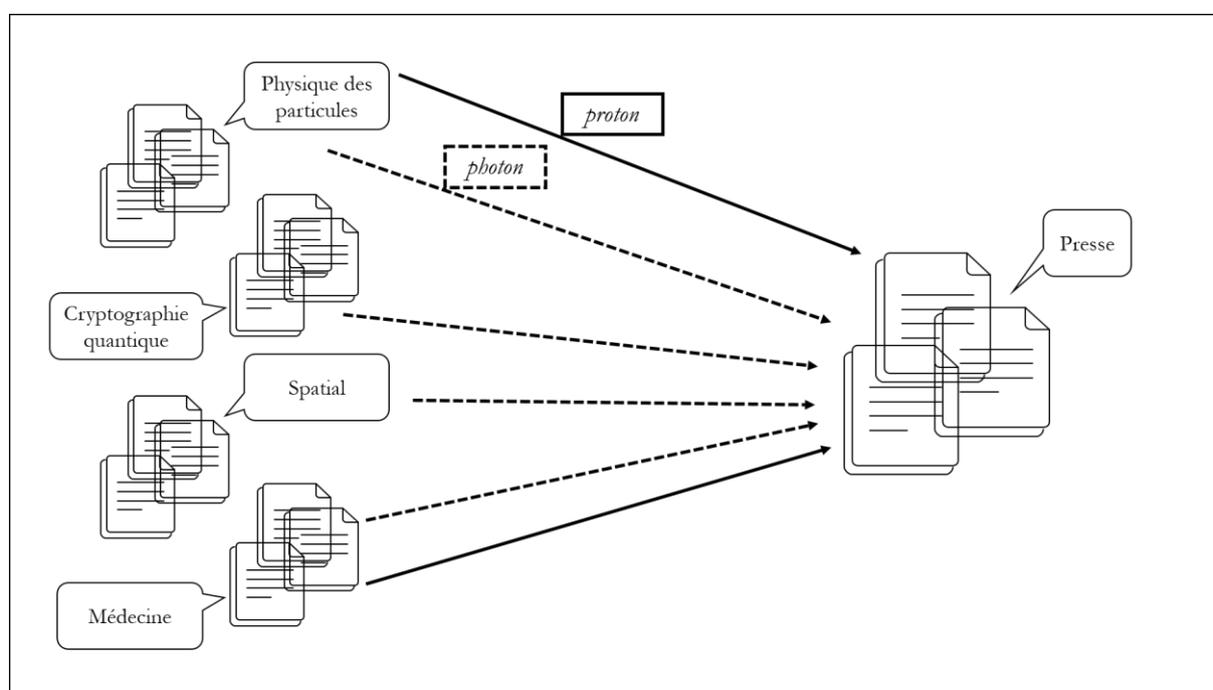


Figure 8.1 : Représentation schématique du passage de termes dans la presse à partir de textes de domaines variés

La prise en compte des sous-corpus intermédiaires permet ainsi de montrer la complexité du processus de déterminologisation. En effet, le fait que la recherche est souvent interdisciplinaire, qu'elle peut faire intervenir des chercheurs de différentes spécialités, ou encore que certains domaines peuvent démontrer un caractère transversal sont des facteurs qui peuvent avoir une certaine influence sur la déterminologisation (voir *infra*, 8.3.2). Ces constats permettent également de mieux saisir le fonctionnement du processus et, paradoxalement, d'apporter des éléments de réflexion supplémentaires par rapport aux études existantes de ce phénomène. De fait, notre approche focalisée sur le continuum entre langue de spécialité et langue générale et sur la prise en

compte de différents intermédiaires met en lumière la nécessité de tenir compte de l'interdisciplinarité de la recherche qui caractérise la majorité des domaines de recherche scientifique. En outre, notre approche permet également de mettre en évidence le décalage entre la délimitation d'un domaine d'étude en terminologie, théoriquement et méthodologiquement nécessaire mais qui ne renvoie qu'une image imparfaite de la réalité, et la complexité réelle des phénomènes étudiés, vu la diversité des domaines que nous identifions dans notre étude de la déterminologisation en physique des particules. Ainsi, bien qu'en terminologie les termes soient appréhendés à l'intérieur d'un domaine précis, aborder la question du sens des termes dans la déterminologisation impose de déconstruire les limites établies dans un objectif pratique entre chaque domaine, ou du moins de tenir compte de la complexité des dynamiques langagières réelles, qui dépassent ce cloisonnement des domaines.

À la lueur de ces éléments, il semble alors qu'une représentation plus complète de la déterminologisation est désormais possible, à travers la prise en compte de types d'intermédiaires plus diversifiés, que ce soit à l'intérieur d'un même domaine ou par le biais de domaines différents, voire par le biais d'œuvres de fiction. En outre, cette diversité dans les intermédiaires possibles permet d'interroger d'autres aspects de la déterminologisation, qui remettent en question la linéarité supposée de ce processus et qui tiennent compte des différences de comportement de chaque terme interrogé dans notre travail.

8.3. Redéfinir la déterminologisation comme un processus non linéaire

8.3.1. Reconsidérer la continuité du processus

L'hétérogénéité des phénomènes observés dans le corpus ainsi que le rôle de différents types d'intermédiaires dans la déterminologisation nous amènent à aborder ce processus sous un angle nouveau. En effet, au Chapitre 2 nous avons montré comment nous tirons profit des travaux sur la déterminologisation pour définir ce phénomène comme un double processus, à la fois dans le temps et dans le continuum entre langue de spécialité et langue générale. Dans cette perspective, nous avons fait l'hypothèse que les termes intègrent petit à petit la langue générale et que les changements sémantiques se produisent également progressivement à travers ce continuum, de la même manière qu'ils s'effectuent progressivement dans le temps. Or cette idée de progression est en réalité plus complexe.

Il apparaît donc désormais nécessaire de reconsidérer la définition de la déterminologisation comme un processus linéaire entre langue de spécialité et langue générale. Cette idée de linéarité

est d'ailleurs probablement véhiculée par le fait de considérer la déterminologisation comme un processus, étant donné que la notion de processus implique généralement une certaine continuité (cf. Chapitre 2, 2.3.2). Cependant, la diversité des intermédiaires susceptibles d'intervenir dans la déterminologisation et la complexité des phénomènes sémantiques qui s'y produisent nous amènent à définir la déterminologisation comme un passage non linéaire de termes dans la langue générale. Ce passage peut s'effectuer par le biais de différents types d'intermédiaires, comme nous l'expliquons dans la section précédente. Par ailleurs, il semble que ce passage ne s'effectue pas nécessairement à partir des textes les plus spécialisés dans un domaine, mais fait potentiellement intervenir toute une variété de textes produits à l'intérieur de plusieurs domaines, connexes ou non, dans lesquels les termes occupent une certaine place. Aussi, les caractéristiques des différents genres pris en compte et des domaines dans lesquels ils s'inscrivent sont susceptibles d'avoir une influence sur l'usage des termes constaté dans la presse et contribuent donc aux changements sémantiques des termes qui se produisent au cours du processus de déterminologisation.

Parallèlement, si les observations détaillées dans les Chapitres 6 et 7 permettent d'identifier différents intermédiaires, elles ne permettent en revanche pas d'observer un certain ordre ou une certaine hiérarchie dans ces intermédiaires. En effet, bien qu'une certaine progression ait été mise en lumière dans la section 8.2.1, tous les sous-corpus intermédiaires ne contribuent pas de la même manière. Cela signifie qu'un terme n'intégrerait probablement pas la langue générale en suivant un seul « chemin » dont le point de départ serait représenté par les textes les plus spécialisés d'un domaine et le point d'arrivée serait la langue générale, mais intégrerait plutôt la langue générale en suivant plusieurs chemins, qui dépendraient des différents types d'intermédiaires dans lesquels ce terme est susceptible d'apparaître. En outre, nous pensons que ces chemins ne seraient pas totalement indépendants les uns des autres, mais seraient en réalité susceptibles de se croiser, d'interagir, même de fusionner ou de se scinder selon les évolutions propres à chacun des termes.

Ces considérations semblent valables pour les changements sémantiques observés, qui s'effectuent également de manière non linéaire au cours du processus. Cela signifie que le sens des termes dans la presse (ou dans la langue générale) se construit potentiellement à partir des usages observables dans les situations de communication représentées par tous les sous-corpus intermédiaires. C'est d'ailleurs ce que suggèrent nos observations dans le corpus, notamment lorsque certaines dimensions sont observées dans un sous-corpus intermédiaire et dans le SC-P, mais non dans le SC-S ni dans les autres sous-corpus intermédiaires. Ainsi, à chaque étape (consécutive aux autres étapes ou qui intervient « en parallèle » des autres étapes), l'usage des termes peut se colorer de nouvelles nuances ou de nouvelles connotations, par exemple, qui peuvent toutes contribuer à la construction du sens des termes dans un contexte non spécialisé.

De ce point de vue, nous pensons donc que la non-linéarité du processus de déterminologisation se caractérise non seulement par le fait que nombre d'intermédiaires de différents types sont susceptibles d'intervenir, mais également par le fait que chaque terme peut emprunter une multiplicité de parcours pour passer dans la langue générale. Ces parcours peuvent être influencés par différents éléments, mis en évidence à plusieurs reprises dans notre travail. Nous faisons le choix de les appeler des *facteurs de déterminologisation*.

8.3.2. Définir des facteurs de déterminologisation

Avant de détailler et d'illustrer les différents facteurs dont dépend l'intégration d'un terme dans la langue générale, nous souhaitons formuler deux remarques. La première concerne le fait que, bien que nous traitions séparément ces facteurs, ceux-ci semblent intervenir simultanément dans le processus de déterminologisation et sont étroitement liés les uns aux autres, du moins est-il difficile de les dissocier complètement. La seconde remarque concerne le fait que l'identification de ces facteurs repose à la fois sur la littérature sur la circulation des termes et la déterminologisation et sur l'exploration du corpus. Ainsi, nos observations permettent de préciser les raisons, de différentes natures, sous-tendant l'intégration de certains termes dans la langue générale plutôt que d'autres. Cinq facteurs sont repérés, qui ont trait :

- à la centralité d'un terme dans différents domaines ;
- aux enjeux de la recherche pour la société ;
- aux dynamiques propres à chaque domaine ;
- à l'intérêt réel ou supposé des locuteurs pour certaines thématiques de recherche ;
- et à la morphologie des termes.

Le premier facteur se rapporte aux différents domaines dans lesquels un terme est utilisé. En effet, lorsqu'un terme représente une certaine utilité dans plusieurs domaines ou lorsqu'il intervient dans des applications liées à plusieurs domaines, alors ce terme est susceptible d'intégrer la langue générale par le biais de textes relevant de ces différents domaines. De ce point de vue, il n'est pas sûr que le même terme passe dans ces textes pour intégrer la langue générale au même rythme et il semble plus probable que l'évolution s'effectue de manière parallèle dans les différents domaines concernés. Ce facteur fait écho aux questions liées à l'interdisciplinarité de la recherche scientifique, voire à la transversalité de certains domaines. Ainsi, lorsqu'un terme est susceptible d'être utilisé dans différents domaines, ou lorsqu'il représente une certaine utilité pour différents domaines, ou encore lorsqu'il intervient dans des applications développées dans d'autres domaines

que la physique des particules, alors les canaux possibles par lesquels ce terme peut atteindre les locuteurs non spécialistes sont multipliés, favorisant ainsi son intégration dans la langue générale.

Le second facteur concerne les enjeux de certaines thématiques scientifiques pour la société. Comme le souligne Moirand (2014), ces enjeux sont le plus souvent de nature économique et politique. Aussi, lorsqu'une thématique scientifique représente plusieurs enjeux politiques ou économiques pour la société, alors il est probable que cette thématique soit particulièrement présente dans des discours de transmission des connaissances (dans le sens de Beacco et Moirand (1995), cf. Chapitre 1, 1.2.2), et notamment la presse, ce qui peut contribuer à la diffusion des termes liés à ces thématiques auprès du grand public et, par extension, à la déterminologisation. Par ailleurs, il est possible que ces enjeux soient traités de différentes manières selon le type de discours pris en compte. C'est notamment ce que nous constatons dans les verbes utilisés avec le terme *accélérateur*, dont certains permettent de saisir les enjeux financiers, pour la société, de la construction de tels instruments. Plus précisément, si le financement de la conception et de la maintenance des gros instruments dédiés à la recherche en physique des particules s'observe dans deux sous-corpus intermédiaires (SC-C et SC-R), la mention précise des coûts et le besoin de justification de ces sommes sont des caractéristiques du SC-P. Il semble donc que l'explication de certains enjeux de la recherche en physique des particules dans les médias favorise le passage des termes liés dans la langue générale. En effet, si ces enjeux représentent une certaine importance pour la société, alors il est probable qu'ils soient largement couverts dans les médias, ce qui a pour conséquence un accroissement de la fréquence des termes liés dans les médias et favorise leur intégration dans la langue générale. De la même manière, si un terme est lié à plusieurs enjeux dans les médias, alors ce terme peut atteindre la langue générale par différents biais.

Le troisième facteur est lié aux dynamiques propres à chaque domaine de recherche. En effet, la recherche ne s'effectue pas nécessairement au même rythme dans tous les domaines de recherche scientifique et toutes les découvertes ou avancées majeures ne trouvent pas forcément le même écho dans les médias. De ce point de vue, cela signifie qu'un terme utilisé dans les médias pour rapporter une découverte importante ou un évènement particulier apparaît probablement de manière récurrente, vu l'engouement de la part des médias pour les découvertes ou les avancées importantes (cf. Chapitre 1, 1.2.2). C'est ce que nous observons par exemple avec les unités qui évoquent la notion de sensationnalisme dans les contextes distributionnels des termes tels que *boson de Higgs* ou *accélérateur de particules* et *LHC*. Dans ces cas, ce sont précisément les aspects extraordinaires liés à la découverte du boson de Higgs et à la construction du LHC qui sont spectaculaires du point de vue des médias et qui favorisent l'apparition récurrente de ces termes dans les médias.

En outre, le même terme peut être associé à des avancées importantes, voire spectaculaires, dans plusieurs domaines à la fois. C'est par exemple le cas du terme *accélérateur de particules*, qui est attesté dans des contextes du SC-P rendant compte d'analyses inédites d'œuvres d'art au Louvre. Le traitement dans la presse d'évènements importants dans plusieurs domaines, qui entraînent l'utilisation des mêmes termes, contribue ainsi fortement à la diffusion de ces termes. Par exemple, d'un point de vue diachronique, Condamines et Picton (2014a : 178) montrent que lorsque certaines actualités du domaine spatial sont d'une importance particulière, celle-ci peut être reflétée par un accroissement de la fréquence de termes liés à ces actualités dans la presse. Dans cette perspective, lorsqu'un terme est utilisé pour relater plusieurs actualités fortes dans différents domaines, alors il est probable que la fréquence de ce terme dans la presse augmente de façon significative pour chacune de ces actualités. Dans ce cas, néanmoins, seule une analyse en diachronie permettrait d'étayer cette hypothèse¹⁷⁴.

Par ailleurs, toutes les avancées ne présentent pas nécessairement le même intérêt pour le grand public. Ainsi, le quatrième facteur concerne l'intérêt du public pour la science de manière générale et en particulier la fascination qu'exercent certains domaines sur les non-spécialistes. Ce facteur repose sur une hypothèse similaire à celle du facteur précédent : si un terme est utilisé pour rendre compte d'une thématique ou d'un évènement qui intéresse particulièrement le grand public selon les médias, alors sa fréquence dans les médias peut être accrue, ce qui contribue à sa diffusion auprès des locuteurs non spécialistes. Cette hypothèse a par ailleurs été détaillée dans le Chapitre 2 (2.1.1.3), mais il semble que nos observations permettent de la confirmer et de mieux saisir comment cette idée d'intérêt supposé du grand public se réalise dans les textes.

En effet, nous avons mis en évidence à plusieurs reprises dans les Chapitres 6 et 7 différentes manières de tenir compte de cet intérêt. Par exemple, il est reconnu que la presse, et les médias de manière générale, cherchent à capter l'attention des lecteurs, par le recours à différentes stratégies, notamment par un traitement sensationnaliste de certaines informations. Cette caractéristique s'expliquerait par le fait que les médias considèrent que ce sont les évènements les plus incroyables, les plus extraordinaires, les plus spectaculaires qui intéressent le public (cf. Chapitre 4, 4.1.1.1). De ce point de vue, un traitement sensationnaliste de la part de la presse participerait à rendre certains évènements plus intéressants pour les lecteurs.

De la même manière, si « le fait qu'un terme se diffuse dans la langue générale est souvent lié à un intérêt accru du public/interlocuteur pour la thématique à laquelle est lié ce terme » (Condamines et Picton, 2014a : 168), alors l'intérêt du grand public pour certaines thématiques

¹⁷⁴ Une réflexion dans cette direction est initiée dans Picton *et al.* (2021).

scientifiques peut être observé par la récurrence du traitement de ces thématiques dans la presse. C'est ce que nous constatons avec les termes *accélérateur de particules*, *boson de Higgs* ou *matière noire* par exemple, qui reviennent de manière régulière dans les différents phénomènes présentés dans les Chapitres 6 et 7. Aussi, nous pouvons en déduire qu'il est très probable que les thématiques liées à ces termes intéressent fortement le grand public.

Un autre moyen de tenir compte de cet intérêt a été relevé dans des contextes où l'utilité de certains concepts ou de certaines découvertes de physique des particules pour d'autres domaines de recherche et pour la société est explicitée. Cette hypothèse est par ailleurs également formulée par Meyer (2000 : 46), qui explique que « certain computer concepts capture the attention of the media and general public, probably because of exciting future applications ». Ainsi, nous pouvons supposer que l'explicitation des applications possibles des recherches en physique des particules dans d'autres domaines peut être un moyen de la part des médias d'attirer l'attention des lecteurs sur les recherches de ce domaine, de par l'intérêt supposé des lecteurs pour ces applications¹⁷⁵.

L'intérêt du grand public pour la science de manière générale est également particulièrement lié à la fascination que certains domaines de recherche exercent sur lui. Cet aspect ne doit pas être négligé. En effet, la fascination (supposée, du moins) du public pour un domaine ou un concept peut être mise en évidence par la présence de ce domaine ou de ce concept dans la culture de manière générale. C'est ce que nous montrons à l'aide des exemples 197 à 199 *supra*, dans un contexte littéraire, télévisuel ou cinématographique. Mais la présence de la physique des particules dans la culture peut également être appréhendée d'un point de vue artistique, par exemple dans les contextes suivants, relatifs à la sculpture et au théâtre.

201. Aujourd'hui, les bustes de la galerie Saks ne sont de loin pas une illustration de l'**antimatière** mais s'y réfèrent de manière bien plus riche. (SC-P, Le Temps, 2011)

202. Dans une esthétique bricolée où un ressort figure le **LHC** et un bonnet de bain prépare le plongeon dans l'hélium, le spectacle offre des pépites poétiques où la parole en liberté rappelle les échappées de Valère Novarina. (SC-P, Le Temps, 2012)

Enfin, un dernier facteur a été mis en évidence au cours de nos observations, qui tient compte de la morphologie des termes. En effet, nous avons constaté au Chapitre 6 que certains termes sont utilisés de manière métaphorique dans le SC-P, en particulier les termes *accélérateur de particules* et *matière noire*. Nous avons également relevé plusieurs éléments qui suggèrent que ces usages métaphoriques se basent en partie sur les composants des termes qui existent par ailleurs

¹⁷⁵ Le fait de capter l'attention du public et de susciter l'intérêt est par ailleurs un aspect important et récurrent dans les ouvrages dédiés à l'écriture journalistique (Brisson *et al.*, 2016 : Chapitre 1 ; Martin-Lagardette, 2016 : Chapitre 1 parmi d'autres).

dans la langue générale. C'est ce que nous proposons d'aborder *supra* par le biais du potentiel métaphorique des termes dans la presse.

Cette observation rejoint partiellement une hypothèse de Meyer et Mackintosh, selon laquelle les termes qui sont simples en apparence seraient plus aisément adoptés par les locuteurs non spécialistes, par exemple des termes créés par la terminologisation d'unités de la langue générale et qui sont donc le plus souvent métaphoriques (Meyer et Mackintosh, 2000b : 127-128). Une différence doit toutefois être soulevée : Meyer et Mackintosh soulignent ici le potentiel des termes d'être adoptés par des non-experts, indépendamment des changements sémantiques pouvant intervenir dans le processus. Or, dans notre cas, nous souhaitons insister sur la diversité de ces changements et notamment sur les usages métaphoriques constatés dans nos données. Ainsi, nous mettons l'accent sur le fait que, si un terme est composé d'unités existant indépendamment dans la langue générale (que le terme soit métaphorique ou non), alors ce terme est susceptible de donner lieu à des usages créatifs dans la langue générale, par le biais de la presse. Il possède donc un potentiel métaphorique plus important que d'autres termes, qui ne sont pas créés à partir d'unités de la langue générale.

Dans cette perspective, puisque certains termes ont un potentiel métaphorique plus important que d'autres, nous pensons que c'est précisément ce potentiel qui constitue un facteur influençant leur intégration dans la langue générale. En effet, puisque la presse joue un rôle important dans le processus de déterminologisation et puisqu'elle fait usage d'un langage plus imagé que d'autres genres, alors il semble que le choix de ces termes par les journalistes soit privilégié du fait qu'ils possèdent un plus grand pouvoir évocateur et sont donc plus propices à des usages métaphoriques. En ce sens, la presse représente un intermédiaire privilégié par lequel ces usages sont diffusés auprès du grand public, voire sont créés, puisqu'ils ne sont attestés que dans le SC-P. Ainsi, le potentiel métaphorique de certains termes contribuerait à la diffusion de ces termes auprès des non-spécialistes.

Ces différents facteurs permettent donc d'appréhender la complexité du processus de déterminologisation et des changements sémantiques des termes. Nous pensons plus précisément que c'est la combinaison de ces facteurs qui permet de tenir compte de la diversité des fonctionnements des termes observés dans le SC-P et d'en rendre compte, à travers des aspects aussi bien linguistiques qu'extralinguistiques. Ainsi, plusieurs facteurs peuvent expliquer le passage d'un terme dans la langue générale. Ces facteurs n'interviennent pas forcément au même niveau, mais ils restent, pour la plupart, étroitement liés les uns aux autres (il semble par exemple évident qu'il faille envisager l'actualité d'un domaine et l'intérêt du grand public pour ce domaine comme deux facteurs se répondant ou ayant une influence mutuelle). Parallèlement, tous les facteurs ne

concernent pas nécessairement les mêmes termes et l'intégration d'un terme n'est pas nécessairement influencée par l'ensemble de ces facteurs. Par exemple, à partir de nos observations, il semble que tous les facteurs participent au passage du terme *accélérateur de particules* dans la langue générale :

- ce terme est utilisé dans des domaines différents de la physique des particules, comme la médecine ou l'énergie nucléaire,
- il apparaît de manière récurrente dans des contextes explicitant les enjeux financiers de la construction du LHC, de par les importantes sommes nécessaires dans ce but,
- il apparaît également de manière récurrente dans des contextes évoquant la mise en marche du LHC après des années de construction, ce qui a été un événement important pour la physique des particules, ainsi que dans des contextes qui traitent de la recherche et de la découverte du boson de Higgs, puisque c'est au LHC que cette particule a été observée,
- l'intérêt pour des thématiques liées à ce terme est mis en évidence à plusieurs reprises, notamment lorsque celles-ci sont en lien avec des applications dans d'autres domaines, mais également lorsque le terme est utilisé dans un contexte de fiction,
- ce terme possède un fort potentiel métaphorique.

En revanche, moins de facteurs peuvent être identifiés en ce qui concerne le passage du terme *photon*. Nous remarquons en effet que ce terme s'intègre dans des contextes qui renvoient à d'autres domaines (comme la médecine ou le spatial), qu'il est étroitement lié à certaines thématiques représentant des enjeux pour la société, par exemple l'utilisation de photons dans un but de cryptographie, et que certaines applications faisant intervenir des photons suscitent l'intérêt des non-spécialistes, par exemple dans la cryptographie ou dans les télécommunications (les photons sont utilisés dans les réseaux de fibre optique).

Enfin, pour ce qui est du terme *boson de Higgs*, il semble que ce soit essentiellement le facteur lié à l'actualité du domaine de la physique des particules qui ait une influence sur son intégration dans la langue générale. Là aussi, nous pouvons considérer que les facteurs concernant les enjeux pour la société (si nous tenons compte des coûts de la construction des instruments dédiés à la recherche et à l'étude de cette particule) et l'intérêt du grand public interviennent également, puisque ces trois facteurs sont étroitement liés. En revanche, nous ne relevons aucune occurrence dans le SC-P où le terme *boson de Higgs* apparaît dans des contextes qui renvoient à d'autres

domaines ou qui explicitent l'utilité de ce terme pour un autre domaine. De la même manière, aucun usage métaphorique n'a été repéré.

Le comportement de chaque terme peut donc être très différent selon ces facteurs, ce qui révèle, selon nous, une autre facette de la complexité de la déterminologisation. En effet, nous expliquons *supra* que différents types d'intermédiaires sont susceptibles d'intervenir dans la diffusion des termes auprès des locuteurs non spécialistes et dans les changements sémantiques observables dans le SC-P. Cependant, la mise en lumière des facteurs de déterminologisation permet d'ajouter une nouvelle variable à l'appréhension du processus, qui peut s'effectuer de diverses manières selon le terme considéré. Il semble donc que l'ensemble de ces éléments illustre le caractère non linéaire de la déterminologisation.

À l'issue de ce chapitre, les trois axes de réflexion que nous avons détaillés mettent en évidence la nécessité de considérer différents niveaux d'analyse et de description du fonctionnement des termes en situation de déterminologisation, afin de rendre compte de la complexité de ce phénomène, à la fois comme processus et comme résultat. Nous avons particulièrement montré que cette complexité ne concerne pas uniquement les manifestations de la déterminologisation dans la presse, qui se caractérisent par une forte diversité, mais également le processus de passage des termes dans la langue générale. Tous ces éléments participent alors à redéfinir la déterminologisation comme un processus non linéaire de passage de termes dans la langue générale, qui dépend de plusieurs facteurs et qui s'effectue par le biais de différents types d'intermédiaires.

Conclusion

Dans notre thèse, nous avons présenté les enjeux d'une définition de la déterminologisation qui tient compte de son double statut de processus et de résultat. Partant du constat de la nécessité de considérer la déterminologisation également comme un processus de passage de termes d'une langue de spécialité à la langue générale, nous avons cherché à observer comment ce processus se manifeste dans les textes, à l'intérieur du continuum formé par les langues de spécialité et la langue générale.

Afin d'aborder ces questionnements, nous avons commencé par replacer la déterminologisation parmi les notions apparentées qui traitent de la problématique de la circulation des termes en dehors de la sphère spécialisée et de leur fonctionnement dans la langue générale. Dans une réflexion centrée sur la déterminologisation, nous avons montré la transversalité de ce phénomène, qui entretient des rapports étroits aussi bien avec les questions de vulgarisation scientifique et de diffusion des connaissances qu'avec des questions liées à néologie et à la lexicologie. Nous avons ensuite approfondi notre réflexion, en tirant parti des travaux qui ont développé la notion de déterminologisation, du point de vue de la description des termes dans un contexte non spécialisé, de la perspective diachronique et du rôle de textes « intermédiaires ». Nous avons repris à notre compte certaines hypothèses détaillées dans ces travaux, afin de mieux circonscrire la question de la déterminologisation et de nourrir notre réflexion sur les possibilités méthodologiques d'observation de ce processus et sur les données à prendre en considération dans ce but.

Nous avons ensuite mis à profit les méthodes de la linguistique outillée et les principes de la terminologie textuelle pour définir une méthode d'observation du processus de déterminologisation dans le domaine de la physique des particules. Ce parti pris nous a permis d'ancrer notre travail de recherche dans une démarche reproductible ; nous avons assuré la reproductibilité par le recours à une méthode de constitution d'un corpus comparable complexe basée sur des critères transparents et par la définition d'indices objectifs. La particularité du corpus que nous avons compilé réside dans l'organisation des textes en cinq sous-corpus, pensés pour approcher la représentation du continuum entre langue de spécialité et langue générale. Enfin, par la richesse des phénomènes analysés grâce à ces indices et à travers l'ensemble de notre démarche, nous avons pu insister sur la nécessité de construire une interprétation solide qui repose sur des faits objectifs et sur l'intérêt de la linguistique outillée pour l'interrogation de problématiques terminologiques complexes, requérant des analyses linguistiques fines. Par la mise en œuvre systématique des indices dans le corpus, nous avons également illustré la diversité des

fonctionnements des termes et nous avons mis au jour plusieurs éléments qui nous permettent de tendre vers une meilleure compréhension de la déterminologisation.

L'apport de notre travail de thèse s'articule principalement autour de deux axes : un axe méthodologique et un axe descriptif/théorique. D'un point de vue méthodologique, nous avons proposé une réflexion détaillée sur la possibilité d'observer un processus en corpus comparable. L'originalité de notre réflexion est illustrée à travers la perspective d'analyse choisie : afin de limiter les biais inhérents à la représentation d'un processus dans des données figées, nous avons cherché à approcher le continuum langue de spécialité-langue générale dans le corpus. Dans ce but, nous avons construit un corpus composé de cinq sous-corpus, dont deux visent à représenter les extrémités du continuum, par le recours à des textes relevant d'un degré de spécialisation élevé et par le recours à des textes caractéristiques d'une situation de communication non spécialisée. Les trois autres sous-corpus se situent entre ces deux extrémités et tiennent compte de textes relevant d'une situation de communication caractérisée par un degré de spécialisation intermédiaire. La multiplication des points de comparaison dans le continuum a garanti l'adéquation du corpus pour le continuum et l'analyse de ces données selon un schéma précis nous a permis de saisir les dynamiques langagières propres à la déterminologisation. Nous avons démontré la pertinence de notre proposition par l'exploration de deux indices dans le corpus et par la mise au jour et la caractérisation de différents fonctionnements des termes en situation de déterminologisation.

Notre choix de baser l'analyse sur deux approches de la distribution en corpus repose sur la complémentarité de ces approches :

- avec le premier indice, fondé sur les contextes des termes de type « fenêtre de mots », nous avons principalement mis en évidence des différences de nature thématique dans les environnements distributionnels des termes ;
- avec le second indice, fondé sur les contextes syntaxiques des termes, nous avons saisi plus finement les différences de fonctionnement des termes dans le corpus, notamment des différences de conceptualisation ou de point de vue.

À travers la mise en œuvre de ces indices, nous avons également insisté sur la nécessité d'adapter, voire de détourner, certains outils ou certaines méthodes, afin d'être en mesure d'aborder précisément les phénomènes sur lesquels porte notre recherche. Par exemple, l'utilisation détournée d'un repérage automatique de *keywords* avec le premier indice a permis de cibler l'analyse sur les différences les plus saillantes dans les contextes des termes, de manière peu coûteuse en temps. Parallèlement, avec ce choix méthodologique, nous avons pu illustrer l'importance

d'articuler ces indices avec des données quantitatives, afin de fournir des éléments objectifs pour la comparaison des cinq sous-corpus.

D'un point de vue descriptif et théorique, à travers l'observation minutieuse du fonctionnement des termes dans le sous-corpus *Presse*, nous avons relevé différents phénomènes, qui ne renvoient pas nécessairement aux changements sémantiques habituellement décrits dans la littérature sur la déterminologisation. En revanche, cette diversité a mis en évidence des différences plus subtiles, assimilables à des conceptualisations ou des points de vue divergents. Nous les avons rapprochées de la notion de multidimensionnalité et avons particulièrement souligné le fait qu'elles peuvent renvoyer aux différentes dimensions d'un concept, selon le domaine dans lequel il est utilisé, selon le genre textuel ou encore selon le degré de spécialisation. L'usage des termes dans un contexte non spécialisé a alors révélé la multidimensionnalité des concepts sous-jacents. De la même manière, nous avons montré comment la diversité des fonctionnements des termes dans le sous-corpus *Presse* permet d'appréhender les mécanismes donnant lieu à des emplois métaphoriques des termes. Nos données ont mis en lumière le fait que ces emplois se fondent à la fois sur les usages des termes dans la presse, en particulier lorsqu'ils véhiculent de nouvelles connotations, et sur la coexistence dans la langue générale des composants du terme ou d'unités de la même famille morphologique, de leur sens en langue générale, voire de leurs connotations. Notre réflexion nous a menée à proposer la notion de potentiel métaphorique pour aborder ces cas.

En outre, notre analyse a révélé la complexité réelle du processus de déterminologisation. La prise en compte de textes relevant de degrés de spécialisation intermédiaires et de genres textuels participant à la diffusion des connaissances a permis de confirmer leur rôle dans le processus de passage de termes dans la langue générale. Mais nous avons surtout mis en évidence le rôle potentiel de nombreux autres intermédiaires de différentes natures, qu'il s'agisse de discours de transmission des connaissances dans d'autres domaines, lorsque les termes concernés occupent une place centrale dans plusieurs domaines, ou d'intermédiaires d'une forme tout à fait différente, qui s'inscrivent dans des œuvres de fiction et reflètent dans le même temps la popularité d'un domaine de recherche dans la culture ou dans l'imaginaire commun.

L'ensemble de notre travail de thèse nous a amenée à définir la déterminologisation comme un processus complexe et non linéaire de passage de termes dans la langue générale, qui fait intervenir de nombreux intermédiaires et qui est influencé par au moins cinq facteurs. Ces facteurs dépendent :

- de la centralité d'un terme dans plusieurs domaines,

- de l'emploi d'un terme dans des discours explicitant les enjeux de certaines thématiques de recherche pour la société,
- des dynamiques propres à la recherche dans chaque domaine,
- de la place d'un terme dans des textes traitant de thématiques supposées intéresser particulièrement le grand public,
- du potentiel d'un terme de donner lieu à des emplois métaphoriques.

Cette définition large permet alors de mieux rendre compte de la diversité des comportements observés selon les termes étudiés et de mieux comprendre les raisons sous-tendant l'intégration de certains termes dans la langue générale plutôt que d'autres. Il semble en particulier que le potentiel métaphorique d'un terme soit déterminant, dans la mesure où la diffusion des emplois métaphoriques de ce terme contribue à l'émergence puis à la stabilisation d'un nouveau sens de ce terme en langue générale, et donc à l'établissement d'un néologisme sémantique. En ce sens, notre thèse permet également de reconnaître pleinement le rôle de la déterminologisation dans la néologie sémantique.

Tous ces éléments rendent alors possible une réflexion approfondie sur le processus de déterminologisation dans le domaine de la physique des particules et ouvrent, dans le même temps, de nouvelles pistes de recherche tout en posant de nouvelles questions. En effet, il n'est pas certain que le processus de déterminologisation s'effectue par le biais d'intermédiaires dans une même langue uniquement. Il semble même plus probable que des intermédiaires dans plusieurs langues contribuent à ce processus, particulièrement en anglais, puisque cette langue reste dominante dans la recherche scientifique. Le rôle de l'anglais ou d'autres langues dans le processus mériterait donc d'être interrogé, de même que le rôle de différents types de médias, et notamment les médias sociaux, puisqu'une grande partie de la communication vers le grand public passe désormais par ces médias.

Dans notre travail, nous avons proposé une méthode d'observation outillée de la déterminologisation en corpus comparable qui constitue une base solide pour pousser plus avant la réflexion sur ce processus, à partir d'autres types de données, et même dans d'autres domaines. Soulignons l'importance de ce dernier point. Nous avons insisté à plusieurs reprises sur les rapports entre les fonctionnements des termes observés dans la presse, la notion de sensationnalisme et le potentiel métaphorique. Or la physique des particules est un domaine où une part importante de la recherche est dédiée à la confirmation des modèles théoriques, ce qui passe par l'observation, à l'aide de différentes expériences, de particules dont l'existence est seulement postulée par la théorie. Comme dans le cas du boson de Higgs, la preuve expérimentale de l'existence d'une particule peut

être fournie des dizaines d'années après sa prédiction par la théorie, à l'aide d'instruments de recherche toujours plus imposants, puissants et coûteux, tels que le Grand collisionneur de hadrons au CERN. Ces caractéristiques rendent le traitement sensationnaliste de sujets liés à la recherche dans ce domaine particulièrement aisé par les médias. Il n'est pas certain qu'avec un domaine moins « spectaculaire », mais dont les termes sont tout de même susceptibles d'intégrer la langue générale, les mêmes phénomènes soient observés dans la presse. Dans le but de vérifier la validité de la méthode, il serait donc nécessaire de l'éprouver sur des données issues d'autres domaines. Cette perspective permettrait également de tester l'influence du domaine sur les fonctionnements des termes dans la presse et d'affiner les facteurs de déterminologisation mis en lumière dans notre thèse.

L'un des apports majeurs de notre thèse réside dans la clarification des rapports entre la déterminologisation et la néologie, et plus précisément dans la confirmation du rôle de la déterminologisation dans l'émergence de néologismes sémantiques issus de l'usage de termes dans la langue générale. La réflexion développée dans notre thèse repose cependant essentiellement sur l'observation détaillée de deux termes. Parallèlement, certaines métaphores établies dans l'usage que nous nous attendions à observer n'ont pas été repérées par la méthode mise en œuvre, notamment le cas du terme *électron libre*. Aussi, afin de donner plus de poids à nos conclusions et permettre des généralisations à partir de données plus nombreuses, notre réflexion devrait être complétée par l'analyse approfondie d'autres termes employés métaphoriquement dans la presse et par la prise en compte d'un plus grand nombre d'occurrences. D'un point de vue méthodologique, les indices définis et explorés dans notre thèse devraient donc être ajustés pour cibler la méthode plus précisément sur le repérage des usages métaphoriques des termes et pour focaliser la recherche sur leur analyse. Dans le même temps, cette proposition permettrait également de développer davantage la réflexion sur le potentiel métaphorique des termes et, plus largement, sur l'influence de la morphologie du terme sur son potentiel d'intégration dans la langue générale.

Dans notre travail, nous avons posé certains jalons pour l'étude outillée en corpus du fonctionnement des termes dans le processus de déterminologisation, d'un point de vue tant théorique que méthodologique. Notre thèse représente alors un apport important pour les champs de la linguistique outillée et de la terminologie textuelle. Enfin, d'un point de vue théorique, nous avons voulu montrer l'ampleur des questions liées au processus d'intégration et au fonctionnement des termes dans la langue générale. En ce sens, notre thèse contribue à reconnaître la transversalité de la déterminologisation dans les questions de circulation des termes, de diffusion des connaissances et de néologie. Nous espérons ainsi avoir démontré par ce travail l'intérêt et la nécessité de poursuivre la réflexion dans cette direction.

Bibliographie

- Abeillé, A. (Éd.). (2003). *Treebanks: Building and Using Parsed Corpora*. Dordrecht: Springer Netherlands.
- Abeillé, A., Clément, L., & Toussnel, F. (2003). Building a Treebank for French. Dans A. Abeillé (Éd.), *Treebanks: Building and Using Parsed Corpora* (pp. 165-187). Dordrecht: Springer Netherlands.
- Adelstein, A. (1996). Banalización de términos con formantes de origen grecolatino. Dans *Actas del V Simposio de Ríterm*. Ciudad de México: Colegio de México; Unión Latina.
- Adelstein, A., & Cabré, M. T. (2002). The Specificity of Units with Specialized Meaning: Polysemy as Explanatory Factor. *D.E.L.T.A.*, 18, 1-25.
- AFP (Agence France-Presse). (2004). *Manuel de l'agencier*. Paris: AFP.
- Ahmad, K., & Musacchio, M. T. (2003). Enrico Fermi and the Making of the Language of Nuclear Physics. *Fachsprache*, 25(3-4), 120-140.
- Aijmer, K., & Altenberg, B. (Éds). (1991). *English Corpus Linguistics*. Londres: Routledge.
- Alecu, B. P., Thomas, I., & Renahy, J. (2012). La « multi-extraction » comme stratégie d'acquisition optimisée de ressources (non) terminologiques. Dans *Actes de la conférence conjointe JEP-TALN-RECITAL 2012* (pp. 511-518). Grenoble, France.
- Alexandru, C., & Gaudin, F. (2006). Les contextes : À la source du terme ? Dans *Mots, termes et contextes. Actes des septièmes Journées scientifiques du réseau de chercheurs « Lexicologie, Terminologie, Traduction » (LTT 2005)* (pp. 59-67). Paris: Éditions des archives contemporaines.
- Allauzen, A., & Bonneau-Maynard, H. (2008). Training and Evaluation of POS Taggers on the French MULTITAG Corpus. Dans *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008*. Marrakech, Maroc.
- Alonso Campo, A. (2008). Environmental Terminology in General Dictionaries. Dans E. Bernal & J. DeCesaris (Éds), *Proceedings of the XIII Euralex International Congress* (pp. 929-935). Barcelona: Institut universitari de lingüística aplicada (IULA).
- Altmanova, J., Centrella, M., & Russo, K. E. (Éds). (2018). *Terminology & Discourse / Terminologie et discours*. Berne: Peter Lang.
- Alves, I. M. (Éd.). (2010). *Neologia e neologismos em diferentes perspectivas*. São Paulo: Paulistana.
- Anthony, L. (2005). AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom. Dans *IEEE International Professional Communication Conference Proceedings* (pp. 729-737).
- Anthony, L. (2018). *AntConc (Version 3.5.6)*. Tokyo, Japon: Waseda University. <https://www.laurenceanthony.net/software/antconc/> [consulté le 17.06.2018]
- Antoine, G., & Martin, R. (Éds). (1995). *Histoire de la langue française*. Paris: CNRS Éditions.

- Assal, A. (1995). La métaphorisation terminologique. *L'Actualité terminologique*, 28(2), 22-24.
- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus Design Criteria. *Literary and Linguistic Computing*, 7(1), 1-16.
- Atkins, S., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Atkins, S., & Zampolli, A. (Éds). (1994). *Computational Approaches to the Lexicon*. Oxford: Oxford University Press.
- ATS (Agence télégraphique suisse). (2008). *Manuel de l'agencier*. Berne: ATS.
- Aussenac-Gilles, N., Bourigault, D., & Teulier, R. (2003). Analyse comparative de corpus : Cas de l'ingénierie des connaissances. Dans *Actes de la Conférence Ingénierie des Connaissances 2003*. Laval, France.
- Aussenac-Gilles, N., & Condamines, A. (2007). Corpus et terminologie. Dans R. T. Pédaque (Éd.), *La redocumentarisation du monde* (pp. 131-147). Toulouse: Cepadues Éditions.
- Aussenac-Gilles, N., Condamines, A., & Szulman, S. (2002). Prise en compte de l'application dans la constitution de produits terminologiques. Dans *Actes des 2èmes assises nationales du groupe de recherche I3 (Information, Interaction, Intelligence)* (pp. 289-303). Cépaduès, France.
- Baayen, R. H., & Shafei-Bajestan, E. (2019). languageR: Analyzing Linguistic Data: A Practical Introduction to Statistics (version 1.5.0). <https://cran.r-project.org/package=languageR> [consulté le 4.05.2019]
- Baker, P., & Egbert, J. (Éds). (2018). *Triangulating Methodological Approaches in Corpus Linguistic Research*. New York, Londres: Routledge.
- Bakhtine, M. (1984). *Esthétique de la création verbale*. Paris: Gallimard.
- Balvet, A., Stosic, D., & Miletic, A. (2014). TALC-sef, Un corpus étiqueté de traductions littéraires en serbe, anglais et français. Dans F. Neveu, P. Blumenthal, L. Hriba, A. Gerstenberg, J. Meinschaefer, & S. Prévost (Éds), *4e Congrès Mondial de Linguistique française* (pp. 2551-2563). Berlin.
- Baneyx, A., Malaisé, V., Charlet, J., Zweigenbaum, P., & Bachimont, B. (2005). Synergie entre analyse distributionnelle et patrons lexico-syntaxiques pour la construction d'ontologies différentielles. Dans *Conférence TLA-2005*. Rouen.
- Bannour, S., Audibert, L., & Nazarenko, A. (2011). Mesures de similarité distributionnelle entre termes. Dans *IC 2011, 22èmes Journées francophones d'Ingénierie des Connaissances* (pp. 523-538). Chambéry, France.
- Barona, J. L. (1998-1999). Sciences, Language and Social Interactions. *Terminology*, 5(1), 107-119.
- Basile, P., & McGillivray, B. (2018). Exploiting the Web for Semantic Change Detection. Dans L. Soldatova, J. Vanschoren, G. Papadopoulos, & M. Ceci (Éds), *Discovery Science* (Vol. 11198, pp. 194-208). Cham: Springer.

- Bassi, E. (2010). A Contrastive Analysis of Keywords in Newspaper Articles on the « Kyoto Protocol ». Dans M. Bondi & M. Scott (Éds), *Keyness in Texts* (pp. 207-218). Amsterdam, Philadelphie: John Benjamins.
- Beacco, J.-C. (1999). *L'astronomie dans les médias. Analyses linguistiques de discours de vulgarisation*. Paris: Presses Sorbonne nouvelle.
- Beacco, J.-C. (2000). Écriture de la science dans les médias. *Les Carnets du Cediscor*, 6, 15-24.
- Beacco, J.-C., Claudel, C., Doury, M., Petit, G., & Reboul-Touré, S. (2002). Science in Media and Social Discourse: New Channels of Communication, New Linguistic Forms. *Discourse Studies*, 4(3), 277-300.
- Beacco, J.-C., & Moirand, S. (1995). Autour des discours de transmission des connaissances. *Langages*, 117, 32-53.
- Béjoint, H. (1988). Scientific and Technical Words in General Dictionaries. *International Journal of Lexicography*, 1(4), 354-368.
- Béjoint, H., & Maniez, F. (Éds). (2005). *De la mesure dans les termes*. Lyon: Les Presses Universitaires de Lyon.
- Béjoint, H., & Thoiron, P. (2000a). Le sens des termes. Dans H. Béjoint & P. Thoiron (Éds), *Le sens en terminologie* (pp. 3-19). Lyon: Les Presses Universitaires de Lyon.
- Béjoint, H., & Thoiron, P. (Éds). (2000b). *Le sens en terminologie*. Lyon: Les Presses Universitaires de Lyon.
- Bernhard, D., Boisseau, M., Gérard, C., Grass, T., & Todirascu, A. (Éds). (2018). *La néologie en contexte. Cultures, situations, textes*. Paris: Lambert-Lucas.
- Bernier, M.-F. (2004). *Ethique et déontologie du journalisme*. Sainte-Foy: Les Presses de l'Université Laval.
- Bernier-Colborne, G. (2014). Analyse distributionnelle de corpus spécialisés pour l'identification de relations lexico-sémantiques. Dans *Actes de SemDis 2014*. Marseille, France.
- Bernier-Colborne, G., & Drouin, P. (2016). Evaluation of Distributional Semantic Models: A Holistic Approach. Dans *Proceedings of the 5th International Workshop on Computational Terminology* (pp. 52-61). Osaka, Japon.
- Bertaccini, F., Lecci, C., & Bono, V. (2008). Processi di terminologizzazione e determinologizzazione nel dominio della diffusione e distribuzione del libro. *AIDA informazioni. Rivista di Scienze dell'informazione*, 26(1-2), 45-61.
- Bertels, A., & Speelman, D. (2013). 'Keywords Method' versus 'Calcul des Spécificités': A Comparison of Tools and Methods. *International Journal of Corpus Linguistics*, 18(4), 536-560.
- Bertels, A., & Speelman, D. (2016). Analyse sémantique distributionnelle dans un corpus technique : Les enjeux sémantiques dans un jeu de statistiques. Dans *JADT 2016 : 13èmes Journées internationales d'Analyse statistique des Données Textuelles*.

- Bestgen, Y. (2017). Getting Rid of the Chi-square and Log-likelihood Tests for Analysing Vocabulary Differences between Corpora. *Quaderns de Filologia: Estudis Lingüístics*, 22, 33-56.
- Bhatia, V. K. (2004). *Worlds of Written Discourses: A genre-based view*. Londres: Continuum.
- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, 8(4), 243-257.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, D., & Reppen, R. (2015a). Introduction. Dans Douglas Biber & R. Reppen (Éds), *The Cambridge Handbook of English Corpus Linguistics* (pp. 1-8). Cambridge: Cambridge University Press.
- Biber, D., & Reppen, R. (Éds). (2015b). *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.
- Bigras, I., & Simard, I. (1997). Traitement et évolution des termes médicaux dans un dictionnaire de langue générale. Dans M. C. Cormier & A. Francoeur (Éds), *Problèmes et méthodes de la lexicographie québécoise* (pp. 97-110). Québec: CIRAL.
- Bilger, M. (Éd.). (2000). *Linguistique sur corpus. Études et réflexions*. Perpignan: Les Presses de l'Université de Perpignan.
- Boguraev, B., & Pustejovsky, J. (Éds). (1996). *Corpus Processing for Lexical Acquisition*. Cambridge, Massachusetts: The MIT Press.
- Boleda, G. (2020). Distributional Semantics and Linguistic Theory. *Annual Review of Linguistics*, 6, 213-234.
- Bondi, M. (2010). Perspectives on Keywords and Keyness: An introduction. Dans M. Bondi & M. Scott (Éds), *Keyness in Texts* (pp. 1-18). Amsterdam, Philadelphie: John Benjamins.
- Bondi, M., & Scott, M. (Éds). (2010). *Keyness in Texts*. Amsterdam, Philadelphie: John Benjamins.
- Botta, M. (2013). La terminologie de l'environnement en vulgarisation scientifique: La famille lexicale de la régénération des forêts en portugais. *Equivalences*, 40(1), 277-298.
- Bouhandi, M. (2019). Apport des termes complexes pour enrichir l'analyse distributionnelle en domaine de spécialité. Dans *TALN-RECITAL 2019* (pp. 473-486). Toulouse.
- Boulanger, J.-C. (1995). Présentation : Images et parcours de la socioterminologie. *META*, 40(2), 194-205.
- Boulanger, J.-C. (1996). Les dictionnaires généraux monolingues, une voie royale pour les technolèctes. *TradTerm*, 3, 137-151.
- Boulanger, J.-C., & L'Homme, M.-C. (1991). Les technolèctes dans la pratique dictionnaire générale. Quelques fragments d'une culture. *META*, 36(1), 23-40.

- Bourigault, D. (2002). UPERY : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus. Dans *TALN 2002* (pp. 75-84). Nancy.
- Bourigault, D., & Aussenac-Gilles, N. (2003). Construction d'ontologies à partir de textes. Dans *Actes de la 10ème conférence « Traitement Automatique des Langues Naturelles »* (pp. 27-50). Bats-sur-Mer, France.
- Bourigault, D., & Fabre, C. (2000). Approche linguistique pour l'analyse syntaxique de corpus. *Cahiers de grammaire*, 25, 131-151.
- Bourigault, D., Fabre, C., Frérot, C., Jacques, M.-P., & Ozdowska, S. (2005). Syntex, analyseur syntaxique de corpus. Dans *Actes des 12èmes Journées sur le Traitement Automatique des Langues Naturelles*. Dourdan, France.
- Bourigault, D., Jacquemin, C., & L'Homme, M.-C. (Éds). (2001). *Recent Advances in Computational Terminology*. Amsterdam: John Benjamins.
- Bourigault, D., & Slodzian, M. (1999). Pour une terminologie textuelle. *Terminologies Nouvelles*, 19, 19-32.
- Boussidan, A. (2013). *Dynamics of Semantic Change. Detecting, Analyzing and Modeling Semantic Change in Corpus in Short Diachrony*. Thèse de doctorat en Sciences cognitives. Université Lumière, Lyon 2.
- Bouveret, M. (1998). Approche de la dénomination en langue spécialisée. *META*, 43(3), 393-410.
- Bowker, L. (1993). Multidimensional Classification of Concepts for Terminological Purposes. Dans *Proceedings of the 4th ASIS SIG/CR Classification Research Workshop* (pp. 39-56).
- Bowker, L. (1996). Towards a Corpus-based Approach to Terminography. *Terminology*, 3(1), 27-52.
- Bowker, L. (1997a). Multidimensional Classification of Concepts and Terms. Dans S. E. Wright & G. Budin (Éds), *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management* (pp. 133-143). Amsterdam: John Benjamins.
- Bowker, L. (1997b). You say “flatbed colour scanner”, I say “colour flatbed scanner”: A descriptive study of the influence of multidimensionality on term formation and use with special reference to the subject field of optical scanning technology. *Terminology*, 4(2), 275-302.
- Bowker, L. (2017). Lexicography and Terminology. Dans P. A. Fuertes-Olivera (Éd.), *The Routledge Handbook of Lexicography* (pp. 138-151). Londres: Routledge.
- Bowker, L., & Pearson, J. (2002). *Working with Specialized Language. A Practical Guide to Using Corpora*. Londres, New York: Routledge.
- Bravo, J.-M. (Éd.). (2004). *A New Spectrum of Translation Studies*. Valladolid: Universidad de Valladolid, Secretariado de Publicaciones e Intercambio Editorial.
- Bres, J. (2017). Dialogisme, éléments pour l'analyse. *Recherches en didactique des langues et des cultures*, 14(2). <http://journals.openedition.org/rdlc/1842> [consulté le 18.09.2020]
- Brezina, V. (2018). *Statistics in Corpus Linguistics: A practical guide*. Cambridge ; New York: Cambridge University Press.

- Brisson, P., Maltais, R., Parent, A., & Asselin, Y. (2016). *L'écriture journalistique sous toutes ses formes. L'écriture journalistique sous toutes ses formes*. Montréal: Presses de l'Université de Montréal.
- Cabré, M. T. (1994). Terminologie et dictionnaires. *META*, 39(4), 589-597.
- Cabré, M. T. (1998). *La terminologie : Théorie, méthode et applications*. Ottawa: Les Presses de l'Université d'Ottawa et Armand Colin.
- Cabré, M. T. (1999). *Terminology: Theory, Methods and Applications*. Amsterdam: John Benjamins.
- Cabré, M. T. (2003). Theories of Terminology: Their description, prescription and explanation. *Terminology*, 9(2), 163-199.
- Cabré, M. T. (2005). La terminología, una disciplina en evolución: Pasado, presente y algunos elementos de futuro. *Debate terminológico*, 1.
- Cabré, M. T. (2016). La terminologie. Dans W. Forner & B. Thörle (Éds), *Manuel des langues de spécialité* (pp. 68-81). Berlin: De Gruyter.
- Cabré, M. T., Estopà Bagot, R., & Vivaldi Palatresi, J. (2001). Automatic Term Detection: A Review of Current Systems. Dans D. Bourigault, C. Jacquemin, & M.-C. L'Homme (Éds), *Recent Advances in Computational Terminology*. Amsterdam: John Benjamins.
- Calberg-Challot, M. (2007). Quand un vocabulaire de spécialité emprunte au langage courant : Le nucléaire, étude de cas. Dans J. Humbley (Éd.), *Aspects de la recherche en langue de spécialité* (pp. 71-85). Paris: Université Paris VII.
- Calberg-Challot, M. (2015). Variation de la dénomination dans le domaine du nucléaire : Le cas du « cendrier ». Dans *Conférence-hommage à John Humbley « Quo vadis, Terminologia »*, Paris 7.
- Calberg-Challot, M., Candel, D., & Fleury, S. (2006). « Nucléaire » et « Atomique », deux formes concurrentielles dans le domaine du nucléaire ? Dans *JADT' 06 : Actes des 8es Journées internationales d'Analyse statistique des Données Textuelles* (Vol. 1, pp. 223-234). Besançon: Presses universitaires de Franche-Comté.
- Campo, Á. (2012). *The Reception of Eugen Wüster's Work and the Development of Terminology*. Thèse de doctorat en Traduction. Université de Montréal.
- Campos Andrés, O. (2013). Procedimientos de desteterminologización: Traducción y redacción de guías para pacientes. *Panace@*, 14(37), 48-52.
- Candel, D., & Gaudin, F. (Éds). (2006). *Aspects diachroniques du vocabulaire*. Rouen: Publications des Universités de Rouen et du Havre.
- Candel, D. (Éd.). (1994). *Français scientifique et technique et dictionnaire de langue*. Paris: Didier Érudition.
- Candel, D. (1995). Le vocabulaire de la physique (physique quantique, atomique et nucléaire). Dans G. Antoine & R. Martin (Éds), *Histoire de la langue française 1914-1945* (pp. 367-397). Paris: CNRS Éditions.
- Candito, M., Crabbé, B., & Denis, P. (2010). Statistical French Dependency Parsing: Treebank Conversion and First Results. Dans *Proceedings of the seventh International Conference on Language Resources and Evaluation (LREC 2010)* (pp. 1840-1847). La Vallette, Malte.

- Cañete, P., Fernández-Silva, S., & Villena, B. (2016). La difusión de la terminología a través de la prensa escrita: Un acercamiento a través de la neología de El País. Dans Observatori de Neologia (Éd.), *Mots d'avui, mots de demà* (pp. 97-114). Barcelona: IULA, Universitat Pompeu Fabra.
- Carletta, J. (1996). Assessing Agreement on Classification Tasks: The Kappa Statistics. *Computational Linguistics*, 22(2), 249-254.
- Carras, C. (2002). *Le vocabulaire économique et commercial dans la presse brésilienne (années 1991 et 1992) : Étude comparative et proposition de dictionnaire bilingue portugais/français*. Thèse de doctorat en Lexicologie, Terminologie Multilingue et Traduction. Université Lumière, Lyon 2.
- Cartier, E. (2016). Neoveille, système de repérage et de suivi des néologismes en sept langues. *Neologica*, 10, 101-131.
- Cartier, E., Sablayrolles, J.-F., Boutmgharine, N., Humbley, J., Bertocci, M., Jacquet-Pfau, C., Kübler, N., & Tallarico, G. (2018). Détection automatique, description linguistique et suivi des néologismes en corpus : Point d'étape sur les tendances du français contemporain. Dans *Congrès Mondial de Linguistique Française (CMLF 2018)*.
- Celotti, N., & Musacchio, M. T. (2004). Un regard diachronique en didactique des langues de spécialité. *Ela. Études de Linguistique Appliquée*, 135, 263-270.
- Ciapuscio, G. E. (1997). Lingüística y divulgación de ciencia. *Quark: Ciencia, Medicina, Comunicación y Cultura*, 7, 19-28.
- Ciapuscio, G. E. (2003). Formulation and Reformulation Procedures in Verbal Interaction Between Experts and (Semi)laymen. *Discourse Studies*, 5(4), 207-233.
- Clas, A. (1994). Collocations et langues de spécialité. *META*, 39(4), 576-580.
- Cobos López, Í. (2019). Traducir para el paciente: Acercamiento y adaptación como modalidad de traducción. *Quaderns de Filologia: Estudis Lingüístics*, 24, 211-228.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- Condamines, A. (1994). Terminologie et représentation des connaissances. *Didaskalia*, 6, 29-40.
- Condamines, A. (2000). Les bases théoriques du groupe toulousain « Sémantique et Corpus » : Ancrages et perspectives. *Cahiers de grammaire*, 25, 5-28.
- Condamines, A. (2002). Corpus Analysis and Conceptual Relation Patterns. *Terminology*, 8(1).
- Condamines, A. (2003). *Sémantique et Corpus Spécialisés : Constitution de bases de connaissances terminologiques*. Mémoire d'Habilitation à diriger des recherches. Toulouse: Université Toulouse le Mirail.
- Condamines, A. (2005a). Linguistique de corpus et terminologie. *Langages*, 157, 36-47.
- Condamines, A. (Éd.). (2005b). *Sémantique et corpus*. Paris: Hermès Science Publications.

- Condamines, A. (2017). Terminological Knowledge Bases: From texts to terms, from terms to texts. Dans P. A. Fuertes-Olivera (Éd.), *The Routledge Handbook of Lexicography* (pp. 335-350). London: Routledge.
- Condamines, A. (2018). Nouvelles perspectives pour la terminologie textuelle. Dans J. Altmanova, M. Centrella, & K. E. Russo (Éds), *Terminology & Discourse / Terminologie et discours* (pp. 93-112). Berne: Peter Lang.
- Condamines, A., & Dehaut, N. (2011). Mise en œuvre des méthodes de la linguistique de corpus pour étudier les termes en situation d'innovation disciplinaire : Le cas de l'exobiologie. *META*, 56(2), 266-283.
- Condamines, A., Dehaut, N., & Picton, A. (2012). Rôle du temps et de la pluridisciplinarité dans la néologie sémantique en contexte scientifique. Études outillées en corpus. *Cahiers de lexicologie*, 100(1), 161-184.
- Condamines, A., Humbert-Droz, J., & Picton, A. (à paraître). Néologie par déterminologisation : méthode de repérage et caractérisation en corpus dans le domaine de la physique des particules. Dans *La néologie des langues romanes : nouvelles dynamiques et enjeux*. Berne: Peter Lang.
- Condamines, A., & Picton, A. (2014a). Des communiqués de presse du Cnes à la presse généraliste. Vers un observatoire de la diffusion des termes. Dans P. Dury, J. Carlos de Hoyos, J. Makri-Morel, F. Maniez, V. Renner, & M. B. Villar Diaz (Éds), *La néologie en langue de spécialité : Détection, implantation et circulation des nouveaux termes* (pp. 165-188). Lyon: Centre de recherche en terminologie et traduction, Université Lumière Lyon 2.
- Condamines, A., & Picton, A. (2014b). Étude du fonctionnement des nominalisations déverbiales dans un contexte de désécialisation. Dans *4e Congrès Mondial de Linguistique française* (pp. 697-711). Berlin.
- Condamines, A., & Picton, A. (2015). Terminologie outillée : Analyse de corpus spécialisés dans différentes situations de néologie. Dans *Conférence-hommage à John Humbley « Quo vadis, Terminologia »*, Paris 7.
- Condamines, A., & Picton, A. (à paraître). Textual Terminology: Origins, Principles and New Challenges. Dans M.-C. L'Homme & P. Faber (Éds), *Theoretical Approaches to Terminology*. Amsterdam, Philadelphie: John Benjamins.
- Condamines, A., & Rebeyrolle, J. (1997). Point de vue en langue spécialisée. *META*, 42(1), 174-184.
- Condamines, A., Rebeyrolle, J., & Soubeille, A. (2004). Variation de la terminologie dans le temps : Une méthode linguistique pour mesurer l'évolution de la connaissance en corpus. Dans *Actes d'Euralex' 2004* (pp. 547-557).
- Cori, M., & David, S. (2008). Les corpus fondent-ils une nouvelle linguistique ? *Langages*, 171, 111-129.
- Cormier, M. C., & Francoeur, A. (Éds). (1997). *Problèmes et méthodes de la lexicographie québécoise*. Québec: CIRAL.

- Cortès, C. (2004). Terminologie et syntaxe de la classification. *Les Cahiers du CIEL. Des fondements théoriques de la terminologie*, 135-161.
- Costes, M., & Laval, S. (2019). Le vocable de la schizophrénie au prisme de la presse quotidienne nationale française : Un effet de mode ? *Les Cahiers du journalisme—Recherches*, 2(3), 29-44.
- Cottez, H. (1994). La place des termes techniques et scientifiques dans un dictionnaire de langue. Dans D. Candel (Éd.), *Français scientifique et technique et dictionnaire de langue* (pp. 13-26). Paris: Didier Érudition.
- Culpeper, J., & Demmen, J. (2015). Keywords. Dans D. Biber & R. Reppen (Éds), *The Cambridge Handbook of English Corpus Linguistics* (pp. 90-105). Cambridge: Cambridge University Press.
- Daille, B. (2005). Variations and Application-Oriented Terminology Engineering. *Terminology*, 11(1), 181-197.
- Daille, B. (2017). *Term Variation in Specialised Corpora*. Amsterdam, New York: John Benjamins.
- Damerau, F. J. (1993). Generating and Evaluating Domain-oriented Multi-word Terms from Texts. *Information Processing & Management*, 29(4), 433-447.
- Deignan, A. (2005). *Metaphor and corpus linguistics*. Amsterdam: John Benjamins.
- Del Tredici, M., Fernández, R., & Boleda, G. (2019). Short-Term Meaning Shift: A Distributional Exploration. *Proceedings of NAACL-HLT 2019 (Annual Conference of the North American Chapter of the Association for Computational Linguistics)*, 2069-2075.
- Delavigne, V. (2001). *Les mots du nucléaire. Contribution socioterminologique à une analyse des discours de vulgarisation*. Thèse de doctorat en Linguistique. Université de Rouen.
- Delavigne, V. (2003). Quand le terme entre en vulgarisation. Dans *Terminologie et Intelligence artificielle, TLA 2003* (pp. 80-91). Strasbourg.
- Delavigne, V. (2006). La formation du vocabulaire de la physique nucléaire : Quelques jalons. Dans *Aspects diachroniques du vocabulaire* (pp. 89-107). Rouen: Publications des Universités de Rouen et du Havre.
- Delavigne, V. (2013). Quand le patient devient expert : Usages des termes dans les forums médicaux. Dans *Terminologie et Intelligence artificielle, TLA 2013*. Paris, France.
- Delavigne, V. (2020). De l'(in)constance du métalinguistique dans un corpus de vulgarisation médicale. *Corela. Cognition, représentation, langage*, HS-31. <http://journals.openedition.org/corela/11031> [consulté le 15.10.2020]
- Delavigne, V., & Bouveret, M. (2000a). Présentation. Dans V. Delavigne & M. Bouveret (Éds), *Sémantique des termes spécialisés* (pp. 7-15). Rouen: Publications de l'Université de Rouen.
- Delavigne, V., & Bouveret, M. (Éds). (2000b). *Sémantique des termes spécialisés*. Rouen: Publications de l'Université de Rouen.
- Deleu, C. (2006). Les journalistes et la science : Représentation des pratiques professionnelles dans les journaux d'information des médias télévisuels généralistes. *Les Cahiers du journalisme*, 15, 66-85.

- Denis, P., & Sagot, B. (2010). Exploitation d'une ressource lexicale pour la construction d'un étiqueteur morphosyntaxique état-de-l'art du français. Dans *Traitement Automatique des Langues Naturelles (TALN 2010)*. Montréal.
- Depecker, L. (1994). L'inscription des termes officiels dans les dictionnaires de langue. *Terminologies Nouvelles*, 12, 151-155.
- Depecker, L. (2002). *Entre signe et concept : Éléments de terminologie générale*. Paris: Presses Sorbonne nouvelle. <https://books.openedition.org/psn/3388> [consulté le 22.03.2019]
- Desagulier, G. (2017). *Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics*. Cham: Springer.
- Desagulier, G. (2019). Can Word Vectors Help Corpus Linguists. *Studia Neophilologica*, 91(2), 219-240.
- Díaz Hormigo, M. T. (2020). Precisiones para una caracterización lingüística de la neología semántica. *ELUA*, 34, 73-94.
- Drouin, P. (2003). Term Extraction Using Non-Technical Corpora as a Point of Leverage. *Terminology*, 9(1), 99-117.
- Drouin, P. (2007). Identification automatique du lexique scientifique transdisciplinaire. *Revue française de linguistique appliquée*, 12(2), 45-64.
- Drouin, P. (2010). From a Bilingual Transdisciplinary Scientific Lexicon to Bilingual Transdisciplinary Scientific Collocations. Dans *Proceedings of the 14th EURALEX International Congress* (pp. 296-305). Fryske Akademy.
- Drouin, P. (2015). Acquisition automatique de termes : Simuler le travail du terminologue. *Ela. Études de linguistique appliquée*, 180(4), 417-427.
- Drouin, P. (2018). Étude du profil lexical d'un corpus de l'environnement. Dans *11es journées du réseau Lexicologie, Terminologie, Traduction*. Grenoble.
- Drouin, P., Francoeur, A., Humbley, J., & Picton, A. (Éds). (2017). *Multiple Perspectives on Terminological Variation*. Amsterdam, New York: John Benjamins.
- Drouin, P., & Langlais, P. (2006). Évaluation du potentiel terminologique de candidats termes. Dans *JADT 2006 : 8es Journées internationales d'Analyse statistique des Données Textuelles* (pp. 389-400). Besançon, France.
- Drouin, P., L'Homme, M.-C., & Robichaud, B. (2018). Lexical Profiling of Environmental Corpora. Dans *11th edition of the Language Resources and Evaluation Conference (LREC 2018)* (pp. 3419-3425). Miyazaki, Japon.
- Dubois, J. (1966). Les problèmes du vocabulaire technique. *Cahiers de lexicologie*, 9(2), 103-112.
- Dubois, J., Giacomo, M., Guespin, L., Marcellesi, C., Marcellesi, J.-B., & Mével, J.-P. (2002). *Dictionnaire de linguistique*. Paris: Larousse.
- Durieux, C. (1996-1997). Pseudo-synonymes en langue de spécialité. *Les Cahiers du CIEL*, 89-114.

- Dury, P. (1997). *Étude comparative et diachronique de l'évolution de dix dénominations fondamentales du domaine de l'écologie en anglais et en français*. Thèse de doctorat en Linguistique, Lexicologie et Terminologie multilingues. Louis-Lumière, Lyon 2.
- Dury, P. (2000). Les variations sémantiques en terminologie : Étude diachronique et comparative appliquée à l'écologie. Dans V. Delavigne & M. Bouveret (Éds), *Sémantique des termes spécialisés* (pp. 17-32). Rouen: Publications de l'Université de Rouen.
- Dury, P. (2005). Terminology and Specialized Translation: The Relevance of the Diachronic Approach. *LSP & Professional Communication*, 5(1), 31-41.
- Dury, P. (2008). The Rise of Carbon Neutral and Compensation Carbone. A Diachronic Investigation into the Migration of Vocabulary from the Language of Ecology to Newspaper Language and Vice Versa. *Terminology*, 14(2), 230-248.
- Dury, P., de Hoyos, J.C., Makri-Morel, J., Maniez, F., Renner, V., & Villar Diaz, M. B. (Éds). (2014). *La néologie en langue de spécialité: Détection, implantation et circulation des nouveaux termes*. Lyon: Centre de recherche en terminologie et traduction, Université Lumière Lyon 2.
- Dury, P., & Lervad, S. (2008). La variation synonymique dans la terminologie de l'énergie : Approches synchronique et diachronique, deux études de cas. *LSP & Professional Communication*, 8(2), 66-79.
- Dury, P., Maniez, F., Arlin, N., & Rougemont, C. (Éds). (2009). *La métaphore en langues de spécialité*. Grenoble: Les Presses Universitaires de Grenoble.
- Egbert, J., & Biber, D. (2019). Incorporating Text Dispersion into Keyword Analyses. *Corpora*, 14(1), 77-104.
- Egbert, J., Larsson, T., & Biber, D. (2020). *Doing Linguistics with a Corpus: Methodological Considerations for the Everyday User*. Cambridge: Cambridge University Press.
- Engwall, G. (1994). Not Chance but Choice: Criteria in Corpus Creation. Dans B. T. S. Atkins & A. Zampolli (Éds), *Computational Approaches to the Lexicon* (pp. 49-82). Oxford: Oxford University Press.
- Estopà, R. (2016). La neología especializada: Términos médicos en la prensa española. Dans C. Sánchez Manzanares & D. Azorín Fernández (Éds), *Estudios de Neología del Español* (pp. 109-129). Universidad de Murcia: Servicio de Publicaciones.
- Fabre, C. (2015). Sémantique distributionnelle automatique : La proximité distributionnelle comme mode d'accès au sens. *Ela. Etudes de linguistique appliquée*, 180(4), 395-405.
- Fabre, C., & Lenci, A. (2015). Distributional Semantics Today: Introduction to the special issue. *Traitement Automatique des Langues*, 56(2), 7-20.
- Falk, I., Bernhard, D., Gérard, C., & Potier-Ferry, R. (2014). Étiquetage morpho-syntaxique pour des mots nouveaux. Dans *Traitement Automatique des Langues Naturelles (TALN 2014)*. Marseille.
- Fernández-Silva, S. (2013). Punto de vista y variación denominativa. *Debate terminológico*, 9, 11-37.

- Fernández-Silva, S. (2016). The Cognitive and Rhetorical Role of Term Variation and its Contribution to Knowledge Construction in Research Articles. *Terminology*, 22(1), 52-79.
- Fernández-Silva, S. (2018). The Cognitive and Communicative Functions of Term Variation in Research Articles: A Comparative Study in Psychology and Geology. *Applied Linguistics*, 40(4), 624-645.
- Fernández-Silva, S., Freixa, J., & Cabré, M. T. (2011). A Proposed Method for Analysing the Dynamics of Cognition through Term Variation. *Terminology*, 17(1), 49-74.
- Firth, J. R. (1957). *Studies in Linguistics Analysis*. Oxford: Blackwell.
- Fišer, D., & Ljubešić, N. (2019). Distributional Modelling for Semantic Shift Detection. *International Journal of Lexicography*, 32(2), 163-183.
- Fleiss, J. L. (1972). Measuring Nominal Scale Agreement among Many Raters. *Psychological Bulletin*, 76(5), 378.
- Förner, W., & Thörle, B. (Éds). (2016). *Manuel des langues de spécialité*. Berlin: De Gruyter.
- Freixa, J. (2006). Causes of Denominative Variation in Terminology. A Typology Proposal. *Terminology*, 12(1), 51-77.
- Freixa, J., & Fernández-Silva, S. (2017). Terminological Variation and the Unsaturation of Concepts. Dans P. Drouin, A. Francoeur, J. Humbley, & A. Picton (Éds), *Multiple Perspectives on Terminological Variation* (pp. 155-180). Amsterdam, New York: John Benjamins.
- Fuertes-Olivera, Pedro A. (Éd.). (2017). *The Routledge Handbook of Lexicography*. London: Routledge.
- Gabrielatos, C. (2018). Keyness Analysis. Nature, metrics and techniques. Dans C. Taylor & A. Marchi (Éds), *Corpus Approaches to Discourse: A Critical Review* (pp. 225-258). London: Routledge.
- Galisson, R. (1978). *Recherches de lexicologie descriptive : La banalisation lexicale. Le Vocabulaire du football dans la presse sportive. Contribution aux recherches sur les langues techniques*. Paris: Nathan.
- Galisson, R. (1979). *Lexicologie et enseignement des langues*. Paris: Hachette.
- Galy, E., & Bourigault, D. (2005). Analyse distributionnelle de corpus de langue générale et synonymie. Dans *4èmes Journées de la linguistique de corpus*, Lorient.
- Gambier, Y. (1991a). Présupposés de la terminologie : Vers une remise en cause. *Cahiers de linguistique sociale, Terminologie et sociolinguistique*, 18, 31-58.
- Gambier, Y. (1991b). Travail et vocabulaire spécialisés : Prolégomènes à une socio-terminologie. *META*, 36(1), 8-15.
- Gamer, M., Lemon, J., & Fellows, I. (2012). irr: Various coefficients of interrater reliability and agreement (version 0.84.1). <https://cran.r-project.org/web/packages/irr> [consulté le 25.05.2020]

- García Palacios, J., de Sterck, G., Linder, D., Maroto, N., Sánchez Ibañez, M., & Torres del Rey, J. (Éds). (2016). *La neología en las lenguas románicas: Recursos, estrategias y nuevas orientaciones*. Berne: Peter Lang.
- Gardin, B., Lefèvre, G., Tardy, M., & Mortureux, M.-F. (1974). À propos du « sentiment néologique ». *Langages*, 36, 45-52.
- Gaudin, F. (1990). De l'interaction à la terminologie : Le travail scientifique. *Cahiers de linguistique sociale*, 16, 149-160.
- Gaudin, F. (1991). Terminologie et travail scientifique : Mouvement des signes, mouvement des connaissances. *Cahiers de linguistique sociale, Terminologie et sociolinguistique*, 18, 101-131.
- Gaudin, F. (1993). *Pour une socioterminologie : Des problèmes sémantiques aux pratiques institutionnelles*. Rouen: Publications de l'Université de Rouen.
- Gaudin, F. (1997). De la lecture des textes à l'appropriation des savoirs. *Le langage et l'homme*, 34(4), 57-73.
- Gaudin, F. (2003). *Socioterminologie : Une approche sociolinguistique de la terminologie*. Bruxelles: De Boeck-Duculot.
- Gaudin, F. (2005). La socioterminologie. *Langages*, 157(1), 81-93.
- Gaudin, F., & Guespin, L. (2000). *Initiation à la lexicologie française : De la néologie aux dictionnaires*. Bruxelles: Duculot.
- Gentilhomme, Y. (1994a). L'éclatement du signifié dans les discours technoscientifiques. *Cahiers de lexicologie*, 64(1), 5-35.
- Gentilhomme, Y. (1994b). Regards sur la terminologisation en lexicologie. *META*, 39(4), 546-560.
- Gérard, C. (2018). Variabilité du langage et productivité lexicale. Problèmes et propositions méthodologiques. *Neologica*, 12, 23-45.
- Gérard, C., Bruneau, L., Falk, I., Bernhard, D., & Rosio, A.-L. (2016). Le Logoscope : Observatoire des innovations lexicales en français contemporain. Dans J. García Palacios, G. de Sterck, D. Linder, N. Maroto, M. Sánchez Ibañez, & J. Torres del Rey (Éds), *La neología en las lenguas románicas: Recursos, estrategias y nuevas orientaciones*. Berne: Peter Lang.
- Gerbig, A., & Mason, O. (Éds). (2008). *Language, People, Numbers: Corpus Linguistics and Society*. Amsterdam: Rodopi.
- Gilbert, P. (1973). Remarques sur la diffusion des mots scientifiques et techniques dans le lexique commun. *Langue Française*, 17, 31-43.
- Girardi, B. (2006). Lorsque la science entre dans les gratuits sans y être vraiment invitée... *Les Cahiers du journalisme*, 15, 112-125.
- Gonon, F., Dumas-Mallet, E., & Ponnou, S. (2019). La couverture médiatique des observations scientifiques concernant les troubles mentaux. *Les Cahiers du journalisme—Recherches*, 2(3), 45-63.

- Gouadec, D. (1990). *Terminologie : Constitution des données*. Paris: AFNOR.
- Grefenstette, G. (1996). Evaluation Techniques for Automatic Semantic Extraction: Comparing Syntactic and Window Based Approaches. Dans B. Boguraev & J. Pustejovsky (Éds), *Corpus Processing for Lexical Acquisition* (pp. 205-216). Cambridge, Massachusetts: The MIT Press.
- Gries, S. Th. (2006). Exploring Variability within and between Corpora: Some methodological considerations. *Corpora*, 1(2), 109-151.
- Gries, S. Th. (2015). Quantitative Designs and Statistical Techniques. Dans D. Biber & R. Reppen (Éds), *The Cambridge Handbook of English Corpus Linguistics* (pp. 50-72). Cambridge: Cambridge University Press.
- Gries, S. Th. (2016). *Quantitative Corpus Linguistics with R: A Practical Introduction*. New York, London: Routledge.
- Gries, S. Th., & Hilpert, M. (2008). The Identification of Stages in Diachronic Data: Variability-based Neighbour Clustering. *Corpora*, 3(1), 59-81.
- Grishman, R., & Kittredge, R. (Éds). (1986). *Analyzing Language in Restricted Domains: Sublanguage description and processing*. London: Lawrence Erlbaum.
- Grothendieck, G. (2018). gsubfn: Utilities for Strings and Function Arguments (version 0.7). <https://cran.r-project.org/package=gsubfn> [consulté le 12.05.2020]
- Guespin, L. (1985). Matériaux pour une glottopolitique. *Cahiers de linguistique sociale*, 7, 13-32.
- Guespin, L. (1991). La circulation terminologique et les rapports entre science, technique et production. *Cahiers de linguistique sociale, Terminologie et sociolinguistique*, 18, 59-79.
- Guilbert, L. (1965). *La formation du vocabulaire de l'aviation*. Thèse de doctorat en Lettres. Faculté des Lettres et Sciences Humaines, Université de Paris, Paris: Larousse.
- Guilbert, L. (1971). La néologie scientifique et technique. *La Banque des Mots*, 1, 45-54.
- Guilbert, L. (1973a). La spécificité du terme scientifique et technique. *Langue Française*, 17, 5-17.
- Guilbert, L. (1973b). Théorie du néologisme. *Cahiers de l'Association internationale des études françaises*, 25, 9-29.
- Guilbert, L. (1975). *La créativité lexicale*. Paris: Larousse.
- Guillot, C., Heiden, S., & Prévost, S. (Éds). (2006). *À la quête du sens : Études littéraires, historiques et linguistiques en hommage à Christiane Marchello-Nizija*. Lyon: ENS Éditions.
- Habert, B. (2000). Des corpus représentatifs : De quoi, pour quoi, comment ? Dans M. Bilger (Éd.), *Linguistique sur corpus. Études et réflexions* (pp. 11-58). Perpignan: Les Presses de l'Université de Perpignan.
- Habert, B. (2004). Outiller la linguistique : De l'emprunt de techniques aux rencontres de savoirs. *Revue française de linguistique appliquée*, 9(1), 5-24.

- Habert, B. (2006). Portrait de linguiste(s) à l'instrument. Dans C. Guillot, S. Heiden, & S. Prévost (Éds), *À la quête du sens : Études littéraires, historiques et linguistiques en hommage à Christiane Marchello-Nizia* (pp. 163-173). Lyon: ENS Éditions.
- Habert, B., & Fabre, C. (1998). Relations de dépendance syntaxique et sémantique distributionnelle. Dans S. Mellet & M. Vuillaume (Éds), *Mots chiffrés et déchiffrés : Mélanges offerts à Etienne Brunet* (pp. 113-123). Paris: Honoré Champion.
- Habert, B., Fabre, C., & Issac, F. (1998). *De l'écrit au numérique : Constituer, normaliser, exploiter les corpus électroniques*. Paris: InterEditions/Masson.
- Habert, B., Illouz, G., & Folch, H. (2005). Des décalages de distribution aux divergences d'acception. Dans A. Condamines (Éd.), *Sémantique et corpus* (pp. 277-318). Paris: Hermès Science Publications.
- Habert, B., Nazarenko, A., & Salem, A. (1997). *Les linguistiques de corpus*. Paris: Armand Colin/Masson.
- Habert, B., & Zweigenbaum, P. (2002). Contextual Acquisition of Information Categories. What has been done and what can be done automatically? Dans B. E. Nevin & S. B. Johnson (Éds), *The Legacy of Zellig Harris: Language and information into the 21st century* (Vol. 2: Mathematics and computability of language, pp. 203-231). Amsterdam: John Benjamins.
- Habert, B., & Zweigenbaum, P. (2003). Classer les mots : Sémantique à gros grain et méthodologie harrissienne. *Revue de Sémantique et Pragmatique*, 12, 101-119.
- Hacken, P. T., & Panocová, R. (Éds). (2015). *Word Formation and Transparency in Medical English*. Cambridge: Cambridge Scholars Publishing.
- Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23-34.
- Halskov, J. (2005). Probing the Properties of Determinologization: The DiaSketch. *Lambda*, 29, 39-63.
- Harris, Z. S. (1954). Distributional Structure. *Word*, 10(2-3), 146-162.
- Harris, Z. S. (1982). Discourse and Sublanguage. Dans R. Kittredge & J. Lehrberger (Éds), *Sublanguage: Studies of language in restricted semantic domains* (pp. 230-236). Berlin, New York: W. de Gruyter.
- Harris, Z. S. (1991). *A Theory of Language and Information: A Mathematical Approach*. Oxford: Clarendon Press; Oxford University Press.
- Hasselgard, H., Oksefjell Ebeling, S., & Ebeling, J. (Éds). (2012). *Corpus Perspectives on Patterns of Lexis*. Amsterdam, Philadelphie: John Benjamins.
- Hazem, A., & Daille, B. (2015). Méthode semi-compositionnelle pour l'extraction de synonymes des termes complexes. *Traitement Automatique des Langues*, 56(2), 51-76.
- Heiden, S., Magué, J.-P., & Pincemin, B. (2010). Une plateforme logicielle open-source pour la textométrie – conception et développement. Dans S. Bolasco, I. Chiari, & L. Giuliano

- (Éds), *Proceedings of 10th International Conference on the Statistical Analysis of Textual Data—JADT 2010* (Vol. 2, pp. 1021-1032). Rome: Edizioni Universitarie di Lettere Economia Diritto.
- Hellrich, J. (2019). *Word Embeddings: Reliability & Semantic Change*. (S.l.): IOS Press.
- Heylen, K., & Bertels, A. (2016). Sémantique distributionnelle en linguistique de corpus. *Langages*, 201(1), 51-64.
- Horecký, J. (1956). *Základy slovenskej terminológie*. Bratislava: Vydavateľstvo Slovenskej akadémie vied.
- Humbert-Droz, J., Picton, A., & Condamines, A. (2019). How to Build a Corpus for a Tool-Based Approach to Terminologisation in the Field of Particle Physics. *Research in Corpus Linguistics*, 7, 1-17.
- Humbley, J. (2005). La traduction des métaphores dans les langues de spécialité : Le cas des virus informatiques. *Lynx*, 52, 49-62.
- Humbley, J. (Éd.). (2007). *Aspects de la recherche en langue de spécialité*. Paris: Université Paris VII.
- Humbley, J. (2018a). *La néologie terminologique*. Limoges: Lambert-Lucas.
- Humbley, J. (2018b). Socioterminology. Dans J. Humbley, G. Budin, & C. Laurén (Éds), *Languages for Special Purposes: An International Handbook* (pp. 469-488). Berlin, Boston: De Gruyter. (page: 469-488 container-title: Languages for Special Purposes).
- Humbley, J., Budin, G., & Laurén, C. (Éds). (2018). *Languages for Special Purposes: An International Handbook*. Berlin, Boston: De Gruyter.
- Hundt, M., Nesselhauf, N., & Biewer, C. (Éds). (2007). *Corpus Linguistics and the Web. Corpus Linguistics and the Web*. Amsterdam: Rodopi.
- Hunston, S. (2008). Collecting Strategies and Design Decisions. Dans A. Ludeling & M. Kytö (Éds), *Corpus Linguistics: An International Handbook* (pp. 154-168). Berlin, New York: Mouton de Gruyter.
- Incelli, E. (2018). Popularising the Higgs Boson: A Corpus-Assisted Approach to Reporting Scientific Discovery in Online Media. *Corpora*, 13(2), 169-203.
- Inkpen, D., Paribakht, T. S., Faez, F., & Amjadian, E. (2016). Term Evaluator: A Tool for Terminology Annotation and Evaluation. *International Journal of Computational Linguistics and Applications*, 7(2), 145-165.
- ISO (Organisation internationale de normalisation). (2019). ISO 1087:2019. Travail terminologique et science de la terminologie—Vocabulaire.
- Jaccard, P. (1901). Etude de la distribution florale dans une portion des Alpes et du Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37, 547-579.
- Jacobi, D. (1984). Auteurs et lecteurs de la recherche : Une illustration de la thèse de la continuité. *Bulletin des bibliothèques de France (BBF)*, 6, 484-491.
- Jacobi, D. (1986). *Diffusion et vulgarisation : Itinéraires du texte scientifique*. Paris: Les belles lettres.

- Jacobi, D. (1988). Le discours de vulgarisation scientifique : Problèmes sémiotiques et textuels. Dans D. Jacobi & B. Schiele (Éds), *Vulgariser la science : Le procès de l'ignorance* (pp. 87-117). Seyssel: Champ Vallon.
- Jacobi, D., & Schiele, B. (Éds). (1988). *Vulgariser la science : Le procès de l'ignorance*. Seyssel: Champ Vallon.
- Jacques, M.-P. (2005). Pourquoi une linguistique de corpus ? Dans *La linguistique de corpus* (pp. 21-30). Rennes: Presses universitaires de Rennes.
- Jacques, M.-P. (2006). L'emploi de termes réduits comme révélateur de la centralité dans le domaine. Dans *Mots, termes et contextes. Actes des septièmes Journées scientifiques du réseau de chercheurs « Lexicologie, Terminologie, Traduction » (LTT 2005)* (pp. 299-308). Paris: Éditions des archives contemporaines.
- Jacques, M.-P. (2016). Une linguistique outillée, pour quels objets ? *Histoire Épistémologie Langage*, 38(2), 87-99.
- Jacques, M.-P. (2017). *La dynamique du texte, corpus, outils, analyses*. Mémoire d'habilitation à diriger des recherches. Université Grenoble Alpes.
- Jedlička, A., Formánková, V., & Rejmánková, M. (1970). *Základy české stylistiky*. Prague: Státní pedagogické nakl.
- Jiménez-Crespo, M. A. (2017). Combining Corpus and Experimental Studies: Insights into the Reception of Translated Medical Texts. *The Journal of Specialized Translation*, 28, 2-22.
- Jiménez-Crespo, M. Á., & Tercedor Sánchez, M. (2017). Lexical Variation, Register and Explicitation in Medical Translation: A comparable corpus study of medical terminology in US websites translated into Spanish. *Translation and Interpreting Studies*, 12(3), 405-426.
- Josselin-Leray, A. (2005). *Place et rôle des terminologies dans les dictionnaires généraux unilingues et bilingues. Étude d'un domaine de spécialité: Volcanologie*. Thèse de doctorat en Lexicologie et Terminologie multilingues. Université Lumière, Lyon 2.
- Josselin-Leray, A. (2010). Affiner la description des termes dans les dictionnaires généraux : L'apport d'un corpus de vulgarisation. *Lexis*, (4), 65-104.
- Josselin-Leray, A., & Roberts, R. P. (2013). De la sélection des termes pour inclusion dans le dictionnaire général. État des lieux général et analyse critique de la terminologie informatique dans le New Oxford Dictionary of English (2000). Dans *Le terme scientifique et technique dans le dictionnaire général. Actes de la 7e édition des RIL (Rencontres Internationales de Lexicographie)* (pp. 85-120). Lyon: Université Lyon 2.
- Kageura, K. (1997). Multifaceted/Multidimensional Concept Systems. Dans S. E. Wright & G. Budin (Éds), *Handbook of terminology management. Volume 1: Basic aspects of terminology management* (pp. 119-132). Amsterdam: John Benjamins.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. Londres, New York: Longman.

- Kilgarriff, A. (1996). Comparing Word Frequencies across Corpora: Why Chi-square doesn't work, and an improved LOB-Brown comparison. Dans *Proceedings of ALLC-ACH Conference* (pp. 169-172).
- Kilgarriff, A. (2001). Comparing Corpora. *International Journal of Corpus Linguistics*, 6(1), 97-133.
- Kilgarriff, A. (2005). Language is never, ever, ever random. *Corpus Linguistics and Linguistic Theory*, 1(2), 263-275.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychlý, P. & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7-36.
- Kilgarriff, A., & Grefenstette, G. (2003). Introduction to the Special Issue on the Web as Corpus. *Computational Linguistics*, 29(3), 333-347.
- Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). The Sketch Engine. Dans *Proceedings of the 11th EURALEX International Congress* (pp. 105-116).
- Kim, M. (2017). *Variation terminologique en francophonie. Élaboration d'un modèle d'analyse des facteurs d'implantation terminologique*. Thèse de doctorat en Linguistique. Paris IV Sorbonne.
- Kittredge, R. (1982). Variation and Homogeneity of Sublanguages. Dans R. Kittredge & J. Lehrberger (Éds), *Sublanguage: Studies of language in restricted semantic domains* (pp. 106-137). Berlin, New York: W. de Gruyter.
- Kittredge, R., & Lehrberger, J. (1982). *Sublanguage: Studies of language in restricted semantic domains*. Berlin, New York: W. de Gruyter.
- Kocourek, R. (1991). *La langue française de la technique et de la science : Vers une linguistique de la langue savante*. Wiesbaden: O. Brandstetter Verlag.
- Kutuzov, A., Øvrelid, L., Szymanski, T., & Velldal, E. (2018). Diachronic Word Embeddings and Semantic Shifts: A survey. Dans *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 1384-1397). Santa Fe, États-Unis: Association for Computational Linguistics.
- Labasse, B. (2012). Sexe, sang et physique des particules : Le « sensationnalisme » est-il partout... Ou nulle part ? *Les Cahiers du journalisme*, 24, 114-149.
- Lafon, P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots. Les langages du politique*, 1, 127-165.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174.
- Landragin, F., Poibeau, T., & Victorri, B. (2016). Analyse des références et des transitions référentielles : L'apport de la linguistique outillée. Dans Laure Sarda, D. Vigier, & B. Combettes (Éds), *Connexion et indexation. Ces liens qui tissent le texte* (pp. 123-135). Lyon: ENS Éditions.
- Larivière, L. (1998). Valeur sémantique du verbe dans les collocations verbales spécialisées. *TTR : traduction, terminologie, rédaction*, 11(1), 173-197.

- Lázaro Gutiérrez, R., del Mar Sánchez Ramos, M., & Vigier Moreno, F. J. (Éds). (2015). *Investigación emergente en Traducción e Interpretación*. Albolote (Granada): Comares.
- Le Draoulec, A., Péry-Woodley, M.-P., & Rebeyrolle, J. (2014). Glissements progressifs de « sémantique ». *Le Discours et la Langue*, 1(6), 109-126.
- Ledouble, H. (2019). Vulgarisation scientifique et médiatisation de la science. Instabilité terminologique dans le domaine de la lutte biologique. *Terminology*, 25(1), 60-92.
- Ledouble, H. (2020). Term Circulation and Conceptual Instability in the Mediation of Science: Binary framing of the notions of biological versus chemical pesticides. *Discourse & Communication*, 14(5), 466-488.
- Leech, G. (1991). The State of the Art in Corpus Linguistics. Dans K. Aijmer & B. Altenberg (Éds), *English Corpus Linguistics* (pp. 8-29). London: Routledge.
- Leech, G. (2007). New Resources, or Just Better Old Ones ? The Holy Grail of Representativeness. Dans M. Hundt, N. Nesselhauf, & C. Biewer (Éds), *Corpus Linguistics and the Web* (pp. 133-149). Amsterdam: Rodopi.
- Lehrberger, J. (1982). Automatic Translation and the Concept of Sublanguage. Dans R. Kittredge & J. Lehrberger (Éds), *Sublanguage: Studies of language in restricted semantic domains* (pp. 81-106). Berlin, New York: W. de Gruyter.
- Lenci, A. (2008). Distributional Semantics in Linguistic and Cognitive Research. *Rivista di Linguistica*, 20(1), 1-31.
- Lenci, A. (2018). Distributional Models of Word Meaning. *Annual Review of Linguistics*, 4, 151-171.
- Léon, J. (2018). TAL et linguistique : Application, expérimentation, instrumentalisation. *Ela. Etudes de linguistique appliquée*, 190(2), 195-203.
- León-Araúz, P. (2015). Term Variation in the Psychiatric Domain: Transparency and Multidimensionality. Dans P. Hacken & R. Panocová (Éds), *Word Formation and Transparency in Medical English* (pp. 33-54). Newcastle-upon-Tyne: Cambridge Scholars Publishing.
- León-Araúz, P. (2017). Term and Concept Variation in Specialized Knowledge Dynamics. Dans P. Drouin, A. Francoeur, A. Picton, & J. Humbley (Éds), *Multiple Perspectives on Terminological Variation* (pp. 213-258). Amsterdam, New York: John Benjamins.
- León-Araúz, P., & Reimerink, A. (2014). From Term Dynamics to Concept Dynamics: Term Variation and Multidimensionality in the Psychiatric Domain. Dans C. V. Andrea Abel & N. Ralli (Éds), *Proceedings of the XVI EURALEX International Congress* (pp. 657-667). Bolzano.
- Lerat, P. (1995). *Les langues spécialisées*. Paris: Les Presses Universitaires de France.
- Levshina, N. (2015). *How to Do Linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.
- L'Homme, M.-C. (1998). Caractérisation des combinaisons lexicales spécialisées par rapport aux collocations de langue générale. Dans *Proceedings of EURALEX 1998* (pp. 513-522).

- L'Homme, M.-C. (2004). *La terminologie : Principes et techniques*. Montréal: Les Presses de l'Université de Montréal. <https://books.openedition.org/pum/10693> [consulté le 12.05.2019]
- L'Homme, M.-C. (2017). Combinatoire spécialisée : Trois perspectives et des enseignements pour la terminologie. *TTR : traduction, terminologie, rédaction*, 30(1-2), 215-241.
- L'Homme, M.-C. (2020). *Lexical Semantics for Terminology*. Amsterdam, Philadelphie: John Benjamins.
- L'Homme, M.-C., & Polguère, A. (2007). Mettre en bons termes les dictionnaires spécialisés et les dictionnaires de langue générale. Dans *Colloque en l'honneur d'Henri Béjoint*. Lyon.
- Light, R. J. (1971). Measure of Response Agreement for Qualitative Data: Some generalizations and alternatives. *Psychological Bulletin*, 76(5), 365-377.
- Loffler-Laurian, A.-M. (1983). Typologie des discours scientifiques : Deux approches. *Étude de Linguistique appliquée*, 51, 8-20.
- Loffler-Laurian, A.-M. (1984). Vulgarisation scientifique : Formulation, reformulation, traduction. *Langue française*, 64, 109-125.
- Ludeling, A., & Kytö, M. (Éds). (2008). *Corpus Linguistics: An International Handbook*. Berlin, New York: Mouton de Gruyter.
- Maniez, F. (2002). Distinguer les termes des collocations : Étude sur corpus du patron <Adjectif—Nom> en anglais médical. Dans *Actes de la conférence Traitement automatique de la langue naturelle—TALN (2002)* (pp. 345-350). Nancy.
- Marchello-Nizia, C. (2004). Linguistique historique, linguistique outillée : Les fruits d'une tradition. *Le français moderne*, 72(1), 58-70.
- Marín Pérez, M. J. (2014). Evaluation of Five Single-word Term Recognition Methods on a Legal English Corpus. *Corpora*, 9(1), 83-107.
- Marín Pérez, M. J. (2016). Measuring the Degree of Specialisation of Sub-Technical Legal Terms through Corpus Comparison. *Terminology*, 22(1), 80-102.
- Martin-Lagardette, J.-L. (2016). *Le guide de l'écriture journalistique* (7e édition). Paris: La Découverte.
- Masár, I. (1991). *Průručka slovenskej terminológie*. Bratislava: Veda.
- Mayaffre, D., Pincemin, B., & Poudat, C. (2019). Explorer, mesurer, contextualiser. Quelques apports de la textométrie à l'analyse de discours. *Langue française*, 203(3), 101-115.
- Mazière, F. (1981). Le dictionnaire et les termes. *Cahiers de lexicologie*, 39(2), 79-104.
- McEneaney, T. (2016). Keywords. Dans P. Baker & J. Egbert (Éds), *Triangulating Methodological Approaches in Corpus Linguistic Research* (pp. 20-32). New York, London: Routledge.
- McEneaney, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge, New York: Cambridge University Press.
- Meillet, A. (1905). Comment les mots changent de sens. *L'année sociologique*, 9, 1-38.

- Mellet, S., & Vuillaume, M. (Éds). (1998). *Mots chiffrés et déchiffrés : Mélanges offerts à Etienne Brunet*. Paris: Honoré Champion.
- Meyer, I. (2000). Computer Words in Our Everyday Lives: How are they interesting for terminography and lexicography? Dans *Proceedings of EURALEX 2000* (pp. 39-58). Stuttgart, Allemagne.
- Meyer, I. (2001). Extracting Knowledge-Rich Contexts for Terminography: A conceptual and methodological framework. Dans D. Bourigault, C. Jacquemin, & M.-C. L'Homme (Éds), *Recent Advances in Computational Terminology* (pp. 279-302). Amsterdam: John Benjamins.
- Meyer, I., & Mackintosh, K. (1996). Refining the Terminographer's Concept-Analysis Methods: How Can Phraseology Help? *Terminology*, 3(1), 1-26.
- Meyer, I., & Mackintosh, K. (2000a). « L'étirement » du sens terminologique : Aperçu du phénomène de la déterminologisation. Dans H. Béjoint & P. Thoiron (Éds), *Le sens en terminologie* (pp. 198-217). Lyon: Les Presses Universitaires de Lyon.
- Meyer, I., & Mackintosh, K. (2000b). When Terms move into Our Everyday Lives: An Overview of De-terminologization. *Terminology*, 6(1), 111-138.
- Meyer, I., Mackintosh, K., & Varantola, K. (1998). From Virtual Sex to Virtual Dictionaries: On the Analysis and Description of a De-terminologized Word. Dans *Proceedings of EURALEX 1998* (Vol. II, pp. 645-654). Liège, Belgique.
- Meyer, I., Zaluski, V., & Mackintosh, K. (1997). Metaphorical Internet Terms: A Conceptual and Structural Analysis. *Terminology*, 4(1), 1-33.
- Milizia, D. (2010). Keywords and Phrases in Political Speeches. Dans M. Bondi & M. Scott (Éds), *Keyness in Texts* (pp. 127-145). Amsterdam, Philadelphie: John Benjamins.
- Moirand, S. (1993). Autour de la notion de didacticité. *Les Carnets du Cediscor*, 1, 9-20.
- Moirand, S. (1997). Formes discursives de la diffusion des savoirs dans les médias. *Hermès*, 21(1), 33-44.
- Moirand, S. (1999). Éléments de théorisation d'une linguistique du discours : L'exemple du discours sur les sciences dans les médias. *Modèles linguistiques*, 40, 5-21. <https://journals.openedition.org/ml/1400> [consulté le 27.09.2020]
- Moirand, S. (2007). *Les discours de la presse quotidienne. Observer, analyser, comprendre*. Paris: Presses universitaires de France, Linguistique nouvelle.
- Moirand, S. (2011). Le dialogisme : De la réception du concept à son appropriation en analyse du discours. *Cahiers de praxématique*, (57), 69-100.
- Moirand, S. (2014). Vers de nouvelles configurations discursives. *Les Carnets du Cediscor*, (12), 141-149.
- Moirand, S., Reboul-Touré, S., & Ribeiro, M. P. (2016). La vulgarisation scientifique au croisement de nouvelles sphères d'activité langagière. *Bakhtiniana. Revista de Estudos do Discurso*, 11(2), 145-169.

- Montalt-Resurrecció, V., & González Davies, M. (2014). *Medical Translation Step by Step: Learning by Drafting* (2e édition). Londres, New York: Routledge.
- Montalt-Resurrecció, V., & Shuttleworth, M. (2012). Research in translation and knowledge mediation in medical and healthcare settings. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, 11, 9-29.
- Morlane-Hondère, F. (2013). *Une approche linguistique de l'évaluation des ressources extraites par analyse distributionnelle automatique*. Thèse de doctorat en Linguistique. Université Toulouse le Mirail, Toulouse II.
- Mortureux, M.-F. (1985). Linguistique et vulgarisation scientifique. *Social Science Information*, 24(4), 825-845.
- Mortureux, M.-F. (1988). La vulgarisation scientifique : Parole médiane ou dédoublée. Dans D. Jacobi & B. Schiele (Éds), *Vulgariser la science : Le procès de l'ignorance* (pp. 118-147). Seyssel: Champ Vallon.
- Mortureux, M.-F. (1993). Paradigmes désignationnels. *Semen*, 8. <https://journals.openedition.org/semen/4132> [consulté le 17.06.2020]
- Mortureux, M.-F. (1995). Les vocabulaires scientifiques et techniques. *Les Carnets du Cediscor*, 3, 13-25. <http://journals.openedition.org/cediscor/463> [consulté le 8.10.2019]
- Mpouli, S. (2019). Chronique d'un échec : Identification des métaphores dans les écrits des géographes. *Traitement Automatique des Langues*, 60(3), 89-111.
- Nelson, M. (2010). Building a Written Corpus. Dans A. O'Keeffe & M. McCarthy (Éds), *The Routledge Handbook of Corpus Linguistics* (pp. 53-65). London: Routledge.
- Neveu, F. (2004). *Dictionnaire des sciences du langage*. Paris: Armand Colin.
- Nevin, B., & Johnson, S. B. (Éds). (2002). *The Legacy of Zellig Harri : Language and information into the 21st century. Vol. 2: Mathematics and computability of language*. Amsterdam: John Benjamins.
- Nicolae, C., & Delavigne, V. (2013). Naissance et circulation d'un terme : Une histoire d'exoplanètes. Dans G. Williams (Éd.), *Acte des sixièmes Journées de la Linguistique de Corpus* (pp. 217-229). Lorient: Université de Bretagne-Sud.
- Nová, J. (2018). Terms Embraced by the General Public: How to Cope with Determinologization in the Dictionary? Dans *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts* (pp. 387-398). Ljubljana, Slovénie.
- Nowakowska, A., & Sarale, J.-M. (2011). Le dialogisme : Histoire, méthodologie et perspectives d'une notion fortement heuristique. *Cahiers de praxématique*, (57), 9-20.
- Observatori de Neologia. (2016). *Mots d'anni, mots de demà*. Barcelona: IULA, Universitat Pompeu Fabra.
- O'Keeffe, Anne, & McCarthy, M. (Éds). (2010). *The Routledge Handbook of Corpus Linguistics*. London: Routledge.
- Oliveira, I. (2005). La métaphore terminologique sous un angle cognitif. *META*, 50(4).

- Oliveira, I. (2009). *Nature et fonctions de la métaphore en science. L'exemple de la cardiologie*. Paris: L'Harmattan.
- Ortego Antón, M. T. (2015). El conocimiento experto multilingüe y su tratamiento en los diccionarios generales bilingües de español-francés: El caso del léxico de la informática. Dans R. Lázaro Gutiérrez, M. del Mar Sánchez Ramos, & F. J. Vigier Moreno (Éds), *Investigación emergente en Traducción e Interpretación* (pp. 129-146). Albolote (Granada): Comares.
- Ortego Antón, M. T., & Fernández Nistal, P. (2015a). Aproximación a las unidades con significado en el campo de la informática adquirido por extensión metafórica en los diccionarios generales de inglés y español. *Revista de Lingüística y Lenguas Aplicadas*, 10, 44-54.
- Ortego Antón, M. T., & Fernández Nistal, P. (2015b). La nomenclatura especializada en los diccionarios generales bilingües. Un estudio de caso: La informática. *Debate terminológico*, 13, 18-33.
- Pearson, J. (1998). *Terms in Context*. Amsterdam, Philadelphie: John Benjamins.
- Pédauque, R. T. (Éd.). (2007). *La redocumentation du monde*. Toulouse: Cepadues Éditions.
- Peirsman, Y., Geeraerts, D., & Speelman, D. (2010). The automatic identification of lexical variation between language varieties. *Natural Language Engineering*, 16(4), 469-491.
- Peirsman, Y., Heylen, K., & Speelman, D. (2007). Finding Semantically Related Words in Dutch. Cooccurrences versus syntactic contexts. Dans *Proceedings of the CoSMO workshop*. Roskilde, Danemark.
- Pelletier, J. (2012). *La variation terminologique: Un modèle à trois composantes*. Thèse de doctorat en Langues, linguistique et traduction. Université Laval, Québec.
- Péraldi, S. (2011). *Indétermination terminologique et multidimensionnalité dans le domaine de la chimie organique. Analyse à partir d'un corpus spécialisé de langue anglaise*. Thèse de doctorat en Linguistique Théorique, Descriptive et Automatique. Université Paris Diderot, Paris 7.
- Péraldi, S. (2012). L'anglais de spécialité en chimie organique: Entre indétermination terminologique et multidimensionnalité. *ASp*, (62), 5-24. <https://journals.openedition.org/asp/3066> [consulté le 17.02.2021]
- Périnet, A., & Hamon, T. (2014). Analyse et proposition de paramètres distributionnels adaptés aux corpus de spécialité. Dans *12e Journées internationales d'Analyse statistique des Données Textuelles (JADT 2014)* (pp. 507-518). Paris.
- Périnet, A., & Hamon, T. (2015). Analyse distributionnelle appliquée aux textes de spécialité— Réduction de la dispersion des données par abstraction des contextes. *Traitement Automatique des Langues*, 56(2), 77-102.
- Péry-Woodley, M.-P. (1995). Quels corpus pour quels traitements automatiques? *Traitement Automatique des Langues*, 36(1-2), 213-232.
- Petit, G. (2001). L'introuvable identité du terme technique. *Revue française de linguistique appliquée*, 6(2), 63-79.

- Peynaud, C. (2014). Définition et explication dans la presse généraliste aux États-Unis : De la médiation du discours expert à la création d'un intertexte médiatique. *ASp*, (65), 21-43. <https://journals.openedition.org/asp/4186> [consulté le 5.05.2021]
- Peynaud, C. (2016). Fukushima and its treatment in the press. *Progress in Industrial Ecology, An International Journal*, 10(2/3), 209-228.
- Phal, A. (1971). *Vocabulaire général d'orientation scientifique (VGOS). Part du lexique commun dans l'expression scientifique*. Paris: Didier.
- Philip, G. (2010). Metaphorical Keyness in Specialised Corpora. Dans M. Bondi & M. Scott (Éds), *Keyness in Texts* (pp. 185-203). Amsterdam, Philadelphie: John Benjamins.
- Picton, A. (2009). *Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial*. Thèse de doctorat en Sciences du Langage. Université Toulouse 2.
- Picton, A. (2018). Terminologie outillée et diachronie : Éléments de réflexion autour d'une réconciliation. *ASp*, 74, 27-52.
- Picton, A., Condamines, A., & Humbert-Droz, J. (2021). Analyse diachronique du processus de déterminologisation. Une réflexion en diachronie courte en physique des particules. *Cahiers de lexicologie*, 118, 11-43.
- Picton, A., & Dury, P. (2017). Diastatic Variation in Language for Specific Purposes. Observations from the Analysis of two Corpora. Dans P. Drouin, A. Francoeur, J. Humbley, & A. Picton (Éds), *Multiple Perspectives in Terminological Variation* (pp. 57-79). Amsterdam, New York: John Benjamins.
- Pierrejean, B. (2020). *Qualitative Evaluation of Word Embeddings: Investigating the instability in neural-based models*. Thèse de doctorat en Sciences du Langage. Université Toulouse - Jean Jaurès.
- Pierrejean, B., & Tanguy, L. (2018). Étude de la reproductibilité des word embeddings : Repérage des zones stables et instables dans le lexique. Dans *Actes de la conférence TALN 2018* (Vol. Volume 1, pp. 33-46). Rennes.
- Poibeau, T. (2005). Parcours interprétatifs et terminologie. Dans *Terminologie et Intelligence artificielle, TLA 2005*. Rouen.
- Pojanapunya, P., & Watson Todd, R. (2018). Log-likelihood and Odds Ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, 14(1), 133-167.
- Pontrandolfo, G. (2016). La traducción a la vista en el ámbito médico-sanitario: Un caso didáctico. *MonTI*, 3, 367-398.
- Portelance, C. (1991). Fondements linguistiques de la terminologie. *META*, 36(1), 64-70.
- Poštolková, B. (1977). K vlivu odborné terminologie na národní jazyk. *Slovo a slovesnost*, 38, 112-120.
- Poštolková, B. (1984). *Odborná a běžná slovní zásoba současné češtiny*. Prague: Academia.
- Poudat, C., & Landragin, F. (2017). *Explorer un corpus textuel : Méthodes, pratiques, outils*. Louvain-la-Neuve: De Boeck Supérieur.

- Prévost, S. (2005). Constitution et exploitation d'un corpus de français médiéval: Enjeux, spécificités et apports. Dans A. Condamines (Éd.), *Sémantique et corpus* (pp. 147-176). Paris: Hermès Science Publications.
- Pruvost, J., & Sablayrolles, J.-F. (2019). *Les néologismes* (4e édition). Paris: Presses Universitaires de France et Humensis.
- R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/> [consulté le 3.07.2019]
- Rastier, F. (2009). *Sémantique interprétative*. Paris: Presses Universitaires de France.
- Rastier, F. (2011). *La mesure et le grain : Sémantique de corpus*. Paris: Honoré Champion.
- Rastier, F., & Valette, M. (2009). De la polysémie à la néosémie. *Le français moderne*, 77, 97-116.
- Rayson, P. (2008). From Key Words to Key Semantic Domains. *International Journal of Corpus Linguistics*, 13(4), 519-549.
- Rayson, P., Leech, G. N., & Hodges, M. (1997). Social Differentiation in the Use of English Vocabulary: Some Analyses of the Conversational Component of the British National Corpus. *International Journal of Corpus Linguistics*, 2(1), 133-152.
- Reboul-Touré, S. (2004). Écrire la vulgarisation scientifique aujourd'hui. Dans *Colloque Sciences, Médias et Société* (pp. 195-212). Lyon.
- Renouf, A. (2012). A Finer Definition of Neology in English: The life-cycle of a word. Dans H. Hasselgard, S. Oksefjell Ebeling, & J. Ebeling (Éds), *Corpus Perspectives on Patterns of Lexis*. Amsterdam, Philadelphie: John Benjamins.
- Renouf, A. (2014). Semantic Neology: The Challenges for Automatic Identification. *Neologica*, 8, 185-220.
- Renouf, A. (2017). Some Corpus-Based Observations on Terminologisation. *Neologica*, 11, 21-48.
- Resche, C. (2004). Approche « terminométrique » du cycle économique : Implications et prolongements. *META*, 49(2), 343-359.
- Reutenauer, C. (2012a). Nouveau sens et évolution des domaines d'emploi : Méthodologie pour l'acquisition lexicale. Dans *Congrès Mondial de Linguistique Française - CMLF 2012*, Lyon.
- Reutenauer, C. (2012b). *Vers un traitement automatique de la néosémie : Approche textuelle et statistique*. Thèse de doctorat en Sciences du Langage. Université de Lorraine.
- Roberts, R. P. (2004). Terms in General Dictionaries. Dans José-María Bravo (Éd.), *A New Spectrum of Translation Studies* (pp. 121-140). Valladolid: Universidad de Valladolid, Secretariado de Publicaciones e Intercambio Editorial.
- Roberts, R. P., & Josselin-Leray, A. (2005). Le traitement des termes dans les dictionnaires généraux. Dans H. Béjoint & F. Maniez (Éds), *De la mesure dans les termes* (pp. 324-348). Lyon: Les Presses Universitaires de Lyon.

- Rogers, M. (2004). Multidimensionality in Concepts Systems: A Bilingual Textual Perspective. *Terminology*, 10(2), 215-240.
- Rogers, M. (2018). Towards a Typology of Terminological Variation: A modest proposal for specialised translation. Dans J. Altmanova, M. Centrella, & K. E. Russo (Éds), *Terminology & Discourse / Terminologie et discours*. Berne: Peter Lang.
- Rondeau, G. (1981). *Introduction à la terminologie*. Montréal: Gaëtan Morin.
- Rossi, M. (2009). L'emploi de la métaphore comme ressource pour la néologie terminologique : Le cas du langage de la dégustation du vin. Dans P. Dury, F. Maniez, N. Arlin, & C. Rougemont (Éds), *La métaphore en langues de spécialité* (pp. 199-227). Grenoble: Les Presses Universitaires de Grenoble.
- Rossi, M. (2014). Émergence et traitement des métaphores terminologiques dans les dictionnaires. *Neologica*, 8, 45-62.
- Rossi, M. (2016). Pour une typologie des avatars métaphoriques dans les terminologies spécialisées. *Langue française*, 189(1), 87-102.
- Rouillard, C.-A. (2016). Usage spécialisé ou usage courant des désignations de la maladie mentale dans le discours des non-spécialistes : Le cas des termes autisme et hystérie. *ScriptUM : la revue du colloque VocUM*, (2), 21-40.
- RStudio Team. (2019). *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc. <http://www.rstudio.com> [consulté le 13.07.2019]
- Rudis, B., Kennedy, P., Reiner, P., Wilson, D., Adam, X., Google, IBM, Impallari Type, Barnett, J., Leeper, T., & Meys, J. (2020). hrbthemes: Additional Themes, Theme Components and Utilities for « ggplot2 » (version 0.8.0). <https://cran.r-project.org/package=hrbthemes> [7.06.2020]
- Sablayrolles, J.-F. (2000). *La néologie en français contemporain, examen du concept et analyse de productions néologiques récentes*. Paris: Champion.
- Sablayrolles, J.-F. (2010). Néologisme homonymique, néologisme polysémique et évolution de sens. Pour une restriction de la néologie sémantique. Dans I. M. Alves (Éd.), *Neologia e neologismos em diferentes perspectivas* (pp. 83-100). São Paulo: Paulistana.
- Sablayrolles, J.-F. (2011). Quelques remarques sur une typologie des néologismes : Amalgamation ou télescopage : Un processus aux productions variées (mots valises, détournements...) et un tableau hiérarchisé des matrices. Dans *Actes de CINEO II. Sao Paulo, 5-8 décembre 2011*.
- Sablayrolles, J.-F. (2012). Extraction automatique et types de néologismes : Une nécessaire clarification. *Cahiers de lexicologie*, 100, 37-53.
- Sablayrolles, J.-F. (2018). Les néologismes ne naissent pas dans les choux. Dans D. Bernhard, M. Boisseau, C. Gérard, T. Grass, & A. Todirascu (Éds), *La néologie en contexte. Cultures, situations, textes* (pp. 23-38). Limoges: Lambert-Lucas.
- Sager, J. (1997). Term Formation. Dans S. E. Wright & G. Budin (Éds), *Handbook of Terminology Management. Volume 1: Basic aspects of terminology management* (pp. 25-41). Amsterdam: John Benjamins.

- Sager, N. (1982). Syntactic Formatting of Science Information. Dans R. Kittredge & J. Lehrberger (Éds), *Sublanguage: Studies of language in restricted semantic domains* (pp. 9-26). Berlin, New York: W. de Gruyter.
- Sager, N. (1986). Sublanguage: Linguistic Phenomenon, Computational Tool. Dans R. Grishman & R. Kittredge (Éds), *Analyzing Language in Restricted Domains* (pp. 1-17). Londres: Lawrence Erlbaum.
- Sagi, E., Kaufmann, S., & Clark, B. (2009). Semantic Density Analysis: Comparing word meaning across time and phonetic space. Dans *Proceedings of the EACL GEMS Workshop* (pp. 104-111). Athènes.
- Sahlgren, M. (2008). The Distributional Hypothesis. *Italian Journal of Linguistics*, 20(1), 1-18.
- Sahlgren, M., & Lenci, A. (2016). The Effects of Data Size and Frequency Range on Distributional Semantic Models. Dans *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 975-980). Austin, Texas: Association for Computational Linguistics.
- Sajous, F., Josselin-Leray, A., & Hathout, N. (2018). The Complementarity of Crowdsourced Dictionaries and Professional Dictionaries viewed through the Filter of Neology. *Lexis*, (12).
- Sánchez Manzanares, C., & Azorín Fernández, D. (Éds). (2016). *Estudios de Neología del Español*. Universidad de Murcia: Servicio de Publicaciones.
- Sarda, L., Vigier, D., & Combettes, B. (Éds). (2016). *Connexion et indexation. Ces liens qui tissent le texte*. Lyon: ENS Éditions.
- Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. Dans *Proceedings of International Conference on New Methods in Language Processing*. Manchester, UK.
- Scott, M. (1997). PC Analysis of Key Words—And Key Key Words. *System*, 25(2), 233-245.
- Scott, M. (2010). Problems in Investigating Keyness, or Clearing the Undergrowth and Marking out Trails... Dans M. Bondi & M. Scott (Éds), *Keyness in Texts* (pp. 43-57). Amsterdam, Philadelphie: John Benjamins.
- Scott, M. (2016). *WordSmith Tools version 7, Stroud: Lexical Analysis Software*. <https://www.lexically.net/wordsmith/> [consulté le 22.04.2017]
- Scott, M., & Tribble, C. (2006). *Textual Patterns: Key words and corpus analysis in language education*. Amsterdam, Philadelphie: John Benjamins.
- Siepmann, D., Bürgel, C., & Diwersy, S. (2017). The Corpus de Référence du Français Contemporain (CRFC) as the First Genre-Diverse Mega-Corpus of French. *International Journal of Lexicography*, 30(1), 63-84.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (1994). Corpus Typology: A Framework for Classification. *EAGLES document*, 1-18.
- Sinclair, J. (2008). Borrowed Ideas. Dans A. Gerbig & O. Mason (Éds), *Language, People, Numbers: Corpus Linguistics and Society* (pp. 21-41). Amsterdam: Rodopi.

- Slodzian, M. (2000). L'émergence d'une terminologie textuelle et le retour du sens. Dans H. Béjoint & P. Thoiron (Éds), *Le sens en terminologie* (pp. 61-83). Lyon: Les Presses Universitaires de Lyon.
- Soldatova, L., Vanschoren, J., Papadopoulos, G., & Ceci, M. (Éds). (2018). *Discovery Science* (Vol. 11198). Cham: Springer.
- Stefanowitsch, A. (2006). Words and their Metaphors: A corpus-based approach. Dans A. Stefanowitsch & S. T. Gries (Éds), *Corpus-Based Approaches to Metaphor and Metonymy* (pp. 63-105). Berlin: Mouton de Gruyter.
- Stefanowitsch, A. (2020). *Corpus Linguistics: A guide to the methodology*. Berlin: Language Science Press.
- Stefanowitsch, A., & Gries, S. T. (Éds). (2006). *Corpus-Based Approaches to Metaphor and Metonymy*. Berlin: Mouton de Gruyter.
- Stengers, I. (1987). *D'une science à l'autre : Des concepts nomades*. Paris: Éditions du Seuil.
- Taillet, R., Villain, L., & Febvre, P. (2018). *Dictionnaire de physique* (4ème édition). Louvain-la-Neuve: De Boeck Supérieur.
- Tanguy, L., Brunet, P., & Ferret, O. (2019). Comparaison qualitative et extrinsèque d'analyseurs syntaxiques du français : Confrontation de modèles distributionnels sur un corpus spécialisé. Dans *Actes de la Conférence sur le Traitement Automatique des Langues Naturelles (TALN)* (pp. 39-53). Toulouse.
- Tanguy, L., & Fabre, C. (2014). Évolutions de la linguistique outillée : Méfaits et bienfaits du TAL. *L'information grammaticale*, 142, 15-23.
- Tanguy, L., Fabre, C., Ho-Dac, L.-M., & Rebeyrolle, J. (2011). Caractérisation des échanges entre patients et médecins : Approche outillée d'un corpus de consultations médicales. *Corpus*, 10, 137-154.
- Tanguy, L., Sajous, F., & Hathout, N. (2015). Évaluation sur mesure de modèles distributionnels sur un corpus spécialisé : Comparaison des approches par contextes syntaxiques et par fenêtres graphiques. *Traitement Automatique des Langues*, 56(2), 103-127.
- Taylor, C. (2008). What is corpus linguistics ? What the data says. *Icame Journal*, (32), 179-200.
- Taylor, C., & Marchi, A. (Éds). (2018). *Corpus Approaches to Discourse: A Critical Review*. London: Routledge.
- Temmerman, R. (2000). *Towards New Ways of Terminology Description. The sociocognitive approach*. Amsterdam, Philadelphie: John Benjamins.
- Tercedor, M. (2011). The Cognitive Dynamics of Terminological Variation. *Terminology*, 17(2), 181-197.
- Tetet, C. (1994). La terminologie de l'alpinisme dans les dictionnaires. *META*, 39(4), 651-661.
- Thoiron, P. (1991). Texte spécialisé et non-spécialiste : Un problème de stratégie de décodage. *Revue belge de Philologie et d'Histoire*, 69(3), 629-643.

- Thoiron, P. (1998). Place et rôle de la terminologie dans les dictionnaires bilingues non spécialisés. Le cas de la terminologie médicale dans le dictionnaire Hachette-Oxford (français-anglais). Dans S. Mellet & M. Vuillaume (Éds), *Mots chiffrés et déchiffrés : Mélanges offerts à Etienne Brunet* (pp. 621-650). Paris: Honoré Champion.
- Thoiron, P., & Béjoint, H. (2010). La terminologie, une question de termes? *META*, 55(1), 105-118.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam, Philadelphie: John Benjamins.
- Tolédano, V., & Candel, D. (2006). Mouvance terminologique, l'espace d'une courte diachronie, dans les dictionnaires et dans la presse : Le cas de « mondialisation-globalisation ». Dans D. Candel & F. Gaudin (Éds), *Aspects diachroniques du vocabulaire* (pp. 157-183). Rouen: Presses universitaires de Rouen et du Havre.
- Tutin, A. (2007a). Autour du lexique et de la phraséologie des écrits scientifiques. *Revue française de linguistique appliquée*, 12(2), 5-14.
- Tutin, A. (2007b). Traitement sémantique par analyse distributionnelle des noms transdisciplinaires des écrits scientifiques. Dans *TALN 2007* (pp. 283-292). Toulouse.
- Tutin, A. (2010). *Sens et combinatoire lexicale, de la langue au discours*. Mémoire d'habilitation à diriger des recherches. Université Grenoble 3.
- Tutin, A., & Grossmann, F. (2002). Collocations régulières et irrégulières : Esquisse de typologie du phénomène collocatif. *Revue française de linguistique appliquée*, 8(1), 7-25.
- Tutin, A., & Grossmann, F. (Éds). (2013). *L'écrit scientifique : Du lexique au discours. Autour de Scientext*. Rennes: Presses universitaires de Rennes.
- Tutin, A., Grossmann, F., Falaise, A., & Kraif, O. (2009). Autour du projet Scientext : Étude des marques linguistiques du positionnement de l'auteur dans les écrits scientifiques. Dans *6e Journées Internationales de Linguistique de Corpus*. Lorient.
- Ungureanu, L. (2006). *L'interpénétration langue générale-langue spécialisée dans le discours d'internet*. Paris: Connaissances et Savoirs.
- Urieli, A. (2013). *Robust French Syntax Analysis: Reconciling Statistical Methods and Linguistic Knowledge in the Talismane Toolkit*. Thèse de doctorat en Sciences du Langage. Université Toulouse le Mirail, Toulouse II.
- Valette, M. (2016). Analyse statistique des données textuelles et traitement automatique des langues. Une étude comparée. Dans *JADT 2016 : 13èmes Journées internationales d'Analyse statistique des Données Textuelles*.
- Van Campenhoudt, M. (2006). Que nous reste-t-il d'Eugen Wüster? Dans *Eugen Wüster et la terminologie de l'École de Vienne*, Université de Paris 7.
- Van de Cruys, T. (2008). A Comparison of Bag of Words and Syntax-based Approaches for Word Categorization. Dans M. Baroni, S. Evert, & A. Lenci (Éds), *Proceedings of the Lexical Semantics Workshop: Bridging the gap between semantic theory and computational simulations* (pp. 47-54).

- Varantola, K. (1986). Special Language and General Language: Linguistic and Didactic Aspects. *Unesco Aised-LSP Newsletter*, 9(2), 10-20.
- Vega Moreno, É., & Llopert Saumell, E. (2017). Delimitación de los conceptos de novedad y neologicidad. *Rilce: Revista de Filología Hispánica*, 33(3), 1416-1451.
- Vicari, S. (2018). Ces termes qui ne vont pas de soi ou de la circulation de la terminologie des énergies renouvelables dans les forums en ligne. *Ela. Études de linguistique appliquée*, 192(4), 447-455.
- Villedieu, Y. (1996). Le sensationnalisme et le journalisme scientifique. *Québec français*, 102, 68-69.
- Viprey, J.-M., & Schepens, P. (2010). Dérivation lexicale et dérive du discours : « mutualisation, mutualiser ». Dans *Proceedings of 10th International Conference Journées d'Analyse Statistique des Données Textuelles (JADT 2010)* (pp. 489-498). Rome.
- Vogh, K., & Courbon, B. (2018). « For me, it was very genetic » : Usage de termes issus du domaine de la génétique dans le discours en ligne de diabétiques anglophones. Dans C. Roche (Éd.), *Actes de la conférence TOTh 2016* (pp. 83-103). Chambéry, France: Université Savoie Mont Blanc.
- Watine, T. (2006). De la multiplication des procédés interactionnels dans les contenus de presse : Vers un journalisme de conversation... *Les Cahiers du journalisme*, 16, 70-103.
- Wauquier, M. (2018). Analyse des noms agentifs dans les espaces vectoriels distributionnels. Dans *Actes de la conférence TALN 2018* (Vol. Volume 2, pp. 27-39). Rennes.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., ... RStudio. (2020). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* (version 3.3.1). <https://cran.r-project.org/package=ggplot2> [consulté le 12.06.2020]
- Wickham, H., François, R., Henry, L., Müller, K., & RStudio. (2020). *dplyr: A Grammar for Data Manipulation* (version 1.0.0). <https://cran.r-project.org/package=dplyr> [consulté le 7.06.2020]
- Williams, G. (1998). Collocational Networks: Interlocking patterns of lexis in a corpus of plant biology research articles. *International Journal of Corpus Linguistics*, 3(1), 151-171.
- Williams, G. (Éd.). (2005). *La linguistique de corpus*. Rennes: Presses universitaires de Rennes.
- Wright, S.E., & Budin, G. (Éds). (1997). *Handbook of terminology management. Volume 1: Basic aspects of terminology management*. Amsterdam: John Benjamins.
- Zweigenbaum, P., & Habert, B. (2004). Accès mesurés aux sens. *Mots. Les langages du politique*, 74, 93-106.

Index des auteurs

A

Abeillé, A. 113
Adelstein, A. 19, 35, 39, 53, 56, 66
Ahmad, K. 125
Alecú, B.P. 110, 111
Alexandru, C. 20, 25
Allauzen, A. 111
Alonso Campo, A. 28, 29
Anthony, L. 91, 107, 172
Assal, A. 13
Atkins, S. 128, 129, 132, 135
Aussenac-Gilles, N. 100, 101, 104, 175

B

Baayen, R.H. 221
Bakhtine, M. 237
Balvet, A. 111
Baneyx, A. 162
Bannour, S. 162, 179, 181
Barona, J.L. 16
Basile, P. 241
Bassi, E. 170, 171, 173
Beacco, J.-C. 25, 26, 120, 121, 137, 298, 299, 306
Béjoint, H. 19, 20, 28, 30, 33, 34, 39, 96, 98, 99
Bernier, M.-F. 120, 121, 162, 164, 167
Bernier-Colborne, G. 162, 164, 167
Bertaccini, F. 62, 77
Bertels, A. 93, 110, 161, 162, 164, 165, 168
Bestgen, Y. 170, 172, 181
Bhatia, V.K. 133, 134
Biber, D. 86, 87, 128, 129, 133, 170, 171, 181, 185
Bigras, I. 21, 28, 29
Boleda, G. 158, 162, 163, 164, 166, 168, 176
Bondi, M. 170, 171
Bonneau-Maynard, H. 111
Botta, M. 11, 12, 28, 39, 62
Bouhandi, M. 241
Boulangier, J.-C. 11, 17, 21, 28, 30, 31, 56, 96, 97, 99
Bourigault, D. 96, 97, 98, 99, 100, 104, 160, 161, 162, 173
Boussidan, A. 78
Bouveret, M. 11, 16, 19, 28
Bowker, L. 15, 16, 29, 56, 73, 75, 99, 128, 129, 133, 134, 135, 136, 140, 282, 283
Bres, J. 237
Brezina, V. 88, 114, 181, 185, 209
Brisson, P. 199, 308

C

Cabré, M.T. 11, 12, 13, 14, 15, 16, 17, 19, 26, 29, 30, 33, 56, 61, 77, 96, 97, 98, 99
Calberg-Challot, M. 13, 14, 16, 125
Campo, Á. 29, 97
Campos Andrés, O. 60

Candel, D. 29, 125
Candito, M. 113
Cañete, P. 26, 35
Carletta, J. 206
Carras, C. 26, 56
Cartier, E. 35, 36
Ceci, M. 258
Celotti, N. 11, 13
Ciapuscio, G.E. 24
Clas, A. 251
Cobos López, Í. 60
Cohen, J. 209
Condamines, A. 16, 26, 58, 67, 69, 70, 72, 73, 74, 75, 77, 78, 88, 92, 94, 95, 96, 97, 98, 100, 101, 102, 103, 104, 106, 128, 130, 134, 135, 137, 138, 140, 157, 158, 159, 163, 177, 211, 247, 281, 296, 307
Cori, M. 87
Cortès, C. 246
Costes, M. 120
Cottez, H. 21, 31, 32, 33
Courbon, B. 62, 63
Culpeper, J. 170, 171, 172, 173, 184, 185

D

Daille, B. 99, 110, 134, 162
Damerau, F.J. 185
David, S. 87
Dehaut, N. 101, 103
Deignan, A. 280
Del Tredici, M. 163, 164, 168, 175, 203, 280
Delavigne, V. 9, 11, 14, 15, 16, 19, 21, 22, 23, 24, 25, 27, 28, 34, 39, 63, 99, 124, 125, 135, 136, 138, 248
Deleu, C. 120
Demmen, J. 170, 171, 172, 173, 184, 185
Denis, P. 111
Depecker, L. 11, 14, 16, 17, 19, 29, 30, 33
Desagulier, G. 87, 93, 114, 165, 166, 190
Díaz Hormigo, M.T. 36, 56, 62
Drouin, P. 56, 91, 94, 99, 108, 151, 153, 164, 167, 246
Dubois, J. 39, 53, 56
Durieux, C. 12, 13, 14, 16
Dury, P. 11, 13, 14, 16, 19, 20, 66, 67, 68, 69, 70, 72, 75, 77, 135, 177, 200, 282

E

Egbert, J. 106, 114, 170, 171, 181, 183, 185
Engwall, G. 128
Estopà, R. 26, 35, 37, 248

F

Fabre, C. 88, 90, 93, 94, 95, 107, 158, 161, 163, 164, 165, 166, 173, 174, 175, 179, 180, 188
Falk, I. 111
Fernández Nistal, P. 29
Fernández, R. 29, 200, 248, 282, 283

Fernández-Silva, S..... 200, 248, 282, 283
 Firth, J.R. 157
 Fišer, D. 161, 163, 175, 178
 Fleiss, J.L. 209
 Freixa, J.....200, 282

G

Gabrielatos, C. 171, 172, 173, 181, 182, 183, 184, 185
 Galisson, R..... 30, 53, 54, 55, 56
 Galy, E. 162
 Gambier, Y..... 11, 19, 22, 96, 97, 99
 Gamer, M. 209
 Gardin, B. 35
 Gaudin, F.10, 11, 13, 14, 15, 16, 19, 20, 22, 23, 24, 25,
 26, 27, 28, 34, 35, 39, 63, 96, 97, 98, 99
 Gentilhomme, Y. 11, 53, 56
 Gérard, C. 36
 Gilbert, P. 34, 39
 Girardi, B..... 120
 Gonon, F.....120, 121
 González Davies, M. 60
 Gouadec, D.....57, 58
 Grefenstette, G.....130, 131, 188
 Gries, S.Th. 93, 114, 159, 171, 176, 181, 185
 Grossmann, F..... 251
 Grothendieck, G. 221
 Guespin, L..... 11, 35, 136
 Guilbert, L..... 9, 10, 11, 14, 28, 34, 35, 39, 40, 53, 78,
 123

H

Habert, B. 86, 87, 88, 89, 90, 91, 94, 128, 130, 158,
 159, 160, 171, 173, 174, 175, 187, 188, 203, 240,
 241, 242
 Hallgren, K.A.....206, 209
 Halskov, J.62, 70, 71, 72, 74, 75, 77, 78, 135, 137, 138,
 175, 177, 189, 296
 Hamon, T.162, 165, 167, 240, 241
 Hardie, A.86, 87, 88, 132
 Harris, Z.S. 157, 158, 159, 163
 Hazem, A..... 162
 Heiden, S. 91
 Hellrich, J..... 160, 163, 165, 166
 Heylen, K. 161, 162, 165
 Hilpert, M..... 176
 Horecký, J..... 46
 Humbley, J. 10, 13, 23, 61
 Hunston, S. 128, 132, 146, 179

I

Incelli, E.120, 121
 Inkpen, D. 110

J

Jaccard, P..... 241
 Jacobi, D..... 22, 23, 24, 55, 56
 Jacques, M.-P..... 86, 87, 88, 90, 93, 94, 95, 203, 206
 Jedlička, A..... 46

Jiménez-Crespo, M.A.....60
 Josselin-Leray, A. 12, 13, 14, 16, 21, 28, 29, 30, 31, 32,
 33, 52, 53, 99

K

Kageura, K.....283
 Kennedy, G. 86, 128, 130, 131, 135, 136, 146, 179
 Kilgarriff, A. 91, 130, 131, 181, 184, 185
 Kim, M.200, 282
 Kittredge, R.158
 Koch, G.G.....209
 Kocourek, R.15
 Kutuzov, A.163

L

Labasse, B. 120, 121
 Lafon, P.....93
 Landis, J.R.209
 Landragin, F. 86, 87, 88, 89, 90, 93, 94, 128
 Langlais, P.....56
 Larivière, L.....251
 Laval, S.120
 Le Draoulec, A.....27
 Ledouble, H..... 12, 26, 62
 Leech, G..... 128, 129, 130, 131, 132, 134
 Lehrberger, J..... 158
 Lenci, A. .. 158, 159, 160, 161, 162, 163, 164, 165, 166,
 167, 173, 174, 179, 180, 189
 Léon, J. 88, 89
 León-Araúz, P.283
 Lerat, P. 14, 15
 Lervad, S..... 200, 282
 Levshina, N.....93, 114, 240, 241
 Light, R.J.209
 Ljubešić, N.....161, 163, 175, 178
 Llopart Saumell, E. 35, 62
 Loffler-Laurian, A.-M. 24, 135

M

Mackintosh, K. 12, 13, 21, 34, 39, 46, 48, 49, 50, 51,
 52, 55, 57, 60, 62, 64, 65, 66, 67, 68, 69, 74, 75, 77,
 123, 128, 129, 134, 135, 168, 175, 177, 204, 219,
 220, 254, 260, 276, 278, 281, 285, 290, 309
 Maniez, F.....251
 Marchello-Nizia, C.88
 Marín Pérez, M.J.....110
 Martin-Lagardette, J.-L. 199, 308
 Masár, I.46
 Mayaffre, D.....89, 90, 91, 93, 94
 Mazière, F.29, 57, 58
 McEnery, T..... 86, 87, 88, 132, 171
 McGillivray, B.241
 Meillet, A.....10, 38, 39
 Meyer, I. .. 12, 13, 21, 26, 34, 39, 42, 45, 46, 47, 48, 49,
 50, 51, 52, 55, 57, 60, 61, 62, 63, 64, 65, 66, 67, 68,
 69, 70, 74, 75, 77, 79, 103, 123, 128, 129, 134, 135,
 168, 175, 177, 204, 219, 220, 254, 260, 276, 278,
 281, 285, 290, 308, 309
 Milizia, D.....170, 171

Moirand, S. 23, 24, 25, 26, 27, 28, 34, 39, 120, 137,
237, 248, 298, 299, 306
Montalt-Resurrecció, V. 60
Morlane-Hondère, F. 157, 158, 159, 160, 161, 162,
163, 165, 168, 173, 179, 180, 186, 189
Mortureux, M.-F. 19, 23, 24, 28, 39, 248
Mpouli, S. 280
Musacchio, M.T. 11, 13, 125

N

Nelson, M. 128, 132
Neveu, F. 276
Nicolae, C. 11, 135, 138
Nová, J. 46
Nowakowska, A. 237

O

Oliveira, I. 11, 277, 278
Ortego Antón, M.T. 29, 30

P

Pearson, J. 15, 16, 19, 27, 34, 39, 56, 73, 75, 85, 99,
128, 129, 130, 133, 134, 135, 136, 140, 143
Peirsman, Y. 160, 161, 163, 173
Pelletier, J. 200, 282
Péraldi, S. 14, 19, 281, 283, 285, 286, 288, 289
Périnet, A. 162, 165, 167, 240, 241
Péry-Woodley, M.-P. 87, 130, 131
Petit, G. 14, 19, 57, 77, 232, 278
Peynaud, C. 62, 63, 76, 212, 298
Phal, A. 151
Philip, G. 171, 280
Picton, A. 26, 58, 67, 69, 70, 72, 73, 74, 75, 77, 78, 88,
95, 96, 100, 101, 102, 103, 104, 135, 137, 138, 177,
200, 203, 296, 307
Pierrejean, B. 160, 161, 164, 165, 166
Poibeau, T. 16, 163
Pojanapunya, P. 170, 171, 181, 183, 184, 185
Polguère, A. 19, 29, 39, 62
Pontrandolfo, G. 60
Portelance, C. 68, 246
Poštolková, B. 46
Poudat, C. 86, 87, 88, 89, 90, 93, 94, 128
Prévost, S. 132
Pruvost, J. 35

R

Rastier, F. 58, 59
Rayson, P. 171, 185
Rebeyrolle, J. 281
Reboul-Touré, S. 23, 24
Reimerink, A. 283
Renouf, A. 35, 36, 39, 65, 66, 68, 69, 70, 71, 72, 74, 77,
135, 175
Reppen, R. 86, 87
Resche, C. 56, 62
Reutenauer, C. 36, 37, 38, 40, 41, 51, 59, 78, 290, 291
Roberts, R.P. 21, 28, 29, 30, 31, 33

Rogers, M. 283
Rondeau, G. 12, 13, 15, 19, 129
Rossi, M. 13, 29, 31, 32, 33
Rouillard, C.-A. 27
Rudis, B. 225
Rundell, M. 129, 135

S

Sablayrolles, J.-F. 35, 36, 37, 39, 40, 56, 62, 78
Sager, J. 13, 19
Sager, N. 158
Sagi, E. 162, 163
Sagot, B. 111
Sahlgren, M. 158, 162, 163, 165, 166, 179, 180
Sajous, F. 62
Sarale, J.-M. 237
Schepens, P. 41
Schmid, H. 111, 112, 181
Scott, M. 91, 93, 170, 172, 173, 181, 182, 184
Shafei-Bajestan, E. 221
Shuttleworth, M. 60
Siepmann, D. 135
Simard, I. 21, 28, 29
Sinclair, J. 85, 128, 135, 140
Slodzian, M. 96, 97, 98, 100, 104
Speelman, D. 93, 110, 164, 168
Stefanowitsch, A. 87, 93, 280
Stengers, I. 13

T

Taillet, R. 125, 154
Tanguy, L. 88, 90, 93, 94, 95, 107, 160, 164, 165, 166,
168, 173, 180, 181, 189
Taylor, C. 88
Temmerman, R. 13, 96, 97, 98, 99
Tercedor Sánchez, M. 60
Tercedor, M. 60, 283
Tetet, C. 29
Thoiron, P. 14, 15, 16, 17, 19, 20, 21, 29, 33, 96, 98, 99
Tognini-Bonelli, E. 86, 87, 128, 157
Tolédano, V. 29
Tribble, C. 170
Tutin, A. 89, 90, 92, 93, 135, 151, 167, 189, 240, 241,
251

U

Ungureanu, L. 11, 13, 15, 16, 19, 21, 28, 34, 39, 64, 66
Urieli, A. 112, 113, 189

V

Valette, M. 58, 59, 88, 94, 106
Van Campenhoudt, M. 96
Van de Cruys, T. 160, 173, 179, 188
Varantola, K. 11, 15, 28, 39
Vega Moreno, É. 35, 62
Vicari, S. 11, 16, 23
Villedieu, Y. 120
Viprey, J.-M. 41

Vogh, K.62, 63

Wickham, H.....221, 225

Williams, G.251, 252

W

Watine, T. 199

Watson Todd, R..... 170, 171, 181, 183, 184, 185

Wauquier, M. 166

Z

Zweigenbaum, P.94, 158, 159, 160, 173, 187, 240, 241

Index thématique

A

accord inter-annotateurs 106, 206, 209
analyse de discours 23, 237
analyse distributionnelle 157, 169, 178, 203
analyse distributionnelle automatique 161, 165
AntConc 107
approche outillée en corpus .. *Voir* linguistique outillée

B

banalisation *Voir* banalisation lexicale
banalisation lexicale 53, 59
banalisation lexicale (lexicographie) 30, 52

C

centralité (terme) 215, 219, 235, 262, 284, 305
centralité (thématique) 203
changement sémantique 27, 33, 38, 50, 66, 77, 163,
219, 245, 276, 289, 292
circulation des termes 9, 27, 60, 62, 99, 102
circulation langagière *Voir* circulation des termes
circulation terminologique. *Voir* circulation des termes
collocation spécialisée 251, 265
connotation 50, 204, 232, 279, 292
contexte de type « fenêtre de mots » 160, 169
contexte distributionnel 159, 179
contexte syntaxique 160, 173, 186, 239
continuum (langue de spécialité-langue générale) ... 14,
21, 78, 139
corpus 85, 100, 128, 132, 166
corpus comparable 102, 105, 147
corpus-based 87
corpus-driven 87, 155, 172
créativité lexicale *Voir* néologie

D

dédomanialisation 58
degré de spécialisation. 16, 74, 133, 139, 285, 288, 295
démarche outillée en corpus . *Voir* linguistique outillée
dépendance syntaxique *Voir* contexte syntaxique
déspecialisation *Voir* dé-spécialisation
dé-spécialisation 57
déterminologisation 13, 42, 46, 60, 74, 275, 303
déterminologisation (processus) 69, 74, 77, 292, 303
déterminologisation (résultat) 64, 74, 276
dialogisme 237
dictionnaires généraux *Voir* lexicographie
didacticité 25, 296, 299
diffusion des connaissances 23
dimension diachronique 69
dimension temporelle *Voir* dimension diachronique
discours de transmission des connaissances 25, 42,
137, 296
domaine 119, 285, 301, 305

E

équilibre 128, 131
évolution de la langue 77, 296
expert de domaine 100, 104, 105

F

facteurs de déterminologisation 305

G

genre de textes *Voir* genre textuel
genre textuel 133, 137, 259, 288
glissement de sens *Voir* changement sémantique
glissement sémantique .. *Voir* changement sémantique

H

hyperonyme 247
hyponyme 247
hypothèse distributionnelle 158, 162

I

indice de type distributionnel *Voir* indice
distributionnel
indice distributionnel 103, 169
interdisciplinarité 213, 215, 219, 229, 235, 262, 301,
305
intermédiaire 299
intermédiaire (degré de spécialisation) 136
intermédiaire (sous-corpus) 140, 293, 294
interprétation 94, 101, 104

K

keyness 181
keyword 169, 178, 197

L

langage banalisé 53
langue de spécialité 14
langue générale 14, 135
lexicographie 33, 42, 51, 52
lexique scientifique transdisciplinaire 151
linguistique de corpus 86, 170
linguistique outillée 85, 88, 92, 95
logiciel 91
LST *Voir* lexique scientifique transdisciplinaire

M

médias 26, 119, 142, 306
mesures statistiques *Voir* méthodes statistiques
métaphore 39, 46, 71, 168, 276, 279, 290

métaphorisation (lexicographie)	31
méthode statistique.....	92, 93
méthodes statistiques.....	95
mot	18
multidimensionnalité	281, 282, 285, 288
MySQL.....	115

N

néologicit�	37
néologie.....	34, 65, 289
néologie s�mantiq�	13, 38, 42, 58, 290
néologisme	35, 71
néologisme s�mantiq�	38, 291
néos�mie.....	<i>Voir</i> n�ologie s�mantiq�
nomadisation	13, 64

O

optique lexico-s�mantiq�	97
outil.....	89, 91
outils d'exploration de corpus	<i>Voir</i> outils de traitement de corpus
outils de traitement automatique des langues	90
outils de traitement de corpus	90, 95

P

paradigme d�finitionnel	24, 297
paradigme d�signationnel	24, 297
point de vue	281
potentiel m�taphorique	276, 292, 309
presse.....	26, 36, 136, 199, 237, 297, 298
presse g�n�raliste.....	<i>Voir</i> presse

R

R	114
rep�rage semi-automatique des candidats-termes..	148
repr�sentativit�	128, 132
reproductibilit�	94, 107
ressource terminographique.....	<i>Voir</i> ressource terminologique
ressource terminologique.....	100, 152
RStudio	114

S

s�mantiq� distributionnelle.....	<i>Voir</i> analyse distributionnelle automatique
sensationnalisme	120, 201, 236, 249, 263, 306
sentiment n�ologique	35
similarit� de Jaccard.....	240, 259
soci�t� du savoir.....	47
sociodiffusion des connaissances.....	56
sociodiffusion des sciences	<i>Voir</i> sociodiffusion des connaissances
socioterminologie	11, 23, 97

T

TAL.....	<i>Voir</i> traitement automatique des langues
Talismane	112
terme	18, 20
terme m�taphorique	32, 46, 277
terminologie outill�e.....	96
terminologie sociocognitive.....	97
terminologie textuelle.....	97, 100, 102, 105
terminologisation	13, 64
TermoStat	108
th�orie communicative de la terminologie	61, 97
Th�orie G�n�rale de la Terminologie	96
traitement automatique des langues	92
traitement automatique des langues	88
traitement automatique des langues	161
transmission des connaissances	<i>Voir</i> diffusion des connaissances
TreeTagger.....	111

V

variante d�nominative.....	200, 248, 297
variante terminologique.....	103
variation conceptuelle	282
variation d�nominative	283
variation des termes	<i>Voir</i> variation terminologique
variation terminologique	97
VGOS.....	<i>Voir</i> Vocabulaire g�n�ral d'orientation scientifique
Vocabulaire g�n�ral d'orientation scientifique	151
vulgarisation.....	<i>Voir</i> vulgarisation scientifique
vulgarisation scientifique	22, 58, 122, 257, 269, 296

Annexes

Annexe A – Constitution du corpus

Cette annexe précise la constitution du corpus, d'abord en listant les différentes sources des textes constitutifs de chaque sous-corpus, dans le tableau A.1, puis en indiquant la répartition des occurrences dans les sous-corpus, selon les sources et selon les périodes, dans le tableau A.2.

Sous-corpus <i>Spécialisé</i>		
Genres		
Article de revue spécialisée	Publications	Bulletin de la Société Française de Physique (jusqu'en août 2006) Reflets de la physique (dès septembre 2006)
	Source	Société Française de Physique
	Pays	France
	URL	https://www.refletsdelaphysique.fr/
	Date d'extraction	Août-septembre 2016
Thèse de doctorat	Universités	Université Montpellier 2 Université Claude Bernard Lyon-1 Université de la Méditerranée, Aix-Marseille II Université Paris VI Université de Grenoble Université Paris-Sud Université de Strasbourg Université Blaise Pascal, Clermont-Ferrand
	Pays	France
	Date d'extraction	Août-septembre 2016
Sous-corpus <i>Communiqués</i>		
Genre		
Communiqué de presse	Sources	Organisation européenne pour la recherche nucléaire (CERN) Institut national de physique nucléaire et de physique des particules (IN2P3)
	Pays	Suisse, France
	URL	https://home.cern/fr https://www.in2p3.cnrs.fr/fr/institut-national-de-physique-nucleaire-et-de-physique-des-particules-0
	Date d'extraction	Janvier-février 2017
Sous-corpus <i>Rapports</i>		
Genre		
Rapport de laboratoire « grand public »	Sources	Organisation européenne pour la recherche nucléaire (CERN) Laboratoire de l'Accélérateur Linéaire (LAL) Laboratoire de Physique Subatomique & Cosmologie (LPSC)
	Pays	Suisse, France
	URL	https://home.cern/fr https://www.lal.in2p3.fr/ http://lpsc.in2p3.fr/index.php/fr/
	Périodicité	CERN : un rapport/an

		LAL : un rapport/deux ans (jusqu'en 2004) ; un rapport/trois ans LPSC : un rapport/deux ans
	Date d'extraction	Septembre-novembre 2016
Sous-corpus <i>Vulgarisation</i>		
Genres		
Article de revue de vulgarisation	Revues	La Recherche Pour la Science Élémentaire
	Pays	France
	URL	https://www.larecherche.fr/ https://www.pourlascience.fr/ https://elementaire.lal.in2p3.fr/
	Date d'extraction	Décembre 2016-mars 2017
Sites Web de vulgarisation	Sites	CERN LHC-France
	Pays	Suisse, France
	URL	https://home.cern/fr https://www.lhc-france.fr/
	Date d'extraction	Décembre 2016
Sous-corpus <i>Presse</i>		
Genre		
Article de presse généraliste	Journaux	Le Temps Le Monde Les Échos Le Figaro
	Pays	Suisse, France
	Date d'extraction	Juin-juillet 2017

Tableau A.1 : Source des textes constitutifs du corpus

Sous-corpus <i>Spécialisé</i>			
Périodes	Sources	Nombre de textes	Nombre d'occurrences
2003-2007	Bulletin de la Société Française de Physique Reflets de la physique Université Claude Bernard Lyon-1 Université de la Méditerranée, Aix-Marseille II Université Paris VI Université Paris-Sud	3 1 3 2 1 1	314 658
2008-2011	Reflets de la physique Université Montpellier 2 Université Claude Bernard Lyon-1 Université de la Méditerranée, Aix-Marseille II Université Paris-Sud Université Blaise Pascal, Clermont-Ferrand	4 1 1 2 1 1	330 975
2012-2016	Reflets de la physique Université Montpellier 2 Université Claude Bernard Lyon-1 Université de Grenoble Université Paris-Sud Université de Strasbourg	4 1 2 1 1 1	349 242
Sous-corpus <i>Communiqués</i>			
Périodes	Sources	Nombre de textes	Nombre d'occurrences
2003-2007	CERN IN2P3	45 31	70 950
2008-2011	CERN IN2P3	58 28	69 478
2012-2016	CERN IN2P3	55 32	69 892
Sous-corpus <i>Rapports</i>			
Périodes	Sources	Nombre de textes	Nombre d'occurrences
2003-2007	CERN LAL LPSC	5 2 3	516 820
2008-2011	CERN LAL LPSC	4 1 2	302 552
2012-2016	CERN LAL LPSC	4 1 2	322 501
Sous-corpus <i>Vulgarisation</i>			
Périodes	Sources	Nombre de textes	Nombre d'occurrences
2003-2007	La Recherche Pour la Science Élémentaire	15 5 5	216 969
2008-2011	La Recherche Pour la Science	17 16	194 675

	Élémentaire	3	
2012-2016	La Recherche	17	208 401
	Pour la Science	19	
	Élémentaire	1	
	CERN	1	
	LHC-France	1	
Sous-corpus <i>Presse</i>			
Périodes	Sources	Nombre de textes	Nombre d'occurrences
2003-2007	Le Temps	118	367 378
	Le Monde	136	
	Les Échos	139	
	Le Figaro	130	
2008-2011	Le Temps	135	365 650
	Le Monde	194	
	Les Échos	140	
	Le Figaro	120	
2012-2016	Le Temps	120	365 680
	Le Monde	164	
	Les Échos	139	
	Le Figaro	109	

Tableau A.2 : Répartition des occurrences selon les sources et les périodes considérées

Annexe B – Liste des termes extraits

Termes extraits et validés par les expertes (à l'issue du quatrième filtre, cf. Chapitre 4, 4.3) :

accélérateur de particules ; accélérateur linéaire ; ALICE ; antihydrogène ; antimatière ; antineutrino ; antiparticule ; antiproton ; antiquark ; asymétrie ; baryon ; boson ; boson de Higgs ; boson de jauge ; CERN ; champ scalaire ; charmé ; collision d'ions ; collision d'ions lourds ; collision proton-proton ; collisionneur ; collisionneur de hadrons ; collisionneur linéaire ; corpusculaire ; désintégration bêta ; électrofaible ; électron ; électron-positron ; électronvolt ; énergie de collision ; faisceau d'électrons ; faisceau de neutrinos ; faisceau de particules ; faisceau de protons ; fermion ; gigaélectronvolt ; gluon ; hadron ; interaction faible ; interaction fondamentale ; interaction forte ; kaon ; kiloélectronvolt ; lepton ; LHC ; masse du Higgs ; matière noire ; méson ; mode de désintégration ; modèle standard ; muon ; muonique ; neutrino ; neutrino électronique ; neutrino muonique ; neutrino stérile ; neutrino tau ; neutron ; nucléon ; particule chargée ; particule élémentaire ; photomultiplicateur ; photon ; physique corpusculaire ; physique subatomique ; pion ; plasma de quarks-gluons ; point de collision ; positron ; proton ; proton-proton ; quark ; quark b ; quark top ; radiofréquence ; rayon cosmique ; relativiste ; relativité restreinte ; saveur ; scintillateur ; sous-détecteur ; spectromètre ; subatomique ; supersymétrie ; supersymétrique ; symétrie électrofaible ; synchrotron ; tau ; télescope ; Tevatron ; trajectographe ; trajectomètre.

Termes sélectionnés pour la mise en œuvre du premier indice (cf. Chapitre 5, 5.3.1.1) :

accélérateur de particules ; accélérateur linéaire ; ALICE ; antimatière ; antiparticule ; antiproton ; asymétrie ; boson ; boson de Higgs ; CERN ; collision proton-proton ; collisionneur ; électron ; électron-positron ; électronvolt ; faisceau de protons ; fermion ; gigaélectronvolt ; gluon ; hadron ; interaction faible ; interaction fondamentale ; interaction forte ; lepton ; LHC ; matière noire ; modèle standard ; muon ; neutrino ; neutrino électronique ; neutrino muonique ; neutrino tau ; neutron ; nucléon ; particule chargée ; particule élémentaire ; photon ; plasma de quarks-gluons ; positron ; proton ; quark ; quark top ; radiofréquence ; rayon cosmique ; relativiste ; relativité restreinte ; saveur ; spectromètre ; supersymétrie ; supersymétrique ; synchrotron ; tevatron.

Termes sélectionnés pour la mise en œuvre du second indice (cf. Chapitre 5, 5.3.2.1) :

accélérateur ; atome ; boson de Higgs ; collisionneur ; détecteur ; électron ; interaction ; ion ; muon ; neutrino ; neutron ; noyau ; particule ; proton ; quark.