



HAL
open science

Les routines de guidage du lecteur dans les écrits scientifiques en français

Yujing Ji

► **To cite this version:**

Yujing Ji. Les routines de guidage du lecteur dans les écrits scientifiques en français. Linguistique. Université Grenoble Alpes [2020-..], 2022. Français. NNT : 2022GRALL023 . tel-03999117

HAL Id: tel-03999117

<https://theses.hal.science/tel-03999117>

Submitted on 21 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : LLSH- Langues, Littératures et Sciences Humaines

Spécialité : Sciences du langage Spécialité Didactique et Linguistique

Unité de recherche : Laboratoire de Linguistique et Didactique des Langues Etrangères et Maternelles

Les routines de guidage du lecteur dans les écrits scientifiques en français

Routines for guiding the reader in scientific writing in French

Présentée par :

Yujing JI

Direction de thèse :

Agnès TUTIN
Université Grenoble Alpes

Directrice de thèse

Olivier KRAIF
Université Grenoble Alpes

Co-directeur de thèse

Rapporteurs :

Mojca PECMAN
PROFESSEUR DES UNIVERSITES, Université Paris 7 Denis Diderot

Shirley CARTER-THOMAS
PROFESSEUR DES UNIVERSITES EMERITE, Mines Paris Tech

Thèse soutenue publiquement le **11 octobre 2022**, devant le jury composé de :

Agnès TUTIN
PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes

Directrice de thèse

Olivier KRAIF
PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes

Co-directeur de thèse

Mojca PECMAN
PROFESSEUR DES UNIVERSITES, Université Paris 7 Denis Diderot

Rapporteuse

Shirley CARTER-THOMAS
PROFESSEUR DES UNIVERSITES EMERITE, Mines Paris Tech

Rapporteuse

Francis GROSSMANN
PROFESSEUR DES UNIVERSITES EMERITE, Université Grenoble Alpes

Examineur

Denis VIGIER
PROFESSEUR DES UNIVERSITES, Université Lyon 2 A&L Lumière

Examineur

Invités :



Remerciements

Je tiens à exprimer ma plus sincère gratitude à tous ceux qui m'ont soutenue, soignée et aidée pendant ces cinq années de thèse à l'Université Grenoble Alpes.

Je tiens à dire d'abord ma sincère gratitude à ma directrice de thèse Madame Agnès Tutin et à mon co-directeur de thèse Monsieur Olivier Kraif. Avant la thèse, j'avais fait un Master en Didactique du Français Langue Étrangère à l'Université Sorbonne Nouvelle. C'est grâce à Madame Cristelle Cavalla, ma directrice de mémoire de Master qui m'a fait découvrir la phraséologie et qui m'a recommandé Madame Tutin, que j'ai fait ma thèse sur ce sujet. Je remercie sincèrement à Madame Tutin d'avoir pris le risque de m'encadrer pour une thèse en linguistique, domaine dans lequel je n'avais pas de base solide au début. Au cours des cinq dernières années, Mme Tutin m'a accordé une attention sans faille dans mes études, et m'a donné toutes les chances possibles de progresser dans mes études et ma formation en linguistique. À l'instar de Madame Tutin, Monsieur Kraif a aussi joué un rôle très important dans l'encadrement de ma thèse. En tant que professeur en Traitement Automatique des Langues, Monsieur Kraif m'a beaucoup aidée dans la manipulation du corpus et des outils exploratoires, dans les analyses quantitatives. Je le remercie de m'avoir appris à utiliser les logiciels et les méthodes expérimentales requis, d'avoir pris le temps de s'occuper de mes problèmes informatiques à de nombreuses reprises, de m'avoir aidée à réaliser des analyses quantitatives à maintes reprises et d'avoir essayé par tous les moyens de m'aider à résoudre les problèmes liés à la méthodologie. Sans lui, je n'aurais jamais maîtrisé les méthodes en linguistique de corpus ou en analyses quantitatives. Je peux dire que chaque étape de ma croissance et chaque réalisation ont été améliorées grâce aux efforts de mes deux directeurs de thèse. J'ai grandement bénéficié des leurs conseils pour l'avancement de ma recherche et pour la révision des articles et de la thèse. Je n'oublierai jamais combien de fois ils ont relu mes propositions de colloques, mes communications scientifiques et mes articles. En résumé, ils m'ont beaucoup aidée à me former la recherche en linguistique, et m'ont inspirée pour devenir enseignant-chercheur. Sans leur enseignement, je ne serais pas là où je suis aujourd'hui.

Je voudrais remercier ensuite les rapporteurs et les examinateurs de ma thèse : Madame Mojca Pecman, Madame Shirley Carter-Thomas, Monsieur Francis Grossmann et Monsieur Denis Vigier pour le temps et les efforts qu'ils ont consacrés à l'examen de ma thèse.

Merci également à Madame Cécile Frérot pour la vérification de la traduction des citations anglaises en français. Votre aide a rendu plus fidèles les idées véhiculées dans ces citations.

Je suis très reconnaissante à l'équipe du laboratoire Lidilem, au sein duquel j'ai passé cinq années merveilleuses et fini mon travail de recherche. Les cours proposés par les enseignants et les services assurés par les personnels administratifs du Lidilem ont rendu ma vie de thèse riche et agréable. Je tiens à remercier mes chers amis et collègues qui forment une grande famille de partage des pensées, des cultures et des joies, que nous nommons « Lidilaime » : Alexis, Azadeh, Lan, Chaeyoung, Célia, Claire (H), Claire (W), Émile, Fabio, Iris, Jingrao, Kim, Laurence, Luca, Marie, Manon, Nicolas, Rose, Roxanne, Qianyun, Wenjie, Xinjiletu ; mais aussi des doctorants en échanges comme Érica, Léandro, Letty, Roland... Je voudrais remercier particulièrement à Claire (W), Claire (H) et Roxanne pour leur relecture minutieuse. Je suis très heureuse d'avoir passé les cinq ans les plus inoubliables de ma vie avec les chers membres du Lidilem. Notre amitié se manifestait partout, dans les soirées, dans les sorties, dans les voyages et aussi dans tous les moments de chaque jour. Leurs sourires et leur amitié sont pour moi la plus grande richesse. Bien que je rentre en Chine après la thèse, mon cœur sera toujours avec eux, avec le Lidilem. Je voudrais citer ici une ligne de poème très célèbre en chinois : 海内存知己, 天涯若比邻 (hǎi nèi cún zhī jǐ, tiān yá ruò bǐ lín) qui signifie que la distance ne peut pas briser notre amitié.

La relecture des brouillons de ma thèse n'a pas été facile. Je suis reconnaissante à Madame Christine Richard, une bénévoles de Coup de Pouce, à Monsieur Stéphane Girard et à mon ami Kylian pour le temps qu'ils ont consacré à la relecture de la première version de ma thèse.

Mes remerciements s'adressent ensuite à mes amis à Grenoble : Congying, Huirui, Jinrao, Lan, Mengxue, Qianyun, Wenjie, Xiaojia, Xinwei, Xinjitelu. Leur soutien et leur accompagnement me sont précieux : ils enrichissent non seulement mes pensées, mais aussi m'ont accompagnée tous les hauts et les bas de la vie.

Un grand merci enfin à ma famille en Chine. Je pense à mes parents qui me soutiennent économiquement et psychologiquement. Malgré la distance, nous avons régulièrement communiqué par Wechat, ce qui m'a apporté beaucoup de soutien. Leur amour est loin, mais profond.

Table des matières

Remerciements	I
Table des matières	III
Index des illustrations	IX
Index des tableaux	XI
Introduction	1
Partie I. Aborder les routines de guidage du lecteur des écrits scientifiques dans les perspectives phraséologique, énonciative et fonctionnelle	7
Chapitre 1. Routines de guidage du lecteur dans la phraséologie étendue : précisions terminologiques	9
1.1. La phraséologie étendue : une extension des objets d'étude et des domaines impliqués	10
1.2. Routines sémantico-rhétoriques et notions connexes	15
1.2.1. Quelques observations sur les routines sémantico-rhétoriques	16
1.2.2. Routines conversationnelles.....	19
1.2.3. Des <i>sentence builders</i> aux <i>p-frames</i> : quelques notions utiles dans la phraséologie de langue anglaise	23
1.2.3.1. Sentence builders (Nattinger & DeCarrico, 1992) et Sentence frames and heads (Lewis, 1996).....	23
1.2.3.2. Lexical bundles (Biber, 2006, 2009), p-frames (Römer, 2010) et lexical frames (Gray & Biber, 2015)	26
1.2.4. Des <i>segments répétés</i> aux <i>motifs</i> : des notions de la lexicométrie et de l'analyse du discours en langue française	30
1.2.4.1. Segments répétés (Salem, 1987, p. 21).....	30
1.2.4.2. Motifs	32
1.2.4.3. Routines discursives.....	34
1.2.5. Synthèse	35
1.2.6. Observations sur les routines sémantico-rhétoriques	38
1.2.6.1. Récurrences des routines.....	38
1.2.6.2. Nature lexicale, sémantique et phrastique des routines	39
1.2.6.3. Fonctions rhétoriques des routines	44
1.2.6.4. Paramètres énonciatifs des routines	45
1.3. Routines de guidage du lecteur dans les écrits scientifiques : état de lieux et définition	47
1.4. Synthèse.....	49
Chapitre 2. Écrits scientifiques dans l'approche dialogique pour l'étude des routines	51
2.1. Le genre des écrits scientifiques	51
2.2. Articles de recherche.....	55
2.2.1. Connaissances scientifiques dans les articles de recherches	56
2.2.2. Organisation des articles de recherches	57
2.3. Les écrits scientifiques : un lieu de dialogisme	58
2.3.1. Monologalité vs dialogalité	59
2.3.2. Le dialogisme dans les écrits scientifiques	61
2.3.2.1. Dialogisme interdiscursif.....	61
2.3.2.2. Dialogisme interlocutif	63
2.3.2.3. Dialogisme intralocutif	65
2.3.2.4. Croisement des dimensions dialogiques	65
2.3.3. Les rôles de l'auteur et l'enrôlement du lecteur dans les articles de recherches	66

2.3.3.1.	Les rôles de l'auteur dans les articles de recherche	66
2.3.3.2.	L'enrôlement du lecteur dans les articles de recherche.....	68
2.4.	Guidage du lecteur dans les écrits scientifiques : procédés extralinguistiques et linguistiques	69
2.5.	Synthèse.....	70
Chapitre 3.	Le métadiscours pour l'étude des fonctions de guidage.....	71
3.1.	Classifications des expressions phraséologiques dans les écrits scientifiques	71
3.1.1.	Le Louvain <i>English Academic Purposes</i> Dictionary	71
3.1.2.	Séquences discursives (Tran, 2014).....	72
3.1.3.	Marqueurs discursifs polylexicaux (Siepmann, 2007).....	73
3.1.4.	<i>Academic Formulas List</i> (Simpson-Vlach & Ellis, 2010)	74
3.1.5.	<i>Academic Phrasebank</i> (Davis & Morley, 2018).....	75
3.1.6.	Classement fonctionnel des routines sémantico-rhétoriques (Chevalier, 2019).....	76
3.2.	Métadiscours	78
3.2.1.	Genèse de la notion de métadiscours	79
3.2.2.	Définition du métadiscours	81
3.2.2.1.	Métadiscours des modèles interactifs et interpersonnels	81
3.2.2.2.	Métadiscours du modèle réflexif.....	85
3.2.2.3.	Définitions du métadiscours de ses différents modèles	88
3.2.3.	Classements fonctionnels des marqueurs métadiscursifs	89
3.2.3.1.	Classements fonctionnels des marqueurs métadiscursifs dans les modèles interactifs/interpersonnels	89
3.2.3.2.	Classements fonctionnels des marqueurs métadiscursifs dans le modèle réflexif	95
3.2.3.3.	Les modèles métadiscursifs et leurs apports à notre étude	100
3.3.	Synthèse.....	101
Conclusion de la partie I.....	102	
Partie II. Recueil, traitement des routines de guidage du lecteur dans le corpus TermITH et élaboration d'un modèle d'analyse	105	
Chapitre 4. Extraction, identification et modélisation des routines de guidage du lecteur.....	107	
4.1.	Linguistique de corpus pour l'étude phraséologique	107
4.2.	Corpus TermITH-transdisciplinaire : constitution, annotation et intérêt.....	109
4.2.1.	Composition du corpus	109
4.2.2.	Annotation effectuée sur le corpus	111
4.3.	Outil d'exploitation et la méthode des Arbres Lexico-syntaxiques Récurrents.....	114
4.3.1.	Outil d'exploitation – Le Lexicoscope	115
4.3.1.1.	Lexicogrammes	115
4.3.1.2.	Concordance	116
4.3.1.3.	Recherche d'expressions polylexicales.....	118
4.3.2.	Méthode des Arbres Lexico-syntaxiques Récurrents.....	118
4.4.	Étape 1. Extraction automatique des candidats de routines.....	120
4.4.1.	Sélection des classes sémantiques dans le LST comme pivots de recherche	121
4.4.1.1.	Ajout des mots courants qui ne sont pas inclus dans le LST.....	122
4.4.1.2.	Prétraitement des mots polylexicaux dans le LST	124
4.4.1.3.	Sélection des classes sémantiques comme pivots de recherche.....	125
4.4.2.	Mise en forme des classes sémantiques pour la recherche	126
4.4.3.	Élaboration des paramètres de l'extraction automatique et lancement de l'extraction	128

4.5. Étape 2. Filtrage, identification et regroupement des occurrences de routines dans les données extraites	131
4.6. Étape 3. Élaboration d'un modèle d'analyse pour des routines de guidage.....	137
4.6.1. Quelques références pour la modélisation des expressions phraséologiques	137
4.6.1.1. Modèles simples	138
4.6.1.2. Modèle à schéma collocationnel de Pecman (2004)	141
4.6.1.3. Modèle de « concept-matching » de Sandor (2007)	144
4.6.2. Élaboration d'un modèle d'analyse pour les routines de guidage du lecteur.....	145
4.6.2.1. Modèle de base	146
4.6.2.2. Notre modèle d'analyse après adaptation.....	149

Conclusion de la partie II..... 154

Partie III. Typologie, modélisation et analyses des routines de guidage du lecteur.....157

Chapitre 5. Routines de structuration textuelle : définition, modélisation et analyses 161

5.1. Structuration dans la linguistique textuelle.....	161
5.2. Fonctions structurantes dans le métadiscours.....	164
5.3. Routines de structuration textuelle.....	165
5.3.1. Routines de fonction de planification de la structure discursive	165
5.3.1.1. Caractéristiques syntaxiques, lexicales et sémantiques de ces routines	168
5.3.1.2. Éléments périphériques et paramètres énonciatifs	171
5.3.1.3. Fonctionnement rhétorique.....	172
5.3.2. Routines d'annonce de l'objectif/l'objet d'étude	174
5.3.2.1. Caractéristiques syntaxiques, lexicales et sémantiques de ces routines.....	177
5.3.2.2. Éléments périphériques et paramètres énonciatifs	180
5.3.2.3. Fonctionnement rhétorique.....	181
5.3.3. Routines de navigation textuelle.....	183
5.3.3.1. Caractéristiques syntaxiques, lexicales et sémantiques de ces routines.....	189
5.3.3.2. Éléments périphériques et paramètres énonciatifs	192
5.3.3.3. Fonctionnement rhétorique.....	193
5.3.4. Routines de conclusion	193
5.3.4.1. Caractéristiques syntaxiques, lexicales et sémantiques de ces routines.....	194
5.3.4.2. Éléments périphériques et paramètres énonciatifs	196
5.3.4.3. Fonctionnement rhétorique.....	196
5.4. Synthèse.....	197

Chapitre 6. Routines à fonction métalinguistique : définitions, modélisations et analyses 199

6.1. Précisions sur le terme « métalinguistique »	199
6.2. Fonctions métalinguistiques dans le métadiscours	201
6.3. Routines métalinguistiques	203
6.3.1. Routines de reformulation	204
6.3.1.1. Caractéristiques syntaxiques, lexicales et sémantiques de ces routines	206
6.3.1.2. Éléments périphériques et paramètres énonciatifs	208
6.3.1.3. Fonctionnement rhétorique.....	209
6.3.2. Routines de réparation	210
6.3.2.1. Caractéristiques syntaxiques, lexicales et sémantiques de ces routines	211
6.3.2.2. Éléments périphériques et paramètres énonciatifs	212
6.3.2.3. Fonctionnement rhétorique.....	213

6.3.3.	Routines de précision terminologique	213
6.3.3.1	Caractéristiques syntaxiques, lexicales et sémantiques de ces routines	217
6.3.3.2	Éléments périphériques et paramètres énonciatifs	221
6.3.3.3	Fonctionnement rhétorique.....	223
6.3.4.	Routines d'emprunt terminologique.....	224
6.3.4.1	Caractéristiques syntaxiques, lexicales et sémantiques de ces routines	225
6.3.4.2	Éléments périphériques et paramètres énonciatifs	227
6.3.4.3	Fonctionnement rhétorique.....	227
6.4.	Synthèse.....	228
Chapitre 7. Routines de topicalisation : définitions, modélisations et analyses		229
7.1.	Précision sur le terme « topicalisation »	229
7.2.	Routines de topicalisation	231
7.2.1.	Routines de mise en évidence	231
7.2.1.1	Caractéristiques syntaxiques, lexicales et sémantiques de ces routines	233
7.2.1.2	Éléments périphériques et paramètres énonciatifs	235
7.2.1.3	Fonctionnement rhétorique.....	237
7.2.2.	Routines de (co)constat.....	237
7.2.2.1	Caractéristiques syntaxiques, lexicales et sémantiques de ces routines	238
7.2.2.2	Éléments périphériques et paramètres énonciatifs	240
7.2.2.3	Fonctionnement rhétorique.....	241
7.2.3.	Routines de focus	241
7.2.3.1	Caractéristiques syntaxiques, lexicales et sémantiques de ces routines	243
7.2.3.2	Éléments périphériques et paramètre énonciatif.....	245
7.2.3.3	Fonctionnement rhétorique.....	246
7.3.	Synthèse.....	246
Chapitre 8. Routines de guidage à travers les parties textuelles, l'actualisation énonciative et les disciplines		249
8.1.	Positions textuelles des routines de guidage.....	249
8.1.1.	Emplacement des routines de structuration dans le texte	250
8.1.2.	Positions textuelles des routines métalinguistiques	255
8.1.3.	Positions textuelles des routines de topicalisation	256
8.1.4.	Synthèse.....	257
8.2.	L'actualisation énonciative des routines de guidage	258
8.2.1.	Les marques explicites d'actualisation énonciative des routines	258
8.2.2.	Le mode implicite d'actualisation énonciative des routines.....	260
8.2.3.	Les pronoms personnels et les rôles d'auteur des routines.....	261
8.2.3.1	Les rôles d'auteur scientifique à travers les pronoms personnels	262
8.2.3.2	L'observation des pronoms personnels et l'enrôlement du lecteur	265
8.2.4.	Synthèse	266
8.3.	Distribution des routines de guidage.....	267
8.3.1.	Répartition des routines dans le corpus global	267
8.3.1.1	Routines de structuration : proportion et fréquences	270
8.3.1.2	Routines métalinguistiques : proportions et fréquences	273
8.3.1.3	Routines de topicalisation : proportion et fréquences	275
8.3.1.4	Synthèse	277
8.3.2.	Routines de guidage à travers les disciplines	278
8.3.2.1	Hypothèses sur les disciplines	279

8.3.2.2 Distances entre les disciplines	281
8.3.2.3 Corrélation entre les disciplines et les routines	286
8.3.3. Synthèse	290
8.4. Synthèse.....	291
Conclusion de la partie III :	294
Conclusion générale	295
Bibliographie	301
Annexe 1	319
Annexe 2	325
Résumé	327
Abstract.....	327

Index des illustrations

Figure 1-1 : Routine pour s'excuser	22
Figure 1-2 : Relation inclusive des routines	47
Figure 2-1 : Le continuum des connaissances scientifiques (Hyland, 2009, p. 63)	56
Figure 3-1 : L'interface du dictionnaire LEAP (Granger & Paquot, 2010, p. 323)	72
Figure 3-2 : Triangle réflexif du métadiscours (Ädel, 2006, p. 18)	86
Figure 3-3 : Configuration du métadiscours chez Ädel (2006, p. 21)	87
Figure 4-1 : Les deux étapes de constitution du corpus TermITH	111
Figure 4-2 : Exemple d'analyse par XIP, repris de Hatier (2016, p. 66)	112
Figure 4-3 : Exemple d'analyse morpho-syntaxique et syntaxique d'une phrase	114
Figure 4-4 : Tableaux de cooccurrences du pivot « définition »	116
Figure 4-5 : Extrait des concordances du pivot définition avec le collocatif proposer du Lexicoscope	117
Figure 4-6 : ALR correspondant à on_PRON + reprendre_VERB + le_DET + définition_NOUN	119
Figure 4-7 : Les démarches pour l'extraction et le traitement de routines de guidage	120
Figure 4-8 : Formatage des classes sémantiques du LST pour le Lexicoscope	127
Figure 4-9 : Un ALR avec classes sémantiques	127
Figure 4-10 : Extraite du fichier HTML pour le pivot \$ PROCESHUMUS	131
Figure 4-11 : Structuration des données après nettoyage	136
Figure 4-12 : Schéma collocationnel de la citation en français (Pecman, 2004, p. 414)	143
Figure 4-13 : Fiche analytique d'une routine (Tutin & Ji, 2022, p. 59–60)	146
Figure 4-14 : Classes sémantiques des verbes dans une routine de planification	152
Figure Partie III-1 : Fonctions des routines de guidages dans les écrits scientifiques	159
Figure 5-1 : Modèle CARS pour l'introduction des articles de recherche en anglais (Swales, 1990, p.141)	173
Figure 6-1 : Les sous-fonctions métalinguistiques	203
Figure 7-1 : Les sous-fonctions de la topicalisation	231
Figure 8-1 : Proportion des fréquences d'occurrence des trois fonctions des routines de guidage	269
Figure 8-2 : Proportion des quatre sous-fonctions dans la structuration	270
Figure 8-3 : Occurrences des routines de planification à travers les disciplines	271
Figure 8-4 : Occurrences des routines de structuration	272
Figure 8-5 : Proportion des quatre sous-fonctions métalinguistiques dans le corpus	273
Figure 8-6 : Occurrences des routines métalinguistiques	275
Figure 8-7 : Répartition des trois sous-fonctions dans topicalisation	276
Figure 8-8 : Occurrences des routines de topicalisation	276

Index des tableaux

Tableau 1-1 : Récapitulatif des caractéristiques des notions.....	35
Tableau 1-2 : Figement syntaxique des notions connexes	37
Tableau 1-3 : L'exemple d'une routine d'annonce de l'objectif d'étude, extrait de notre corpus	45
Tableau 2-1 : Typologie du discours scientifique(Loffler-Laurian, 1983, p. 9–12)	55
Tableau 2-2 : Les valeurs de ON dans les articles de recherche en français (Fløttum et al., 2006, p. 117).....	68
Tableau 3-1 : Fonctions des marqueurs discursifs polylexicaux chez Siepmann (2007).....	73
Tableau 3-2 : Les sous-classes des fonctions d'organisation du discours.....	74
Tableau 3-3 : Les fonctions principales intégrées par l'Academic Phrasebank.....	76
Tableau 3-4 : 12 Fonctions identifiées pour les routines sémantico-rhétoriques dans les écrits scientifiques.....	77
Tableau 3-5 : Les typologies lexicales du métadiscours de Vandekoppe (1985) et de Crismore et al. (1993).....	91
Tableau 3-6 : Typologie du métadiscours de Hyland (2005, p. 49).....	93
Tableau 3-7 : Modèle réflexif du métadiscours d'Ädel (2006, p38).....	96
Tableau 3-8 : Fonctions discursives du métadiscours dans les écrits académiques en anglais (Ädel, 2006).....	98
Tableau 4-1 : Répartition en disciplines du corpus TermITH (reproduction de Hatier, 2016, p. 58).....	109
Tableau 4-2 : Requêtes proposées par le Lexicoscope.....	115
Tableau 4-3 : Définitions dans deux classes sémantiques.....	122
Tableau 4-4 : Les mots les plus fréquents dans le Scientext.....	123
Tableau 4-5 : Verbes ajoutés dans le répertoire du LST pour l'extraction des routines	124
Tableau 4-6 : Adjectifs ajoutés dans le répertoire du LST pour l'extraction des routines.....	124
Tableau 4-7 : Routines de présentation des objectifs (Chevalier, 2019).....	139
Tableau 4-8 : Exemples de patrons lexico-syntaxiques dans l'introduction des écrits scientifiques en anglais.....	140
Tableau 4-9 : Expressions de la fonction 'background knowledge' reprises de Sandor (2007, p. 103–104).....	145
Tableau 4-10 : Correspondance entre la description sémantique et la construction syntaxique de la routine de planification.....	151
Tableau 4-11 : Démonstration des réalisations lexicales d'une routine de guidage(planification de la structure discursive).....	153
Tableau 4-12 : Un exemple de la modélisation adaptée dans notre recherche.....	154
Tableau 5-1 : Routine N° 1 de planification de la structure discursive.....	167
Tableau 5-2 : Routine N° 2 de planification de la structure discursive.....	167
Tableau 5-3 : Routine N° 1 d'annonce de l'objectif/l'objet d'étude.....	175
Tableau 5-4 : Routine N° 2 d'annonce de l'objectif/l'objet d'étude.....	175
Tableau 5-5 : Routine N° 3 d'annonce de l'objectif/l'objet d'étude.....	176
Tableau 5-6 : Routine N° 4 d'annonce de l'objectif/l'objet d'étude.....	176
Tableau 5-7 : Routine N° 5 d'annonce de l'objectif/l'objet d'étude.....	177
Tableau 5-8 : Routine N° 6 d'annonce de l'objectif/l'objet d'étude.....	177
Tableau 5-9 : Routine N° 1 de navigation textuelle	185
Tableau 5-10 : Routine N° 2 de navigation textuelle	186
Tableau 5-11 : Routine N° 3 de navigation textuelle	186
Tableau 5-12 : Routine N° 4 de navigation textuelle	187
Tableau 5-13 : Routine N° 5 de navigation textuelle	187

Tableau 5-14 : Routine N° 6 de navigation textuelle	188
Tableau 5-15 : Routine N° 7 de navigation textuelle	188
Tableau 5-16 : Routine N° 8 de navigation textuelle	189
Tableau 5-17 : Routine N° 1 de conclusion	194
Tableau 5-18 : Routine N° 2 de conclusion	194
Tableau 6-1 : Routine N° 1 de reformulation	205
Tableau 6-2 : Routine N° 2 de reformulation	206
Tableau 6-3 : Routine N° 3 de reformulation	206
Tableau 6-4 : Routine N° 1 de réparation	211
Tableau 6-5 : Routine N° 2 de réparation	211
Tableau 6-6 : Routine N° 1 de précision terminologie (définition)	214
Tableau 6-7 : Routine N° 2 de précision terminologie (définition)	215
Tableau 6-8 : Routine N° 3 de précision terminologie (dénomination)	215
Tableau 6-9 : Routine N° 4 de précision terminologie (dénomination)	216
Tableau 6-10 : Routine N° 5 de précision terminologie (dénomination)	216
Tableau 6-11 : Routine N° 6 de précision terminologie (désignation)	217
Tableau 6-12 : les concepts X à définir et les énoncés définitoires Y dans les exemples 11-15	218
Tableau 6-13 : les marqueurs identifiés avec les routines de précision terminologique	221
Tableau 6-14 : Routine N° 1 d'emprunt terminologie	225
Tableau 6-15 : Routine N° 2 d'emprunt terminologie	225
Tableau 7-1 : Routine N° 1 de mise en évidence	232
Tableau 7-2 : Routine N° 2 de mise en évidence	232
Tableau 7-3 : Routine N° 3 de mise en évidence	233
Tableau 7-4 : Routine de (co)constat	238
Tableau 7-5 : Routine N° 4 de mise en évidence	242
Tableau 7-6 : Routine N° 2 de focus	242
Tableau 8-1 : Positions textuelles particulières de certaines routines de navigation textuelle	252
Tableau 8-2 : Récapitulation des positions textuelles des routines de guidage	257
Tableau 8-3 : Récapitulation des pronoms personnels et des verbes dans les routines de guidage	263
Tableau 8-4 : Récapitulation des rôles d'auteur dans les 15 routines de guidage avec pronom personnel	264
Tableau 8-5 : Renvois des pronoms personnels dans les routines	265
Tableau 8-6 : Fréquences des fonctions de Hyland (2005, p. 92) et d'Ädel (2006, p. 79) citées dans notre étude	268
Tableau 8-7 : Comparaison entre les marqueurs de reformulation (Tran, 2014, p.298) et les routines métalinguistiques	274
Tableau 8-8 : Distribution des fonctions des routines de guidage	277
Tableau 8-9 : Fréquences relatives pour 100 000 mots des routines dans les disciplines	284
Tableau 8-10 : Fréquences relatives par 1 000 000 mots des routines dans les disciplines ...	290
Tableau 8-11 : Fréquences relatives par million de mots de deux routines dans l'histoire et la géographie	290

Introduction

La présente recherche s'intéresse aux expressions préconstruites dans le genre des écrits scientifiques académiques. Elle s'inscrit dans l'Action « Corpus académiques, lexique scientifique transdisciplinaire et phraséologique »¹ de l'axe 1 du laboratoire Lidilem à l'Université Grenoble Alpes. À travers l'utilisation des outils en traitement automatique des langues (TAL) et la description linguistique du lexique scientifique et des expressions phraséologiques dans des corpus scientifiques transdisciplinaires, cette action vise à approfondir des réflexions linguistiques et didactiques sur le discours scientifique à l'écrit et à l'oral. En linguistique, l'étude phraséologique sur le discours scientifique vise un double objectif : d'une part, elle consiste à approfondir la réflexion sur la modélisation des routines, d'autre part, l'étude des expressions phraséologiques doit s'associer à la réflexion sur les fonctions rhétoriques et discursives du discours scientifique. Dans ce cadre, notre travail porte spécifiquement sur les routines de guidage du lecteur dans le discours académique à l'écrit, en particulier dans les articles de recherche.

L'objectif de cette thèse est d'élaborer une typologie fonctionnelle et d'établir une modélisation linguistique des routines de guidage du lecteur. La typologie devrait permettre de classer les routines de guidage du lecteur identifiées dans le corpus à l'aide de critères rhétoriques, lexico-syntaxiques et sémantiques. La modélisation vise à décrire de façon systématique les caractéristiques linguistiques des routines, en synthétisant leurs essentiels.

Dans le but de mieux comprendre l'intérêt de notre étude, nous proposons d'abord d'effectuer un tour d'horizon des études de la phraséologie, notamment dans les écrits scientifiques en français. Une description plus détaillée de l'objectif et de l'intérêt de notre étude sera ensuite développée. Par la suite, les domaines en lien avec la thèse seront brièvement décrits afin de situer notre recherche. Nous finirons par une brève présentation du plan de la thèse.

1. Le contexte de l'étude : la phraséologie scientifique en anglais et en français

Notre étude s'inscrit dans le domaine de la phraséologie qui a connu depuis 20 ans une extension des objets d'étude et des domaines associés. La phraséologie n'est plus vue comme

¹ Pour plus d'informations, voir le site : <https://lidilem.univ-grenoble-alpes.fr/thematiques-recherche/actions-recherche/corpus-academiques-lexique-scientifique-transdisciplinaire-et-phraséologie>.

une branche de la lexicologie qui étudie les phénomènes contraints d'une langue comme les collocation ou les expressions figées. Elle s'intéresse aujourd'hui aux séquences ou aux formules préconstruites plus longues dans le discours. L'intérêt pour ces séquences préfabriquées a conduit les linguistes à articuler la phraséologie et l'analyse du discours, ce qui a permis aux études portant sur les motifs, les routines discursives, les routines sémantico-rhétoriques de voir le jour. Ces objets linguistiques sont des séquences sémantiquement compositionnelles avec des fonctions étroitement liées aux caractéristiques du genre discursif (Sitri & Tutin, 2016, p. 5). Outre l'analyse du discours, la phraséologie fait appel à la linguistique de corpus et au TAL pour adopter une approche plus empirique afin d'identifier des unités phraséologiques dans les corpus numériques. Le croisement entre la phraséologie, l'analyse du discours, la linguistique de corpus et le TAL constitue une interface interdisciplinaire où les études phraséologiques se diversifient, mais génèrent aussi des questions spécifiques sur la délimitation et la reconnaissance de ces unités phraséologiques émergentes, ainsi que leurs modélisations. L'étude de la phraséologie dans le discours scientifique se situe au croisement de ces différents domaines.

La recherche sur la phraséologie scientifique n'est pas récente. Il existe de nombreuses études portant sur la langue anglaise, en particulier pour les domaines du *English Academic Writing* (EAW) ou du *English for Academic Purposes* (EAP) remontant aux années 90. Howarth (1998), Granger (1998), Granger et Meunier (2008) ont tous souligné l'importance des phénomènes phraséologiques dans les études linguistiques et didactiques concernant l'écrit académique en anglais (EAW), de la même façon que les travaux de Pecman (2004) et ceux de Tutin (2007, 2014), Tutin et Kraif (2016) et Tran (2014) sur l'écrit scientifique en Sciences Humaines et Sociales en langue française. Ces études mettent en lumière la présence de séquences polylexicales ou de séquences préfabriquées facilitant le processus d'écriture et de lecture. En même temps, elles montrent que les écrits scientifiques sont dans une certaine mesure un genre de discours codé conventionnellement entre des membres d'une communauté discursive. Non seulement la présence importante des expressions phraséologiques a été soulignée, mais la diversité de ces expressions a également fait l'objet d'études dans de nombreux travaux portant sur la phraséologie scientifique.

Les études sur le discours scientifique en anglais ont produit une grande variété terminologique : des unités phraséologiques étudiées dans la phraséologie traditionnelle comme des collocations, des locutions, ou des phénomènes phraséologiques plus variés et complexes pouvant correspondre à des séquences courtes *in other words*, mais aussi à des propositions

longues *I don't want to*, comme on peut l'observer dans les études sur les *lexical bundles* (Biber & Barbieri, 2007), les *p-frames* (Römer, 2010) et les *lexical frames* (Gray & Biber, 2015), etc. Dans les études sur le discours scientifique en français, l'intérêt porte particulièrement sur la phraséologie transdisciplinaire à laquelle notre laboratoire a consacré plusieurs projets de recherche. On peut se référer au projet ANR Scientext² (2007-2010) ayant pour l'objectif la constitution d'un corpus des écrits scientifiques et la description du raisonnement et du positionnement de l'auteur à travers l'observation de la phraséologie scientifique transdisciplinaire (Grossmann & Tutin, 2013, p. 15), et au projet ANR TermITH³ qui a permis d'identifier des éléments phraséologiques (nous y reviendrons dans la partie II). Ces études se sont intéressées à des séquences polylexicales diverses allant des collocations jusqu'à des expressions plus longues et complexes comme les routines.

2. Le thème, les objectifs et l'intérêt de l'étude

Dans notre étude, nous nous intéressons à un type de phénomènes phraséologiques dans les écrits scientifiques : les routines sémantico-rhétoriques, plus précisément les routines à fonction de guidage du lecteur. Ce sont des expressions préfabriquées avec des variations lexicales et syntaxiques, comme [article|travail |étude] [s'inscrit|s'appuie|se base|inspire|prend appui] [dans|sur], il est [important|essentiel|central|fondamental] de [noter|constater|relever|voir|mentionner|montrer]. Le terme « routines sémantico-rhétoriques » trouve son origine dans la réflexion de Tutin (2014) sur la typologie des phénomènes phraséologique dans les écrits scientifiques en langue française. Notre travail se base en premier lieu sur cette typologie phraséologique qui différencie notre objet d'étude des autres types de phénomènes phraséologiques dans les écrits scientifiques comme les collocations ou les locutions. Selon cette étude, quatre grands types des unités phraséologiques sont identifiés dans les écrits scientifiques en français (Tutin, 2014) : des séquences polylexicales à fonction référentielle (*formuler une hypothèse*), à fonction discursive (*en résumé, par conséquent*), à fonction interpersonnelle (*contre toute attente*) et les expressions récurrentes plus larges et complexes (*nous voudrions souligner que*), nommées « routines sémantico-rhétoriques ».

Dans le cadre de ce travail, nous nous intéressons plus précisément à ce dernier type de phraséologisme, notamment ceux qui ont pour fonction le guidage du lecteur, parce qu'ils ont

² <https://anr.fr/Projet-ANR-06-CORP-0020>

³ <https://anr.fr/Projet-ANR-12-CORD-0029>

été moins étudiés, mais qu'ils sont centraux pour comprendre certaines caractéristiques rhétoriques des écrits scientifiques. Prises dans le champ de la phraséologie étendue, ces routines se trouvent à la frontière du lexique et de la phrase (Legallois & Tutin, 2013, p. 3). Leur rôle est de permettre aux auteurs de guider la lecture tout au long du texte et de faciliter la compréhension du texte.

Nous souhaitons, à travers cette recherche, recenser les routines de guidage du lecteur au moyen du corpus TermITH-transdisciplinaire (nous y revenons au chapitre 4), en proposant une typologie fonctionnelle accompagnée d'une modélisation linguistique, découvrir les stratégies rhétoriques mises en œuvre par ces routines de guidage pour enfin tracer les caractéristiques énonciatives et rhétoriques des écrits scientifiques en français. Notre étude consiste non seulement à fournir une étude systématique des routines de guidage – un type de phraséologisme encore peu étudié dans les écrits scientifiques, mais aussi à compléter les études précédentes dans une perspective énonciative par une approche phraséologique, car ces études ont été habituellement menées à travers des marques énonciatives plus simples comme les pronoms liés à l'auteur ou des marques déictiques. Bien que les routines soient transdisciplinaires, nous mènerons également quelques observations sur les modes d'écritures dans les disciplines, ce qui permet de fournir une perspective disciplinaire à notre étude.

De ce fait, plusieurs questions se posent dans cette thèse :

- 1) Comment caractériser les routines de guidage du lecteur dans les écrits scientifiques ?
- 2) Comment les identifier dans le corpus ?
- 3) Quelles fonctions rhétoriques propres au discours académique peut-on associer à ces routines ?
- 4) Comment modéliser ces routines de guidage ?
- 5) Quelles sont les similitudes et les différences entre les disciplines en sciences humaines et sociales ?

Notre étude vise trois objectifs principaux :

- 1) Une définition plus précise des routines de guidage et l'approfondissement de leurs caractéristiques linguistiques ;
- 2) L'élaboration d'une typologie fonctionnelle ainsi que les critères de classement pour classer les routines de guidage ;
- 3) L'élaboration et l'approfondissement d'une modélisation linguistique permettant de décrire systématiquement les routines de guidage.

Cette étude nécessite des connaissances interdisciplinaires qui concernent trois domaines principaux : la phraséologie, l'analyse du discours et la linguistique de corpus.

3. Les domaines impliqués de notre étude

Notre étude sur l'identification, la modélisation linguistique et la description des routines de guidage se situe au croisement de trois domaines de recherche : la phraséologie, l'analyse du discours et la linguistique de corpus.

Le premier domaine est celui de la description phraséologique. D'abord, il faut noter que les routines de guidage du lecteur ne font pas l'objet de descriptions détaillées dans les ouvrages de phraséologie traditionnelle. Leur donner une définition et les différencier nécessite une observation minutieuse impliquant nombre de notions connexes, comme les *segments répétés*, les *motifs*, les *lexical bundles*, les *routines conversationnelles*, etc.

Le deuxième domaine abordé dans cette recherche est l'analyse du discours, envisagé en particulier dans l'approche énonciative et l'approche fonctionnelle. Comme l'élaboration d'un modèle analysable et d'une typologie fonctionnelle pour des routines de guidage du lecteur dans les écrits scientifiques constitue le but central de notre recherche, nous nous référons d'abord à l'approche énonciative pour comprendre le genre discursif des écrits scientifiques en matière de traits dialogiques, tels qu'ils ont été développés en détail par le courant de la praxématique (Bres, 2005, 2017). Une autre approche fonctionnelle – le métadiscours, a également été exploitée dans le but de mieux diagnostiquer les fonctions rhétoriques des données recueillies.

Le dernier domaine concerne la linguistique de corpus. Les linguistes ont davantage recours aux corpus pour éviter la seule intuition dans leurs recherches en linguistique. Grâce aux innovations informatiques, les linguistes gagnent ainsi en objectivité. L'étude sur les marqueurs phraséologiques tire également bénéfice de l'approche sur corpus, dans laquelle s'inscrit notre recherche. Notre étude utilisera le corpus TermITH-transdisciplinaire en s'appuyant sur des outils d'exploration des usages phraséologiques faisant également appel aux méthodes et outils du traitement automatique des langues (TAL).

4. Le plan de la thèse

La présente étude comporte trois grandes parties. La première partie, qui se compose de trois chapitres, est consacrée à la définition du cadre théorique de notre travail, en faisant appel à des approches et notions centrales. Nous commencerons par esquisser dans le Chapitre 1 le contour des routines de guidage du lecteur dans les écrits scientifiques, et dans les deux

chapitres qui suivent, nous expliquerons les deux approches linguistiques que nous avons adoptées afin d'inscrire l'étude des routines de guidage dans une réflexion théorique plus approfondie. Nous évoquerons dans le Chapitre 2 en particulier le concept de dialogisme en examinant le fonctionnement interactif des écrits scientifiques, qui est la motivation la plus profonde de l'emploi de ces routines dans le genre en question. Par la suite, nous introduirons une approche fonctionnelle dans le Chapitre 3, en abordant plus précisément les sources théoriques et empiriques du métadiscours, une approche qui permet de diagnostiquer des fonctions susceptibles de guider la lecture.

La deuxième partie porte sur la méthodologie d'étude, avec un seul chapitre. Ce chapitre 4 présentera la description de notre corpus, du matériel utilisé, des outils d'exploitation du corpus, des méthodes d'extraction des candidats routines et des critères de sélection. L'objectif est d'approfondir ensuite le modèle d'analyse des routines de guidage dans les écrits scientifiques.

La troisième partie se concentre sur l'élaboration d'une typologie fonctionnelle et les analyses qualitatives et quantitatives des routines de guidage du lecteur dans les écrits scientifiques. Elle se compose de quatre chapitres dont les trois premiers chapitres (5, 6 et 7) se focaliseront sur l'analyse qualitative des caractéristiques linguistiques des routines de guidage en leur proposant des modélisations associées. Le dernier chapitre (8) discutera de façon synthétique les régularités et les singularités des routines à travers les positions textuelles et les disciplines. L'inclusion énonciative et les rôles d'auteur seront aussi examinés dans ce chapitre.

En guise de bilan, nous présenterons les apports, les perspectives, ainsi que les limites de cette étude dans la conclusion. Nous reviendrons sur les résultats les plus significatifs et ouvrirons des pistes à suivre dans le futur.

Partie I. Aborder les routines de guidage du lecteur des écrits scientifiques dans les perspectives phraséologique, énonciative et fonctionnelle

Les écrits scientifiques présentent des caractéristiques de préfabrication lexico-syntaxique associées aux façons de dire routinisées dans la communauté académique, à laquelle les apprentis scripteurs cherchent à s'intégrer. En observant ces expressions préfabriquées, on peut relever des expressions récurrentes qui servent à faciliter la lecture du texte et que l'on appelle « routines de guidage du lecteur ». Ces routines de guidage du lecteur relèvent de la phraséologie scientifique transdisciplinaire et intègrent des fonctions rhétoriques spécifiques de nature métadiscursive dans le discours scientifique. Leur présence ne vise pas à la transmission de contenus informatifs, mais à faciliter la compréhension des contenus transmis lors de la communication. Du point de vue énonciatif, ces routines de guidage reflètent le fonctionnement interactif des écrits scientifiques – en particulier le fonctionnement dialogique. Du point de vue rhétorique, elles peuvent remplir plusieurs fonctions typiques dans ce genre académique, permettant de mieux comprendre les stratégies rhétoriques couramment mobilisées par l'auteur expert. Par conséquent, leur étude nécessite une approche triple : la mise en lumière des caractéristiques interactives et dialogiques des écrits scientifiques, la délimitation de la notion des routines dans la perspective phraséologique combinée avec l'analyse du discours, l'inspection d'un point de vue rhétorique des fonctions métadiscursives dans les écrits scientifiques.

Dans cette partie théorique, nous proposons une synthèse intégrant plusieurs modèles linguistiques qui guident notre recherche. Nous abordons d'abord la notion des routines de guidage du lecteur dans les écrits scientifiques en les observant à l'aide des notions connexes dans plusieurs domaines comme la phraséologie et l'analyse du discours. Comme la question des routines de guidage du lecteur est étroitement liée à la question du genre discursif, nous proposons par la suite d'observer les caractéristiques du fonctionnement interactif des écrits

scientifiques en inscrivant l'étude des routines de guidage dans l'approche énonciative de l'analyse du discours, en particulier le dialogisme. Pour finir, nous introduirons le modèle du métadiscours afin de donner une vue d'ensemble des fonctions métadiscursives fréquentes dans les écrits scientifiques en les associant aux marqueurs/phrases linguistiques. Cela permettra de concevoir notre propre typologie fonctionnelle des routines de guidage du lecteur.

Chapitre 1. Routines de guidage du lecteur dans la phraséologie étendue : précisions terminologiques

La phraséologie transdisciplinaire scientifique comporte des ensembles divers allant des collocations comme *résultat intéressant, formuler une hypothèse* (Tutin, 2007) et des séquences polylexicales de fonctions discursives *contrairement à, en d'autres termes* (Tran, 2014) aux expressions plus longues qui se situent à la frontière entre le lexique et la phrase et constituent une zone étendue de la phraséologie (Legallois & Tutin, 2013), que nous appelons « routines sémantico-rhétoriques » (Tutin, 2014 ; Tutin & Kraif, 2016). L'objet d'étude de cette thèse, les routines de guidage du lecteur relèvent des routines sémantico-rhétoriques qui possèdent certaines caractéristiques linguistiques communes. Ces routines sémantico-rhétoriques, en tant qu'expressions récurrentes, ne répondent pas vraiment au principe de figement et ne font donc pas l'objet d'études dans la phraséologie traditionnelle.

Ces expressions soulèvent un certain intérêt dans le monde francophone dans la perspective d'analyse du discours (Legallois & Tutin, 2013; Longrée & Mellet, 2013; E. Née, Sitri, & Veniard, 2014, 2016; Luodonpää-Manni, Grossmann, & Tutin, 2022) ou de caractérisation des genres textuels (Kraif, Novakova, & Sorba, 2016; Novakova & Sorba, 2017; Bendinelli, 2017; Sitri & Veniard, 2017), comme dans le monde anglophone, où elles sont étudiées à la marge des formules préfabriquées dans la perspective de l'enseignement de l'anglais en tant que langue seconde (Nattinger & DeCarrico, 1992; Lewis, 1996) ou de l'enseignement de l'*Academic English* ainsi que dans l'analyse des registres (Biber, 2006, 2009; Ellis, Simpson-Vlach, & Maynard, 2008; Römer, 2010; Gray & Biber, 2015). Bien sûr, ces expressions récurrentes, sous des désignations variables selon les études, sont proches de plusieurs notions connexes comme les routines, les motifs, les segments répétés, les paquets lexicaux (*lexical bundle*), les *lexical phrases*, les *p-frames*, etc., que nous aborderons plus tard.

Dans ce chapitre, nous décrirons d'abord brièvement la phraséologie étendue qui concerne notre objet d'étude ; nous entrerons ensuite dans l'observation des caractéristiques linguistiques des routines sémantico-rhétoriques en rappelant certaines notions connexes ; nous délimiterons par la suite les caractéristiques linguistiques des routines sémantico-rhétoriques en montrant les paramètres à prendre en compte. Enfin, nous expliquerons la relation entre les routines de guidage du lecteur et les routines sémantico-rhétoriques et donnerons ainsi une définition précise des routines de guidage du lecteur.

1.1. La phraséologie étendue : une extension des objets d'étude et des domaines impliqués

La phraséologie constitue aujourd'hui un domaine qui comporte de nombreux objets d'études et s'inscrit également dans une perspective interdisciplinaire. L'extension du domaine phraséologique donne lieu à une prolifération de termes et de critères définitoires pour des faits phraséologiques. Granger & Paquot (2008, p. 27) notent que les frontières de ce domaine en plein essor sont floues, et distinguent deux approches principales dans les recherches phraséologiques, qu'on appelle « approche phraséologique » et « approche basée sur les corpus »⁴ en français. Il nous semble nécessaire de faire un bref état de l'art de ces deux approches pour délimiter les domaines où s'inscrit notre étude sur les routines de guidage. De manière générale, ces deux approches se distinguent en fonction de leurs critères définitoires. L'approche phraséologique est plus traditionnelle⁵, et distingue les expressions phraséologiques des expressions libres par des principes de figement ou de contrainte (*ibid.*). C'est l'approche phraséologique traditionnelle en langue française, comme celle qui est adoptée dans la Théorie Sens-Texte de Mel'čuk (2012). L'approche basée sur la fréquence ou la cooccurrence est plus répandue dans le monde anglo-saxon, notamment dans les études effectuées par des linguistes britanniques comme J.-R. Firth ou J. Sinclair (Granger & Paquot, 2008, p. 28). Ces deux approches ne sont pas totalement opposées l'une à l'autre, et peuvent être vues comme complémentaires. En effet, on constate que de nombreuses recherches combinent les deux approches.

Ancrée dans la lexicologie et la sémantique, l'approche phraséologique s'intéresse généralement aux unités phraséologiques définies souvent par un principe de figement qui se réalise sur plusieurs plans (Legallois & Tutin, 2013) :

- (a) Sur le plan syntaxique : la fixité de la structure ;
- (b) Sur le plan sémantique : la non-compositionnalité ;
- (c) Sur le plan lexical : la restriction de la substitution lexicale ;
- (d) Sur le plan énonciatif : la non-actualisation des référents.

Les études qui entrent dans ce courant se concentrent davantage sur la description des caractéristiques linguistiques des expressions phraséologiques dans des perspectives

⁴ En anglais, c'est "the phraseological approach and the distributional or frequency-based approach" (Granger & Paquot, 2008, p. 27).

⁵ Cette approche est traditionnelle en Europe, mais certains l'appellent également l'approche continentale (Williams, 2003).

syntaxiques, sémantiques et lexicologiques. La question centrale ici demeure le figement formel ou sémantique (ou non-compositionnalité sémantique) des constituants des phénomènes phraséologiques (*ibid.*). Avec ces critères, ce courant s'intéresse tout particulièrement aux mots composés, aux collocations et aux expressions figées. Mel'čuk (2012), de son côté, a identifié trois grandes classes des phrasèmes en appliquant des contraintes de nature sémantico-lexicale : les idiomes, les collocations et les clichés. Les idiomes manifestent la double contrainte sur les plans lexicaux et sémantico-lexicaux, tandis que les collocations et les clichés sont compositionnels, mais avec une contrainte de sélection lexicale. Par rapport aux critères de figement qui concernent à la fois les plans lexical, syntaxique et sémantique dans la tradition française, la définition des phrasèmes chez Mel'čuk autorise l'inclusion d'éléments phraséologiques comme les collocations et les clichés, ce qui élargit déjà la sphère d'objet d'étude de la phraséologie. Néanmoins, le critère définitoire repose encore sur une forme de contrainte lors du choix des expressions.

Ce petit exemple nous permet de voir que les objets d'études de la phraséologie traditionnelle se concentrent sur une série de phénomènes phraséologiques qui regroupe des expressions non libres qui manifestent au moins une contrainte, soit sur le plan lexical, soit sur les plans syntactique et sémantique. Néanmoins, l'étude de la phraséologie a connu une extension remarquable depuis une trentaine d'années. Les études phraséologiques de langue française dépassent la limite du figement et recourent également aux critères statistiques et fonctionnels utilisés davantage dans les études de la phraséologie sur la langue anglaise dès les années 1990 (Gray & Biber, 2015, p. 125). Elles peuvent s'inscrire par conséquent dans l'approche basée sur les usages et la linguistique de corpus, comme l'utilisation de la fréquence des cooccurrences.

Cette dernière approche privilégie la recherche basée sur de grands corpus et adopte souvent une méthodologie *corpus-driven*, afin de mettre en évidence des chaînes de mots sélectionnés mutuellement. À titre d'exemple, Sinclair *et al.* (2004, p. 10)⁶ proposent une définition statistique des collocations, qui se caractérisent selon eux comme la cooccurrence de deux unités qui se trouvent souvent ensemble. De nombreuses études phraséologiques sur la langue française ont intégré également ce critère quantitatif, même s'il n'est pas dominant. Nous trouvons, par exemple, des travaux sur les collocations dans les écrits scientifiques (Tutin, 2007 ; Cavalla, 2008, 2015 ; Cavalla & Tutin, 2008), mais aussi des travaux qui s'inscrivent

⁶ La première version de cet ouvrage a été publiée en 1970, nous avons utilisé la version rééditée et publiée en 2004.

davantage dans l'analyse du discours comme les études des segments répétés dans le discours politique (Salem, 1987), des motifs dans les discours littéraires (Legallois, 2009 ; Legallois, Longrée & Mellet, 2013 ; Charnois, & Poibeau, 2016 ; Novakova & Sorba, 2017). Ces études phraséologiques ont ajouté au critère de cooccurrences d'autres critères, comme les fonctions discursives, pour les *lexical bundles*, les *lexical phrases*, les *motifs*, ou le critère énonciatif pour les *routines*. Par conséquent, une extension remarquable des phénomènes linguistiques a été observée dans le domaine de la phraséologie : de façon non exhaustive, plus de soixante catégories ont été recensées (Wray, 2000, cité par Wahl & Gries, 2018, p. 88).

En prenant compte de l'abondance des éléments phraséologiques, Gries (2008, p. 4) a résumé six paramètres typiques dans les recherches phraséologiques qui permettent de les définir avec des critères plus flexibles :

- (1) la nature des éléments impliqués dans le phraséologisme
- (2) le nombre des éléments impliqués dans le phraséologisme
- (3) le nombre d'occurrences observées dans le corpus pour qu'il soit inclus en tant que phraséologisme
- (4) la distance tolérée entre les éléments d'un phraséologisme
- (5) le degré de flexibilité syntaxique et lexicale des éléments d'un phraséologisme
- (6) la compositionnalité ou la non-prédictibilité sémantique

Pour résumer, ces six paramètres qui peuvent être réalisés à des degrés divers permettent d'élargir le champ des objets phraséologiques, notamment pour des configurations de cooccurrence plus flexibles (*ibid.*, p. 10), comme notre objet d'étude, les routines de guidage du lecteur.

Observons de plus près ces six critères.

Le premier prête attention à la nature linguistique des éléments inclus dans une expression phraséologique. Cela peut être des items lexicaux (*strong tea*) ou des patrons grammaticaux, ou les deux en même temps (*ibid.*, p. 5). Cela permet d'inclure des expressions plus complexes comme des *lexical bundles*, des routines, des expressions idiomatiques, etc.

Le deuxième paramètre n'est plus un problème épineux aujourd'hui, car on ne restreint pas le nombre de mots dans des expressions phraséologiques à deux éléments, mais on accepte

des expressions plus longues, comme des segments répétés qui comportent souvent 2 à 5 mots, voire plus.

Le troisième paramètre consiste à observer des faits phraséologiques à partir des critères statistiques. Il s'agit d'un critère de répétition des expressions dans des corpus. Cependant, l'exploitation de ce critère aboutit à deux méthodes différentes : la première s'appuie sur la fréquence absolue dans le corpus étudié dont le seuil est souvent fixé arbitrairement par des linguistes. Bien qu'il soit jugé moins fiable par Gries (2008, p. 5), la plupart des études préliminaires en anglais utilisent la fréquence absolue ou relative (le pourcentage d'occurrences dans le corpus) comme pour les *lexical bundles*. La deuxième méthode exige que la fréquence de cooccurrences soit supérieure à la fréquence escomptée de manière aléatoire pour délimiter des phénomènes non imputables au seul hasard (avec un certain seuil de probabilité), qui pourraient ainsi entrer dans la sphère phraséologique.

Ce critère de fréquence constitue un paramètre définitoire central dans nombreuses études phraséologiques comme celles de Sinclair (avec le projet *Cobuild*), mais l'appui sur ce seul critère n'est pas fiable. Il pourrait, d'une part, extraire des chaînes de mots assez fréquentes, mais dépourvues de sens autonome, comme dans l'anglais *of the*, dans le français *et de la*. Ces chaînes de mots ont en effet une fréquence très élevée, mais ne sont pas des expressions phraséologiques. D'autre part, il ne peut pas prendre en compte d'autres facteurs comme le sens de l'expression, les connotations pragmatiques et contextuelles. Wray (2002, p. 31) illustre les limites de l'emploi exclusif du critère fréquentiel par l'expression *keep your hair on* qui n'est pas une unité phraséologique à moins qu'elle ne signifie *calm down* (calmez-vous) au lieu de *don't move your wig* ('ne bougez pas votre perruque'). La fréquence absolue recense l'expression sans différencier ces deux sens. Pour compenser cette lacune, plusieurs études se servent de corpus de contraste, dans le dessein d'identifier des faits phraséologiques spécifiques en menant des analyses qualitatives.

La distance entre des éléments composant une expression phraséologique constitue également un point discutable. Dans de nombreuses études, notamment celles basées sur les *n-grams* ou *cluster*, les composants sont contigus et aucune distance n'est tolérée. C'est le cas d'expressions comme les locutions figées *en résumé*, *dans l'intention de* qui s'emploient souvent en tant qu'expressions lexicalisées. Néanmoins, il faut noter que les expressions phraséologiques ne sont pas nécessairement contiguës. Certaines sont discontinues et peuvent inclure des *slots* variables, comme les *motifs*, les *routines* ou les *p-frames* (p.ex. *is _ to*). Nous reviendrons plus tard sur ce point.

Le cinquième paramètre concerne particulièrement la flexibilité sur les plans syntaxiques et lexicaux des éléments d'un phraséologisme. Cette souplesse permet d'étendre les expressions phraséologiques à des schémas plus larges qui pourraient regrouper des expressions du même champ sémantique et de la même configuration syntaxique. Par exemple, Nattinger & DeCarrico (1992, p. 45) distinguent les *phrasal constraints* (*the _er, the _er, what with _*) et les *sentence builders* (*that reminds me of X, it seems to me (that) X*) qui sont en mesure de tolérer des variables lexicales et syntaxiques (au niveau des propositions subordonnées). Ici, il faut faire la différence entre des expressions phraséologiques qui intègrent des *slots* (souvent, un paradigme lexical) et les constructions grammaticales autorisant des paradigmes lexicaux. En fait, une expression phraséologique doit contenir au moins un item lexical spécialisé qui est fixe, alors que cette condition n'est pas nécessaire pour les constructions grammaticales. Par exemple, la construction grammaticale Verbe transitif + Objet direct + Objet indirect ne présente aucun mot fixe. Nous postulons qu'une telle construction grammaticale ne doit pas être considérée en tant qu'unité phraséologique.

Le dernier paramètre renvoie au caractère compositionnel du sens. Les unités phraséologiques peuvent correspondre à des expressions dont le sens n'est pas compositionnel comme les locutions figées, les expressions idiomatiques, ou les clichés. Mais elles peuvent également être compositionnelles et transparentes. Dans ce cas, le sens des expressions phraséologiques est prédictible. Ainsi, la non-compositionnalité cesse de constituer un critère exclusif pour reconnaître des phénomènes phraséologiques. Cette concession ne signifie cependant pas que l'investigation sur l'aspect sémantique des unités phraséologiques est moins pertinente. En revanche, les séquences phraséologiques doivent être à la fois compositionnelles et utilisables en tant qu'unités fonctionnelles. L'exemple des *sentence builders* cité plus haut représente un cas d'expressions ayant un sens unique et indépendant, mais qui sont tout à fait compositionnel sur le plan sémantique. D'une part, le critère d'autonomie de l'unité sémantique permet d'exclure des chaînes de mots très fréquentes, mais dépourvues de sens comme cité plus haut *of the, et de la*. D'autre part, l'ouverture à la compositionnalité sémantique permet également d'inclure davantage d'expressions polylexicales comme les routines.

Ces six paramètres ne sont bien sûr pas obligatoires pour toutes les unités phraséologiques, mais proposent une définition plus systématique pour délimiter les objets d'étude dans le domaine de la phraséologie. En dehors de ces six paramètres principaux, d'autres paramètres peuvent également s'intégrer dans la définition d'un type d'expressions phraséologiques. Prenons l'exemple des *lexical bundles* (Biber, 2006) dont les fonctions

discursives réalisées dans le discours scientifique constituent un critère central et permettent de caractériser le discours scientifique en anglais. En ce sens, la fonction ajoute aussi une dimension d'analyse du discours à l'étude des *lexical bundles*.

Ce simple exemple montre que les études de phraséologie ont tendance à dépasser la limite de la lexicologie et s'intègrent dans des domaines de plus en plus divers. Cela fait écho au constat que Legallois & Tutin (2013, p. 9) ont dressé dans la revue *Langages* N° 189 :

La phraséologie n'est plus considérée comme un sous-domaine privilégié de la lexicologie ; l'analyse textuelle, la syntaxe, la psycholinguistique ont investi de manières diverses ce champ – non plus en notant la marginalité du phénomène phraséologique, mais bien souvent en montrant son importance cruciale dans la complexité des agencements linguistiques.

Plusieurs domaines peuvent être impliqués dans l'étude de la phraséologie. Notons ici la linguistique textuelle, la syntaxe, la psycholinguistique (Legallois & Tutin, 2013), mais aussi plus largement l'analyse du discours, la linguistique contrastive, la linguistique de corpus ou le traitement automatique des langues. Le croisement entre le domaine de la phraséologie et d'autres domaines contribue également à l'extension des critères définitoires des faits phraséologiques, ce qui rend la définition des faits phraséologiques plus complexe (Legallois & Tutin, 2013). Néanmoins, l'avantage est que l'étude de la phraséologie n'est plus limitée aux domaines sémantiques, morphologiques et lexicologiques, mais s'ouvre à des dimensions plus spécifiques, par exemple, la dimension textuelle ou discursive, et aborde les questions centrales des études sur le genre textuel. À titre d'exemple, dans le domaine de la linguistique textuelle, les expressions phraséologiques sont devenues un moyen de structurer le texte, en particulier des locutions discursives comme *d'une part, d'autre part, quant à*, etc.

Nous avons parcouru dans cette partie quelques éléments montrant comment l'extension du domaine de la phraséologie s'est réalisée. L'objet d'étude de cette thèse – les routines de guidage du lecteur se situent dans cette double extension des objets et des domaines.

1.2. Routines sémantico-rhétoriques et notions connexes

Comme nous l'avons déjà indiqué dans la première partie du chapitre, les objets linguistiques qui nous intéressent ne relèvent pas de la phraséologie classique, dans le cadre de la lexicologie, mais se situent à la limite de la phraséologie, et relèvent autrement dit de « la phraséologie étendue ». Nous voudrions d'abord inspecter la place qu'occupent les routines sémantico-rhétoriques dans le domaine de la phraséologie en choisissant précisément quelques notions qui sont proches de notre objet d'étude, dans l'intention de décrire plus précisément les caractéristiques des routines sémantico-rhétoriques. Notre objectif, dans cette section, consiste

à répondre aux questions suivantes : quelle est la nature phrastique de routines sémantico-rhétoriques ? Quel degré de flexibilité les composants des routines assument-ils et sur quels niveaux se manifeste leur flexibilité ? Quels sont les paramètres énonciatifs à prendre en compte dans l'étude des routines sémantico-rhétoriques ? Quel niveau d'abstraction apparaît nécessaire pour décrire les routines ?

Nous organisons cette section de la façon suivante. Nous étudierons d'abord rapidement les caractéristiques des routines sémantico-rhétoriques à partir de quelques exemples concrets, afin d'en donner une première esquisse. Ensuite, nous sélectionnerons en particulier trois ensembles de notions proches (les routines conversationnelles, les notions à la base de *n-grams* dans les études sur l'anglais, les motifs et les routines discursives en français). L'objectif sera de montrer les similitudes ainsi que les différences entre les routines sémantico-rhétoriques et ces notions connexes. Enfin, nous essayons de répondre aux questions posées plus haut en menant une observation plus approfondie des routines sémantico-rhétoriques et en proposant une définition à la fin de cette section.

1.2.1. Quelques observations sur les routines sémantico-rhétoriques

Dans les écrits scientifiques en français, on rencontre souvent des expressions assez longues qui se situent à la frontière du lexique et de la syntaxe. Voyons quelques exemples (les expressions qui nous intéressent sont mises en gras, les éléments facultatifs ou périphériques sont soulignés ou mis en italique) :

- [1] **L'article s'organise donc autour de 3 parties. La première rappelle** l'intérêt de se focaliser sur les relations sémantiques à partir de leurs traces en corpus pour construire des ressources sémantiques comme les terminologies ou les modèles conceptuels. **La deuxième partie fait état de** résultats en linguistique de corpus relatifs à l'étude des relations conceptuelles et qui nous semblent intéressants pour la modélisation des connaissances. Enfin, **dans la troisième partie, nous montrons** comment ces principes linguistiques guident la spécification d'une méthode et d'un outil d'aide au repérage et à la modélisation des relations conceptuelles : le système Caméléon. (Linguistique)
- [2] **J'emprunte l'expression à Denis Laborde** (1999 : 262 - 299), qui l'emprunte lui-même à Lucy Schuman (1990). (Anthropologie)
- [3] **Je mobilise ici la définition goffmanienne** des cadres, à savoir qu'il s'agit de principes organisateurs qui structurent les événements, orientent les activités et permettent aux participants d'accorder du sens à ce qu'ils font. (Science de l'informatique)
- [4] [8] **Nous reprenons le terme « sémiurgie » forgé par** Roland Barthes (1976, p. 180), qui désigne non pas l'étude des signes (sémiologie), mais qui met l'accent sur l'action, la création démiurgique de sens, la fabrication et

l'utilisation de signes en vue de produire certains effets. (Science de l'information)

- [5] **Nous pouvons constater** dans ce texte **que** la prescription s'exprime de manière explicite à travers des verbes à valeur exhortative. (Linguistique).
- [6] Avant de les évoquer brièvement, **je voudrais souligner que** l'immense majorité des emplois prépositionnels manifestent les propriétés évoquées en 1. (Linguistique).
- [7] **On peut** *cependant* **remarquer** sur le tableau 6a **que** lorsqu'on compare les mêmes points de vue issus des deux situations, on n'obtient alors pas de différence significative. (Psychologie).
- [8] À partir des graphiques 1 et 2, **nous pouvons remarquer que** le niveau d'indemnisation de départ joue de manière différente dans les deux réglementations. (Économie).

Ces expressions fréquentes dans notre corpus sont nommées par Tutin (2014) « les routines sémantico-rhétoriques ». Il s'agit des façons de dire routinisées dans les écrits scientifiques (*ibid.*). Ces routines présentent des caractéristiques communes, observées par Tutin dans son étude postérieure en collaboration avec Kraif (Kraif & Tutin, 2017). De façon sommaire, ce sont des expressions récurrentes souvent construites autour d'un prédicat verbal (tous les exemples 1-8) avec la possibilité d'admettre des variantes aux niveaux syntaxique et paradigmatique : l'exemple 1 dispose de deux expressions analogues :

Groupe 1 :	<i>La</i>		<i>première</i>		<i>rappelle</i>
	<i>La</i>		<i>deuxième partie</i>		<i>fait état de</i>

C'est le même cas pour les phrases en gras dans les exemples 2-4 :

Groupe 2 :	<i>J'</i>		<i>emprunte</i>		<i>l'expression à quelqu'un</i>
	<i>Je</i>		<i>mobilise (ici)</i>		<i>la définition goffmanienne</i>
	<i>Nous</i>		<i>repreons</i>		<i>le terme forgé par quelqu'un</i>

Les exemples 5-8 présentent des caractéristiques comparables :

Groupe 3 :	<i>Nous</i>		<i>pouvons</i>		<i>constater (dans ce texte) que</i>
	<i>Je</i>		<i>voudrais</i>		<i>souligner que</i>
	<i>On</i>		<i>peut (cependant)</i>		<i>remarquer (sur le tableau 6a) que</i>
	<i>Nous</i>		<i>pouvons</i>		<i>remarquer que</i>

Nous pouvons constater selon les trois groupes d'expressions identifiées ci-dessus que chaque groupe partage une structure syntaxique identique, en présentant des flexibilités lexicales sur les axes syntaxiques et paradigmatiques. D'abord, ces expressions ne sont pas forcément linéaires, mais peuvent assumer un certain degré de discontinuité par l'incise des éléments argumentatifs ou déictiques comme *cependant*, *ici*, *dans ce texte*, etc. Ensuite, des variabilités lexicales peuvent être observées sur toutes les parties syntaxiques de chaque groupe :

- Groupe 1 : *la première*|*la deuxième partie* pour le sujet, *rappelle*|*fait état de* pour le

prédicat ;

- Groupe 2 : *je|nous* comme sujet ; *emprunter|mobiliser|reprendre* comme prédicat ; *expression|définition|terme* comme objet ; *à quelqu'un|goffmanienne|forgé par quelqu'un* comme modifieur d'objet.
- Groupe 3 : *je|nous|on* comme sujet ; *vouloir|pouvoir* dans la position des verbes modaux ; *constater|souligner|remarquer* comme verbes prédicatifs.

De plus, ces paradigmes lexicaux appartiennent souvent au même champ sémantique. Par exemple : *la première|la deuxième partie* renvoient aux parties textuelles du discours ; *je|nous|on* renvoient à l'auteur du discours ; *emprunter|mobiliser|reprendre* sont tous des verbes d'USAGE, *constater|souligner|remarquer* expriment un constat. En résumé, nous pouvons dire que les routines sont des groupes d'expressions qui partagent une même structure syntaxique avec des variations lexicales apparentées.

En interrogeant les fonctions de ces expressions dans les exemples cités plus haut, on voit qu'elles remplissent des fonctions rhétoriques spécifiques dans les écrits scientifiques en français. Le groupe 1 vise à annoncer le plan textuel du discours. Le groupe 2 remplit une fonction d'emprunt de la terminologie et le groupe 3 permet de formuler un constat. Ces fonctions rhétoriques constituent un facteur essentiel pour l'étude du genre discursif des écrits scientifiques. Plusieurs travaux antérieurs ont déjà mis en évidence les fonctions rhétoriques que les routines peuvent exercer dans les écrits scientifiques. Par exemple, Grossmann & Tutin (2010) ont identifié la fonction évidentielle (*on peut voir sur le tableau 6 b*) qui correspond à une fonction d'observation et de constat *comme nous avons pu le voir (le constater, l'observer, etc.)* (Grossmann, 2014b). Plus récemment, Ji & Tutin (2019, 2020) ont recensé les fonctions de topicalisation pour mettre l'accent sur le thème (*le|la problème|question central(e)|essentiel [le] est*) ou le constat (*je|on|nous [pouvoir|vouloir] constater|souligner|remarquer que*), et les fonctions métalinguistiques comme la définition (*on|nous [pouvoir] définit|concevoir X comme*), la reformulation (*cela veut dire que*). Ces routines n'ont pas qu'une fonction rhétorique : la participation des certains éléments discursifs comme *enfin, cependant* permet aux routines de s'inscrire dans le processus argumentatif des écrits scientifiques, et de ce fait, de garantir la cohérence et la cohésion du texte.

Dans la perspective énonciative, ces expressions s'inscrivent dans le processus d'actualisation énonciative du discours scientifique. L'utilisation des pronoms personnels *je, nous, on* dans les exemples 2-8, des déictiques de lieu textuel comme *ici, dans ce texte, sur le*

tableau 6 a dans les exemples 3, 5, 8, et les noms *la première*, *la deuxième partie* dans l'exemple 1 renvoient également aux lieux textuels du discours en cours.

En résumé, quatre critères peuvent être retenus pour définir de façon générale les routines sémantico-rhétoriques dans les écrits scientifiques (Tutin & Kraif, 2016) :

- a) Récurrence des expressions ;
- b) Construction autour d'un prédicat verbal avec la flexibilité sur les plans syntaxique et lexical ;
- c) Fonctions rhétoriques liées au discours argumentatif ;
- d) Actualisation énonciative.

Bien que ces quatre critères soient essentiels pour caractériser les routines, il existe encore des aspects à approfondir, car en réalité, ce type d'expressions phraséologiques est plus complexe que ne le montrent les premiers exemples examinés. Parmi les aspects à approfondir et à modéliser, nous pouvons noter la nature phrastique des routines, le degré de flexibilité des composantes des routines, les paramètres énonciatifs, les éléments facultatifs ou périphériques comme les déictiques, les marqueurs discursifs.

Nous proposons maintenant d'examiner le statut théorique des routines sémantico-rhétoriques dans le domaine de la phraséologie de façon à mieux définir les routines. Nous rappellerons quelques termes et notions en usage afin de montrer que les routines sémantico-rhétoriques, en particulier celles de guidage du lecteur, ont leurs propres caractéristiques qui les distinguent d'autres notions proches. Dans cette logique, nous évoquerons principalement trois types de notions : les routines conversationnelles de la phraséologie pragmatique, les notions phraséologiques utilisées dans l'enseignement de l'anglais, les notions utilisées en langue française, puisant leur origine dans la textométrie et puis dans l'analyse du discours. Ces dernières ont également beaucoup influencé la définition de notre objet d'étude.

1.2.2. Routines conversationnelles

Dans cette sous-partie, la notion de routines conversationnelles va être évoquée dans le domaine de la phraséologie pragmatique. Cette notion partage avec les routines sémantico-rhétoriques le même terme « routine », mais renvoie à des expressions très variées, dont certaines présentent une ressemblance avec les routines sémantico-rhétoriques au niveau formel.

Les routines conversationnelles sont des expressions récurrentes dans les conversations. Elles relèvent de ce fait du domaine de la phraséologie pragmatique. La notion de routines conversationnelles s'est développée généralement à travers les études sur les conversations, à

savoir l'interaction face à face. En général, c'est un type d'expressions rituelles permettant la communication transparente entre les interlocuteurs d'une conversation. L'emploi des routines conversationnelles est fortement influencé par l'usage de la communauté discursive ciblée en répondant aux conventions normalisées de la conversation, comme Coulmas (2011, p. 3) l'a indiqué :

Les routines sont des types d'interactions où aucune « négociation » n'est nécessaire entre les individus. Dans la mise en œuvre des routines verbales, la créativité du langage est socialement canalisée en fonction de la réussite des solutions liées aux tâches verbales récurrentes, déterminées par l'adéquation fonctionnelle et la tradition⁷.

En utilisant ces expressions rituelles aux fonctions spécifiquement requises dans une situation communicative, le sujet parlant s'intègre dans la communauté discursive cible. Par conséquent, les routines conversationnelles sont étroitement liées à l'aspect social de l'utilisation du langage. Leurs fonctions consistent à guider une interaction verbale dans la société :

Les routines sont un moyen d'accompagner les individus dans leurs interactions sociales. [...] elles constituent des liens normalisés entre ce que les gens disent effectivement et le type de fonctions communicatives que leurs énoncés permettent d'accomplir⁸. (Coulmas, 2011, p. 6-7)

Les routines conversationnelles sont souvent envisagées dans une perspective fonctionnelle, liée à l'aspect pragmatique et social de la communication. Dans cette perspective, les études mettent en lumière l'organisation conversationnelle à un niveau plus abstrait en examinant les stratégies, les actions et les procédures conversationnelles, autrement dit l'organisation conversationnelle (*ibid.*, p.10). Aijmer (2014, p. 2) a résumé ainsi trois groupes de routines conversationnelles analysés dans le corpus *London-Lund Corpus of Spoken English* :

- a) les routines liées aux actes de langage comme remercier (*thank you*), pardonner (*I beg your pardon*), demander (*if you could, I should be glad to*), saluer (*hello; how do you do?*), etc.;
- b) les routines permettant d'exprimer les sentiments, les attitudes du sujet parlant (*I really like/love*);

⁷ Notre traduction, la citation originale est : Routines are kinds of interactions where no "negotiation" is necessary between individuals. In the enactment of verbal routines, the creativity of language is socially canalized according to successful solutions of recurring verbal tasks, fixed by functional appropriateness and tradition.

⁸ Notre traduction, la citation originale est : Routines are a means of guiding a person's normal participation in social interaction. [...] they constitute standardized links between what people actually say and what sort of communicative functions their utterances serve to perform (Coulmas, 2011, pp. 6–7).

c) les routines liées aux fonctions d'organisation conversationnelle, comme l'énumération (*firstly; in the first place*), la récapitulation (*in a word; briefly*), etc.

En dehors des fonctions pragmatiques, l'étude de ce type d'expressions ne s'intéresse pas moins aux aspects formels et sémantiques qui touchent parfois aux phénomènes de grammaticalisation, de lexicalisation, voire d'idiomatization. Par exemple, les routines comme *I think, to tell you the truth* ont été analysées en tant qu'adverbes à travers le processus de grammaticalisation (Aijmer, 2014, p. 10).

En ce qui concerne l'aspect formel, les routines conversationnelles peuvent recouvrir des catégories très variées. Reprenons l'étude citée plus haut sur les routines d'énumération comme exemple. Ces routines peuvent être des mots simples (*firstly*), des syntagmes prépositionnels (*in the first place, to begin with*), des phrases (*the next thing is*).

En ce qui concerne l'aspect sémantique, on observe aussi une certaine hétérogénéité. Des routines rituelles comme *Good to see you! how do you do?* dans la salutation ne sont pas vraiment compositionnelles au plan sémantique, car leurs fonctions pragmatiques sont plus déterminantes que leurs sens compositionnels (Coulmas, 2011, p. 4). Cependant, il existe des routines comme *to begin with, the next thing is* qui sont transparentes sémantiquement. Ces routines conservent leur sens habituel, respectant les règles grammaticales, syntaxiques, sémantiques observées dans les mots ou expressions libres.

En résumé, les routines conversationnelles regroupent un ensemble d'expressions lexicales ou phrastiques habituellement utilisées dans les conversations qui possèdent des caractéristiques syntaxiques et sémantiques fortement diversifiées par rapport aux routines sémantico-rhétoriques.

Parmi toutes les variations des routines conversationnelles, nous nous intéressons ici en particulier à un type qui présente des traits communs avec les routines sémantico-rhétoriques, d'un point de vue syntaxique et sémantique : ce sont des expressions dans lesquelles différents groupes de mots peuvent s'intégrer. Il s'agit des phrases complètes ou ouvertes, dont certains arguments peuvent présenter des variations catégorielles⁹. Comme dans la routine (Aijmer, 2014, p. 91), on observe des variantes : l'intensifieur peut être *so|very|terribly*, les arguments

⁹ On parle également de « lexicalized sentence stem » (Pawley & Syder, 1983, p.192). Initialement utilisé pour désigner des phrases lexicalisées ou semi-lexicalisées que les natifs de l'anglais ont tendance à employer spontanément dans la conversation (*ibid*).

après *sorry* peuvent varier (*that|to about...*). Ces caractéristiques syntaxiques et sémantiques sont analysables à l'aide des règles linguistiques.

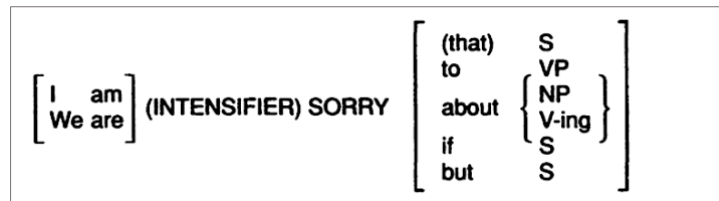


Figure 1-1 : Routine pour s'excuser

En guise de conclusion, nous pouvons voir qu'il existe encore un certain flou qui accompagne cette notion de routine conversationnelle, des plus variables aux plus fixes comme *how do you do* ou *how are you*. On note par ailleurs que la caractéristique commune des routines conversationnelles qui permet de regrouper ces mots ou expressions très variés reste toujours la dimension pragmatique, en termes de fonction exercée dans les conversations.

Nous résumons ici les points communs et les divergences entre les routines conversationnelles et les routines sémantico-rhétoriques (Tutin & Ji, 2022, p. 46) :

- Points communs : toutes ces notions sont envisagées comme des expressions spécifiques à un genre ou un registre de discours. Elles exercent des fonctions spécifiques dans le discours.
- Divergences : les routines conversationnelles consistent en une notion englobant toute expression monolexicale ou plurilexicale, sémantiquement transparente ou non, qui remplit des fonctions pragmatiques dans les conversations. L'aspect pragmatique et social constitue une clé centrale dans leurs définitions. Quant aux routines sémantico-rhétoriques, ce sont plutôt des phrases ou semi-phrases, sémantiquement transparentes, exerçant des fonctions rhétoriques dans un genre codifié.

En un mot, même si ces deux notions incluent le même terme de « routines », elles diffèrent l'une de l'autre. Les routines sémantico-rhétoriques constituent un ensemble d'expressions plus homogène dans leurs formes par rapport aux routines conversationnelles. Nous pensons qu'elles doivent faire l'objet d'une théorisation dans le domaine phraséologique.

Si les *routines conversationnelles* partagent le mot *routine* avec les *routines sémantico-rhétoriques*, nous trouvons aussi des notions phraséologiques proches utilisées dans le domaine de l'enseignement de l'anglais, ou dans les domaines de la lexicométrie et de l'analyse du discours pour la langue française. Ces notions, souvent développées à partir des chaînes de mots

(*segments répétés* en français et *n-grams* en anglais), présentent des ressemblances frappantes avec notre objet d'étude en ce qui concerne les caractéristiques linguistiques et fonctionnelles. Ces notions, à la fois proches et différentes des routines, constituent une référence incontournable pour notre étude. Dans les deux sections qui suivent, nous les discuterons tour à tour.

1.2.3. Des *sentence builders* aux *p-frames* : quelques notions utiles dans la phraséologie de langue anglaise

Pour la langue anglaise¹⁰, plusieurs travaux sur les séquences lexicalisées se sont intéressés à des objets proches de nos routines, comme les *sentence builders* (*I think (that)*¹¹ *X; its' only in X that Y, the _er X, the _er Y*), les *sentence frames and heads* (*secondly, ... and finally; Would you like to...*) (Lewis, 1993, 1996) ou les *lexical bundles* (*to be able to, what I want to*) (Biber, Conrad, & Cortes, 2004 ; Biber & Barbieri, 2007a ; Biber, 2009), les *p-frames* (Römer, 2010) ou *lexical frames* (Gray & Biber, 2015).

De manière générale, nous avons affaire à deux grands types d'unités phraséologiques : soit linéaires et contigus comme les *lexical bundles*, soit linéaires, mais possiblement discontinus, regroupant un petit ensemble de mots qui occupe la même position dans l'expression comme les *sentence builders*, les *sentence frames and head*, les *p-frames* ou les *lexical frames* (Gray & Biber, 2015). Nous aborderons ces notions dans l'ordre chronologique.

1.2.3.1. *Sentence builders* (Nattinger & DeCarrico, 1992) et *Sentence frames and heads* (Lewis, 1996)

Ces deux notions ont été développées dans l'objectif d'enseigner la langue anglaise, notamment sur le plan du lexique. Cette perspective pédagogique conditionne le fait que les unités phraséologiques observées dans ces études ne s'appliquent pas à un genre discursif spécifique, mais sont observables dans plusieurs genres comme la conversation quotidienne, le discours scientifique, etc.

Nattinger et DeCarrico (1992) ont proposé le concept de *lexical phrases* et analysé leurs aspects formels et fonctionnels. Les *sentence builders* qui nous intéressent ici constituent un sous-type des *lexical phrases*. Généralement, le terme *lexical phrases* désigne un type d'unités lexico-grammaticales préfabriquées de longueur variable qui manifeste une fréquence élevée et

¹⁰ Nous ne comptons ici que les études pour l'anglais général ou l'anglais académique (oral ou écrit), dont l'objectif peut être varié : la didactique, l'acquisition, l'analyse du discours.

¹¹ Le mot *that* est facultatif dans cette *sentence builder* : *I think X*.

a un sens plus figé par rapport aux combinaisons libres. Il s'agit d'un composite conventionnel de forme et de fonction¹² (Nattinger & DeCarrico, 1992, p. 42). Le critère de fonction pragmatique est central dans la définition des *lexical phrases*, ce qui les différencie des collocations auxquelles ne sont pas attribuées spécifiquement des fonctions pragmatiques (*ibid.*, p. 36).

Deux types d'expressions préfabriquées ont été inventoriés en tant que *lexical phrases*¹³ selon leurs caractéristiques formelles (*ibid.*, p. 36–37) :

- 1) des chaînes d'items lexicaux spécifiques qui n'admettent pas de variabilité syntaxique ni de substitution paradigmatique (*for the most part; by the way; in part, how are you? Have a nice day!*). Ces chaînes de mots sont des expressions polylexicales ou rituelles (certaines sont les routines conversationnelles discutées plus haut).
- 2) des cadres phrastiques (*frame*) regroupant des items lexicaux fixes et des items lexicaux flexibles correspondant aux catégories syntaxiques et sémantiques (<a N [+ time] + ago> qui correspond aux *a year ago, a month ago*), auxquels appartiennent les *sentence builders*.

Plus précisément, les *sentence builders* fournissent un cadre pour une proposition avec des emplacements libres en matière d'arguments ou de paramètres lexicaux, et permettent d'exprimer une idée complète combinée avec une fonction spécifique (Nattinger & DeCarrico, 1992, p. 42). Du point de vue formel, ces expressions doivent avoir une structure syntagmatique simple qui est facile à interpréter et est en mesure d'inclure des paradigmes lexicaux flexibles en acceptant parfois des trous dans la phrase. Observons l'exemple cité plus haut <modal + you +VP>. Sur l'axe syntagmatique, c'est une phrase avec une structure syntaxique simple composée par un mot modal, le pronom personnel *you* et un syntagme verbal ; en revanche, sur l'axe paradigmatique, plusieurs mots modaux et des syntagmes verbaux peuvent participer à la construction d'une phrase concrète : *could you pass the salt? Would you lend me a dollar?* (Nattinger & DeCarrico, 1992, p. 18). Ces expressions récurrentes peuvent donc être à la fois

¹² Le *lexical phrases* peut exercer à la fois plusieurs fonctions pragmatiques selon le contexte.

¹³ Nattinger & DeCarrico (1992, p. 38-47) ont proposé une typologie pour classer les *Lexical phrases* : 1) mots polylexicaux (*so long, in essence*) ; 2) expressions institutionnalisées : *how do you do ? (Comment ça va ?), a watched pot never boils (s'inquiéter ne fait pas avancer les choses)* ; 3) phrases contraintes courtes et moyennes : *a _ ago ; the _er, the _er*. 4) *Sentence builders* : *not only X, but also Y*.

continues (*my point is that X; I'm a great believer in X*) ou discontinues (*it seems [to me] [that] X; it is|has been [often] asserted/believed/noted that X*).

Les *sentence builders*, peuvent remplir diverses fonctions pragmatiques auxquelles sont confrontés les apprenants de la langue anglaise (*ibid.*, p. 60–65) :

- 1) les fonctions assurant le bon déroulement et introduisant l'objectif d'une interaction sociale (salutation : *I don't catch|get your name*; montrer le désaccord : *I don't [really] agree [with you|X]*, etc.) ;
- 2) Les fonctions de conversation autour des thèmes quotidiens, comme parler du climat (*is it going to X?*), parler de ce que l'on aime (*what do you like to X?*), parler de faire des courses (*I want to buy|see X*), etc. ;
- 3) Les fonctions discursives qui contribuent à l'organisation du sens et de la structure du discours, y compris les fonctions d'exemplification (*it's like X*), de qualification (*it's only in X that Y*), d'évaluation (*I'm [not] absolutely sure|positive|certain [but] _____*), de récapitulation (*my point [here] is that X*).

En résumé, du point de vue formel, les *sentence builders* sont des phrases incomplètes relativement longues, soit continues ou discontinues, présentant une grande flexibilité syntaxique et paradigmatique. Cet aspect formel présente des analogies intéressantes avec les routines sémantico-rhétoriques en termes de flexibilité syntaxique et lexicale. Du point de vue fonctionnel, nous pouvons observer des différences. Comme l'étude porte sur ces expressions présentant un intérêt didactique, elles s'intéressent à la large palette de fonctions pragmatiques que rencontrent les apprenants dans leur utilisation des langues cibles.

Une autre notion proche de la notion de *sentence builders* est celle de *sentence frames and heads* qui est développée par Lewis (1996). Dans l'intention de traiter la question de l'enseignement du lexique en anglais, Lewis (1996) a proposé quatre types d'items lexicaux : les mots simples et polylexicaux, les collocations, les phrases institutionnelles et les *sentence frames and heads*¹⁴. Ce dernier type regroupe l'ensemble des expressions récurrentes ayant une fonction discursive permettant le décodage d'un texte, comme annoncer les objectifs (*we come now to a number of important reservations*) (Lewis, 1996, p. 259). De plus, l'auteur sépare les phrases ouvertes (*sentence heads*) et les phrases d'ossature (*sentence frames*). Il s'agit ici de la question de l'empan des expressions. Les phrases ouvertes sont similaires aux *sentence builders* continues. Par exemple, dans les articles académiques, il s'agit d'expressions comme : *the main*

¹⁴ Cette catégorisation s'est inspirée de celle de Nattinger & DeCarrico (1992) sur les *lexical phrases*.

cause of the problem is. Les phrases d'ossature, quant à elles, ont une portée supérieure qui touche à la fois plusieurs phrases. Reprenons l'exemple suivant tiré d'un article académique donné par Lewis (1997, p. 11) :

There are broadly speaking two views of... The more traditional, usually associated with ... and his/her colleagues, suggests that while the more progressive view, associated with ... suggests... In this paper I wish to suggest a third position, which, while containing elements of the view proposed by ... also takes account of recent developments in ... which have produced evidence to suggest ... and so on.

Les trous marqués par les points de suspension indiquent des positions (cadres paradigmatiques) pour des groupes de mots différents. Toutes les phrases incluant de telles places lexicalement flexibles forment des phrases d'ossature. Nous pouvons constater que les phrases d'ossature sont la combinaison d'au moins deux phrases ouvertes. À l'exception de leurs portées, elles ne manifestent pas de différence radicale en matière des caractéristiques linguistiques. Signalons également que les phrases d'ossature recensent également des cadres phrastiques comme *Firstly... ; Secondly ...; Finally...* qui ne constituent même pas des phrases pour nous. Par conséquent, il nous semble que l'aspect formel des *sentence frames and heads* n'est pas clairement défini, alors que l'aspect fonctionnel est bien envisagé comme pragmatiquement lié aux genres discursifs.

Les notions de *sentence builders* et *sentence frames and heads* sont conçues dans une perspective de didactique du lexique en anglais. Bien que l'objectif initial ne soit guère linguistique, ces études permettent de montrer que les expressions polylexicales phrastiques méritent une étude systématique, et qu'elles sont bien lexicalement flexibles, sémantiquement transparentes, et à interpréter sous l'angle d'une fonction spécifique dans leur contexte concret. Il faut admettre que les éléments définitoires de ces deux notions sont très semblables à ceux des routines sémantico-rhétoriques : la récurrence des phrases, la variabilité lexicale et les fonctions spécifiques. Les *sentences builders* et les *sentence frames and heads* sont les premières unités phraséologiques ayant une portée phrastique que nous avons pu observer.

1.2.3.2. *Lexical bundles*¹⁵ (Biber, 2006, 2009), *p-frames* (Römer, 2010) et *lexical frames* (Gray & Biber, 2015)

Nous nous concentrons dans cette partie sur trois notions phraséologiques développées dans le cadre de l'EAP (*English for Academic Purposes*). Ce sont toutes des séquences récurrentes qui sont définies statistiquement comme des séquences de mots les plus fréquentes

¹⁵ En français, ce terme a été traduit comme « paquets lexicaux » (Cislaru & Olive, 2017).

dans un registre donné, quels que soient leur idiomaticité ou leur statut syntaxique dans le discours (Biber *et al.*, 1999, p. 990). Le critère définitoire essentiel de ces expressions est la fréquence des cooccurrences des mots dans le corpus. Nous pouvons faire correspondre ces notions à la notion de *segments répétés* (Salem, 1987), définie comme des séquences de plus de trois mots qui cooccurrent fréquemment dans un corpus (nous y reviendrons plus tard). Bien que ces trois concepts soient analogues à celui des segments répétés, ils présentent des différences en ce qui concerne les formes linguistiques, notamment sur le plan lexical.

Les *lexical bundles*¹⁶ sont souvent envisagés comme des séquences récurrentes de nature continue dans les textes du registre universitaire (Biber, 2006). Ce sont des suites de mots dont la syntaxe et le sens sont transparents. Par exemple, *do you want to* ou *I don't know what* ne présentent aucune opacité quant à leur description syntaxique et sémantique. Il est à noter que les paquets lexicaux sont souvent des propositions syntaxiquement incomplètes et ouvertes (Biber, 2006, p. 134). Les pourcentages de paquets lexicaux incomplets représentent 85 % dans le discours scientifique oral et plus de 95 % dans les écrits académiques (*ibid.*, p.135). Il existe deux types de paquets lexicaux incomplets :

- a. les paquets lexicaux qui sont constitués d'une phrase encadrant et introduisant une phrase subordonnée, dont le dernier mot sert à débiter cette phrase subordonnée, comme dans les deux exemples cités plus haut ;
- b. les paquets lexicaux qui servent de relateurs entre les éléments qui suivent et qui précèdent, comme des locutions *in the case of*.

Les paquets lexicaux sont continus et ne tolèrent pas l'insertion de trous à l'intérieur de leurs constituants, mais peuvent introduire des éléments à leur suite ou relier d'autres éléments dans une phrase.

Certes, la plupart des études sur les paquets lexicaux se concentrent sur les expressions continues, mais Biber (2009) a identifié des expressions discontinues qui ont une fonction de cadres phrastiques et qui intègrent à la fois d'éléments fixes et variables comme *to the _ of, on the _ of*, etc. Ce type d'expressions discontinues possède également plusieurs dénominations comme *lexical phrases, phrases-frame, lexical frames*, etc. Römer (2010, p. 98) a nommé *phrases frames* ou *p-frame* des *lexical bundles* discontinus qui consistent davantage en une

¹⁶ Le terme *lexical bundles* a été créé dans le livre *Longman Grammar of Spoken and Written English* (Biber *et al.*, 1999).

série d'expressions identiques avec un seul constituant variable comme *at the _ of, in the _ of*¹⁷. Gray & Biber (2015) préfèrent le terme *lexical frame*. Malgré la différence terminologique, ils désignent tous des *lexical bundles* discontinus avec les variations lexicales.

Les définitions de ces unités phraséologiques ne se basent pas que sur la statistique. Les auteurs soulignent également l'importance de leurs fonctions dans le discours. Par exemple, dans le discours scientifique, d'une part, l'identification des fonctions des paquets lexicaux permet de mieux comprendre la production des textes universitaires ; d'autre part, la distribution des fonctions offre des pistes d'analyse des différences en matière de registres et de disciplines universitaires (Biber, 2006). Dans les textes universitaires, trois fonctions principales¹⁸ ont été mises en évidence : fonctions interpersonnelles, en particulier en ce qui concerne le positionnement (stance expressions en anglais), fonctions discursives pour organiser le discours et fonctions référentielles (Biber, 2006, p. 139), comme chez Halliday.

Les fonctions de positionnement se divisent en deux classes plus précises : la fonction épistémique et la fonction d'attitude ou de modalité. La première vise à montrer l'évaluation de l'état des connaissances à transmettre. Elle est souvent en rapport avec la certitude ou non, le degré de possibilité comme *I think it was..., I don't know what...* Le deuxième regroupe des expressions exprimant l'attitude de l'auteur sur les informations à transmettre ; nous recensons des expressions de désir *I don't want to*, d'obligation *you have to do*, d'intention *is going to be, to be able to* (*ibid.*, p.139–142).

Les fonctions référentielles se concentrent sur l'identification d'une entité ou d'un attribut particulièrement important d'une entité. Quatre classes sont mises en place : la fonction de focus *that's one of the problems*, la fonction de référence imprécise *and something like that*, la fonction de spécification *a little bit about, the size of the*, la fonction textuelle *as shown in Figure, locale in the United States*, temporelle *the end of the year*.

Les fonctions qui nous intéressent le plus sont les fonctions discursives. Ces fonctions visent à mieux organiser le texte et sont aussi identifiées pour les routines :

- 1) La fonction d'introduction du thème ou de focalisation. Cette fonction est souvent

¹⁷ Ces expressions avec des cadres sont très proches des « collocational frames » de Renouf & Sinclair (1991, p. 128).

¹⁸ Cette taxinomie ternaire a été inspirée de la théorie de Halliday (1985) sur les métafonctions du langage (nous l'aborderons dans le chapitre 3). En fait, la plupart des études sur la phraséologie dans le discours scientifique en anglais sont influencées par Halliday. Pour plus d'informations sur ce point, se référer à l'article d'Oakey (2020).

présente dans la classe, énoncée par les enseignants désirant annoncer un nouveau thème à aborder par la suite : *I want to talk about, what I want to do is*.

- 2) La fonction d'élaboration ou de clarification du topique qui vise à construire ou clarifier un thème, comme *you know what I mean, on the other hand*, etc.

Ces fonctions se révèlent indispensables pour assurer la cohérence et la cohésion du texte. Nesi et Basturkmen (2006) ont investigué les fonctions des paquets lexicaux permettant de garantir la cohésion textuelle dans 160 cours et conférences issus des corpus *The British Academic Spoken English (BASE)* et *The Michigan Corpus of Academic Spoken English (MICASE)*, d'une taille de 1 270 798 mots. Elles ont recueilli des paquets lexicaux d'une longueur de 4 mots de fréquences au moins égale à 10, et découvert que deux catégories des paquets lexicaux, à savoir les expressions de références et les organisateurs du discours (qui correspondent aux fonctions discursives et références chez Biber, 2006), peuvent contribuer à la cohésion du texte. Entre autres, l'expression *at the same time* crée une relation de contraste entre deux éléments (Nesi & Basturkmen, 2006, p. 35), *and you can see* peut signaler une relation d'apposition entre deux propositions, et elle apparaît souvent en simultanément avec des locutions adverbiales comme *for example, for instance* (Nesi & Basturkmen, 2006, p. 34).

L'étude de ces séquences phraséologiques permet d'étudier les caractéristiques d'un registre. Dans son article de 2007, Biber renvoie à son étude contrastive précédemment effectuée sur les *lexical bundles* dans les classes et les manuels ainsi que dans les conversations (discours oral, des dialogues quotidiens) et les articles académiques. Les résultats montrent que l'emploi des paquets lexicaux dépend des registres. Par exemple, la langue d'enseignement dans les classes a tendance à employer plus de paquets lexicaux que les autres registres, mais elles emploient les *lexical bundles* de positionnement comme le discours scientifique oral, et ceux de fonctions référentielles comme les articles académiques (Biber & Barbieri, 2007 b, p. 265).

Quant aux *p-frames* et *lexical frame*, la partie variable est souvent essentielle pour identifier des différences entre registres. Dans les recensions d'ouvrage en anglais, Römer (2010) a identifié la fonction d'évaluation (*it is _ that*), de structuration (*in the _ chapter*), de référence au contenu d'ouvrage (*the relationship between _ and _*), d'organisation discursive (*on the _ hand*). Le registre de l'écrit académique en anglais a tendance à présenter plus des *p-frames* dont les parties fixes sont des mots-outils et la partie variable est remplie par les mots qui portent un sens crucial pour le contenu, tandis que le registre des conversations quotidiennes

contient plus d'expressions dont les parties fixes et variables sont occupées par des mots-outils (Biber, 2009).

En résumé, les *lexical bundles*, les *p-frames* et les *lexical frames* renvoient aux expressions récurrentes, prédictibles, incomplètes, continues ou non, qui remplissent des fonctions spécifiques comme les fonctions interpersonnelles, discursives, référentielles dans les textes. Leurs réalisations varient en fonction du genre discursif. Nous pouvons dire que les notions de *p-frames* et de *lexical frames* sont plus proches des routines sémantico-rhétoriques sur l'aspect formel. Pour l'aspect fonctionnel, comme les fonctions de ces deux notions sont étudiées sur la base des *lexical bundles*, nous devons dire que les études portant sur les *lexical bundles* fournissent des propositions constructives.

Les études phraséologiques sur ces expressions en anglais s'intéressent principalement à leur distribution et à l'analyse de leurs fonctions discursives dans un registre, ce qui est différent des études sur la phraséologie pour la langue française. Comme nous allons le voir, les linguistes qui travaillent sur la langue française ont souvent des perspectives interdisciplinaires.

1.2.4. Des segments répétés aux motifs : des notions de la lexicométrie et de l'analyse du discours en langue française

Des travaux issus de la tradition lexicométrique française¹⁹ sont consacrés aux objets qui nous intéressent à travers plusieurs notions comme le *segment répété* (Salem, 1987) ou la notion de *motifs* (Longrée & Mellet, 2013), et enfin les routines discursives (Sitri & Veniard, 2017).

1.2.4.1. Segments répétés (Salem, 1987, p. 21)

La première notion proche de notre objet est celle de *segments répétés*. Définis par Salem (1987, p. 21) comme « toutes les suites de formes graphiques non séparées par un caractère délimiteur de séquence, qui apparaissent plus d'une fois dans ce corpus de textes », les *segments répétés* ont été conçus au début en tant qu'une méthode lexicométrique visant à segmenter le corpus et à extraire des formes avec leurs occurrences dans la perspective de la statistique textuelle (Salem, 1986).

¹⁹ Cette tradition s'intéresse à des recherches de l'analyse du discours en sciences humaines et sociales ayant une perspective quantitative. Appelée également « statistique linguistique » (Guiraud, 1959), « statistique lexicale » ou « linguistique quantitative » (Muller, 1969, 1979), « statistique textuelle » (Salem, 1987), la lexicométrie combine l'étude des sciences du langage avec l'étude statistique de l'informatique.

Du point de vue graphique, les *segments répétés* sont souvent des chaînes de mots contiguës sans interruption interne. La forme graphique des segments répétés présente très peu de variations morphologiques (Cislaru & Olive, 2017). Prenons l'exemple du travail de Siepmann & Bürgel (2019) sur les unités phraséologiques du français contemporain. Dans le *Corpus de référence du français contemporain* d'une taille de 310 millions de mots, qui comporte des textes oraux, des textes écrits et des textes écrits oralisés, pendant la période entre 1945 et 2014, ces deux linguistes ont cherché des séquences phraséologiques (n-grammes) entre deux et quatre mots. La plupart des unités phraséologiques retrouvées sont des locutions comme *en plus, par exemple, à partir de*, ou comme des syntagmes simples comme *tout le monde, de plus en plus, une fois de plus*, etc. Ce sont des segments répétés fixes à la fois syntaxiquement et graphiquement, c'est-à-dire, sans actualisation dans le discours. En dehors de ces unités phraséologiques fixes, les auteurs ont repéré des segments répétés lemmatisés qui tolèrent donc des variations morphologiques, comme *dans* + déterminant possessif + *tête* qui peut correspondre aux :

dans ma/sa/leur/votre/nos/vos/tête(s)

La variation des segments répétés est assez restreinte. En résumé, les segments répétés constituent un outil d'extraction phraséologique avec une configuration lexicale assez fixe qui permet de témoigner éventuellement des variations très faibles sur le plan morphologique²⁰.

En général, les segments répétés contiennent au moins deux mots et peuvent atteindre jusqu'à plus de dix mots. L'ouvrage de Salem (1987) recense des segments de deux jusqu'à douze mots dans le discours politique. Les segments de deux mots graphiques sont souvent des termes propres au domaine politique où s'inscrit le corpus SYND²¹ comme *organisation syndicale, sécurité sociale* (*ibid.*, p. 145–146). Les segments répétés plus longs correspondent aux syntagmes plus longs ou aux phrases complètes, par exemple *pays en voie de développement* (*ibid.*, p. 145), *la nature de ces affrontements sera fonction de l'état des rapports* (*ibid.*, p. 85). Mais des segments grammaticaux comme *de la, et de la, à tous les*, etc. sont éliminés des segments répétés, car ils ne constituent pas des « unités de fonctionnement » dans le texte (*ibid.*, p. 51). La notion d'unité de fonctionnement renvoie ici aux unités polylexicales qui ont un sens autonome et qui peuvent s'employer en tant qu'unité lexicalisée sémantique dans le texte.

²⁰ Ces variations morphologiques peuvent être ignorées à l'aide d'une lemmatisation.

²¹ Le corpus SYND regroupe des textes de résolutions confédérales votés par quatre syndicats ouvriers français entre 1971-1976 (Salem, 1987, p. 57). Il s'agit d'un corpus politique.

Les segments répétés sont proches de la notion de *lexical bundles* en ce qui concerne les aspects formels et sémantiques. Ce sont des suites de mots qui cooccurrent dans un corpus et conservent une autonomie sémantique. Contrairement aux *lexical bundles* auxquels les auteurs attribuent des fonctions, les segments répétés ne revendiquent pas une telle attribution. Néanmoins, l'emploi généralisé des segments répétés a suscité un grand intérêt dans le monde francophone, en particulier pour les études sur le genre textuel. Nous pensons ici particulièrement aux motifs (Longrée & Mellet, 2013, 2015 ; Novakova & Sorba, 2017 ; Goossens, Jacquot, & Novakova, 2018) et aux routines discursives (Sitri & Veniard, 2017).

1.2.4.2. *Motifs*

La notion de motif a été développée initialement dans le domaine de l'analyse du discours dans une perspective de classification des textes et de caractérisation du genre discursif (Longrée *et al.*, 2008 ; Mellet & Longrée, 2009, 2012 ; Quiniou *et al.*, 2012 ; Longrée & Mellet, 2013). Dans le travail de Mellet & Longrée (2009), les motifs peuvent servir à faire émerger les caractéristiques textuelles des textes historiques en latin. L'idée est que les textes peuvent être distingués et classés automatiquement par des structures récurrentes, soit les motifs. Dans des études plus récentes de ces deux auteurs, les fonctions discursives des motifs ont été mises en évidence pour distinguer les genres discursifs (nous y reviendrons ci-dessous dans cette section).

Par rapport aux n-grammes, les motifs possèdent des caractéristiques linguistiques plus complexes. Ils contiennent des éléments de nature variable qui ne sont pas nécessairement contigus. Les motifs sont définis comme « l'association récurrente de n éléments d'un ensemble donné » (*ibid.*), c'est-à-dire qu'il s'agit de la cooccurrence de n éléments dans un texte :

si le texte est formé d'un certain nombre d'occurrences des éléments A, B, C, D, E, un motif pourra être la microstructure récurrente ACD ou bien encore AA, etc., sans que l'on préjuge ici de la nature des éléments A, B, C, D, E en question. (Longrée & Mellet, 2013, p. 66)

Ce ne sont donc pas des expressions phraséologiques définies avec le critère de figement linguistique au niveau lexical, syntaxique et sémantique, mais avec un critère statistique de fréquence de cooccurrence. Au-delà de la récurrence, le motif constitue un cadre collocationnel qui permet de « conceptualiser la multidimensionnalité (ou le caractère multiniveau) de certaines formes récurrentes qui sollicitent à la fois le lexique, les catégories grammaticales et la syntaxe, éventuellement la prosodie, la métrique » (Longrée & Mellet, 2013, p. 66). Du point de vue formel, le motif est linéaire sur l'axe syntagmatique. Par contre, la permutation des constituants est acceptée ; comme *ici j'aimerais toutefois préciser* et *ici, toutefois, j'aimerais préciser*. De plus, il mobilise non seulement des éléments fixes, mais aussi des éléments

variables comme des variations lexicales, grammaticales et certaines actualisations en discours. Nous pouvons citer pour exemple *ici j'aimerais toutefois préciser* plusieurs réalisations (Longrée & Mellet, 2013, p. 67) :

<i>Ici</i>		<i>j'</i>		<i>aimerais</i>		<i>toutefois</i>		<i>préciser</i>
<i>Ici</i>		<i>j'</i>		<i>aimerais</i>		(vide)		<i>préciser</i>
(vide)		<i>j'</i>		<i>aimerais</i>		<i>toutefois</i>		<i>préciser</i>
<i>Ici</i>		<i>je</i>		<i>voudrais</i>		<i>toutefois</i>		<i>préciser</i>
<i>Ici</i>		<i>je</i>		<i>souhaiterais</i>		<i>toutefois</i>		<i>préciser</i>
<i>Ici</i>		<i>je</i>		<i>souhaite</i>		<i>toutefois</i>		<i>préciser</i>

La présence d'un élément n'est pas obligatoire, comme *toutefois, ici* dans ces exemples. La variation lexicale se présente à travers les verbes modaux *aimer, vouloir, souhaiter* qui constituent un paradigme pour exprimer le souhait. Le verbe lui-même manifeste également plusieurs formes morphologiques liées aux variations temporelles et modales : *souhaiterais/souhaite*.

Chez Quiniou *et al.* (2012, p. 2), ce type de variations des formes présente un niveau d'abstraction relativement plus faible, appelé « des motifs d'items », moins abstraits par rapport à « des motifs d'itemsets ». Dans un motif d'items, les constituants variables se limitent à des mots lexicaux (*ibid.*), représentés en général par des trous dans le motif séquentiel. Par exemple le motif d'items *des _ plus _ que* correspond aux phrases *il a **des morsures plus venimeuses que celles de ta bouche, des cailloux anguleux plus brillants que des marbres***, identifiées par les auteurs comme étant caractéristiques de leur sous-corpus poétique. En revanche, les motifs d'itemsets généralisent les parties variables en fonction de contraintes manifestées par des trous et les étendent sur le plan catégoriel. Dans cet exemple, le motif d'items se généralise en tant que motif d'itemsets *des N plus ADJ que*. Quels que soient les niveaux d'abstraction des motifs d'items ou d'itemsets, le motif est par essence linéaire, que l'on considère les trous sur le plan formel ou catégoriel grammatical.

À l'instar des *sentence builders*, les motifs sont aussi une combinaison de formes et de fonctions. Ils sont sémantiquement compositionnels et montrent une certaine « stabilité du sens et de la fonction textuelle » (Longrée & Mellet, 2013, p. 67). L'exemple *ici j'aimerais toutefois préciser* est doté d'un sens facile à décoder et assez stable malgré les variations lexicales et grammaticales. La stabilité concerne également la fonctionnalité discursive et le contexte d'utilisation des motifs qui s'avèrent très importants dans la définition des motifs. Les fonctions

structurantes qui servent à organiser les discours et les textes et des fonctions caractérisantes qui contribuent à distinguer des genres discursifs ou des styles d'écriture ont été observées dans les études de Kraif *et al.* (2016) et de Longrée & Mellet (2013). On peut citer par exemple la fonction cohésive, la fonction résumptive, la fonction conclusive dans les textes historiques (Longrée & Mellet, 2013).

Pour résumer, les motifs sont envisagés ainsi au prisme de quatre critères :

- (1) la récurrence d'un motif dans le discours ;
- (2) la linéarité et la variabilité aux niveaux syntaxique, lexical, grammatical, même catégoriel ;
- (3) la prédictibilité au niveau sémantique ;
- (4) la fonctionnalité structurante et/ou caractérisante qui permet de caractériser le discours en question.

Les motifs sont ainsi définis comme des séquences récurrentes, linéaires et contiguës ou discontinues, associées aux fonctions discursives d'un genre discursif (Longrée & Mellet, 2013, p. 65). Il est à constater que les études sur les motifs présentent déjà une réflexion approfondie sur le niveau d'abstraction de ce type d'expressions et sur leurs fonctions dans le discours. Cela a un grand impact sur les études des routines discursives que nous allons voir et notre étude autour des routines de guidage du lecteur.

1.2.4.3. *Routines discursives*

Il nous reste une dernière notion à examiner, celle des routines discursives. Employées dans plusieurs études (Née, Sitri, & Veniard, 2014, 2016 ; Sitri & Veniard, 2017) sur les rapports éducatifs, les routines discursives représentent des patrons syntaxiques avec des catégories spécifiques intégrant étroitement associés aux situations de production d'un genre discursif donné (X V *état dans* + GN : X *est dans un conflit de loyauté, il est dans la dissimulation de ses devoirs*) (Sitri & Veniard, 2017). Sur le plan formel, les routines discursives et les motifs se rapprochent. Ils peuvent construire aussi des cadres phrastiques intégrant des variations lexicales. La différence se trouve davantage sur le plan fonctionnel et le niveau du prolongement discursif. Contrairement aux motifs qui remplissent plutôt des fonctions structurantes et caractérisantes comme nous avons évoqué plus haut, les routines discursives se focalisent sur l'articulation entre les fonctions textuelles et les déterminations discursives. Autrement dit, les études sur les routines discursives engagent une réflexion plus approfondie sur les plans discursifs et sociaux notamment en ce qui concerne la généricité et

l'interdiscursivité (Née, Sitri, & Veniard, 2016, p. 79). Comme l'ont indiqué Née, Sitri, & Veniard, (2016, p. 78) :

Une routine discursive consiste en la mise en relation de séquences linguistiques récurrentes, partiellement figées (i.e. les patrons définis plus haut) avec des déterminations discursives et des fonctions textuelles propres à un genre ou une sphère d'activité.

Dans cette optique, l'analyse des routines discursives nécessite un regard sur le genre discursif et les normes du genre qui permettent d'expliquer la limite du choix linguistique du scripteur dans une situation de production langagière. *In fine*, les motifs ou les routines discursives manifestent des caractéristiques proches étant donné que la notion de routine discursive est développée à partir de celle de motifs, mais avec les perspectives plus discursives et pragmatiques.

1.2.5. Synthèse

De manière générale, outre les routines conversationnelles qui s'éloignent des routines, nous faisons face à deux grands types d'unités phraséologiques : a) l'un est linéaire et contigu comme les segments répétés, les *lexical bundles*, b) l'autre est linéaire, mais peut être discontinu en regroupant un petit ensemble de mots qui occupe la même position dans l'expression, comme les *sentence builders*, les *sentence frames and heads*, les motifs et les routines discursives.

Nous essayons de récapituler d'abord les notions discutées en haut par le tableau 1-1 en rassemblant leurs caractéristiques. Nous proposons ensuite une synthèse sur quatre paramètres : figement, variation, sémantique et fonction, énonciation.

Notions	Nombres de mots	Prédictibilité sémantique	Degré de Flexibilité	Fonctions
Segments répétés	>=2	√	-	-
Lexical bundles (paquet lexical)	Souvent entre 2 et 4	√	-	Discursive, référentielle et interpersonnelle
P-frames/lexical frames	Souvent entre 2 et 4	√	+	Discursive, référentielle et interpersonnelle
Sentence builders	>=3	√	+	Interpersonnelle, thématique, discursive
Sentence frames and heads	>=3	√	++	Discursive
Motifs	>=3	√	++	Structurante du texte et caractérisante du genre discursif
Routines discursives	>=3	√	++	Fonctions textuelles liées aux déterminations discursives et sociales
Routines sémantico-rhétoriques	>=3	√	++	Fonctions rhétoriques spécifiques des écrits argumentatifs

Tableau 1-1 : Récapitulatif des caractéristiques des notions

a. Aspect idiomatique ou de figement

En général, une unité polylexicale peut être considérée en tant qu'unité phraséologique soit par l'idiomaticité (certaines collocations, expressions figées), soit par la cooccurrence des éléments au sein d'une unité ainsi que leur fréquence en tant qu'unité intégrante dans un corpus. Toutes ces notions ne font pas appel au principe de figement linguistique souvent utilisé pour définir une unité phraséologique dans la tradition lexicologique²². Elles relèvent de la phraséologie étendue (Legallois & Tutin, 2013) et ont ainsi une définition plus statistique qui mobilise des critères quantitatifs comme la fréquence et la cooccurrence des séquences de mots (souvent plus de 3 mots graphiques). Les séquences ainsi recensées sont souvent stockées en tant qu'unités entières et autonomes munies d'un sens précis sans être opaques au plan sémantique, car elles suivent les règles grammaticales habituelles et sont sémantiquement transparentes.

b. Aspect syntaxique

Au niveau syntaxique, nous observons que ces termes regroupent des séquences diverses dont la longueur varie entre 3-6 mots ou plus. Mais la plus grande différence réside sur quatre types de flexibilités morphosyntaxiques : (i) continuité-discontinuité syntaxique ; (ii) changement de l'ordre des mots ; (iii) capacité d'admettre des paradigmes ; (iv) construction complète vs incomplète. Les *segments répétés* ou *lexical bundles* sont des séquences continues sur le plan syntaxique, et ils n'admettent donc ni changement de l'ordre des mots ni substitution paradigmatisée des mots (Longrée & Mellet, 2015). A l'opposé, les *p-frames*, les *sentence builders*, les *sentence frames and heads*, les motifs et les routines discursives recouvrent des séquences discontinues avec la possibilité d'avoir une variation lexicale. Par contre, ces séquences phraséologiques ne peuvent pas admettre de changement de l'ordre des mots. Quant aux routines sémantico-rhétoriques, elles peuvent manifester à la fois les variations syntaxiques et lexicales, et atteignent donc au plus haut degré de variabilité. Nous résumons ces éléments dans le tableau 1-2 :

²² La tradition lexicologique de la phraséologie notamment sur la langue française s'intéresse aux unités phraséologiques régies par des contraintes sémantiques ou l'idiomaticité.

	Dis(+)/ continuité(-)	Changement de l'ordre des mots (oui +, non -)	Substitution paradigmatique (oui +, non -)	Construction In(+)/complète(-)
Segments répétés	-	-	-	-+
Lexical bundles	-	-	-	-+
P-frames/lexical frames	+	-	+	-+
Sentence builders	-+	+	-+	-+
Sentence frames and heads	-+	+	-+	-+
Motifs	-	+	+	-+
Routines	-+	+	+	-+

Tableau 1-2 : Figement syntaxique des notions connexes

c. Aspects sémantiques et fonctionnels

L'aspect sémantique des unités phraséologiques est étroitement lié à leurs caractéristiques idiomatiques et syntaxiques. Toutes les notions évoquées ne sont pas liées au principe d'idiomaticité. Leurs sens sont prédictibles (voir tous les exemples cités dans la section 1.2.3). Ces faits phraséologiques sont susceptibles de s'associer à une ou plusieurs fonctions discursives, textuelles, pragmatiques, etc., et forment donc des « figements propres au genre [...], au domaine » (Sitri & Tutin, 2016, p. 11).

d. Actualisation énonciative, quel niveau ?

À la suite de Benveniste (1970, p. 12), nous entendons par actualisation énonciative la mise en énonciation d'une langue par le locuteur dans une situation donnée. Notre première observation des routines sémantico-rhétoriques a mis en évidence le fait que ces séquences s'actualisent dans le discours, un phénomène observable à travers les indices énonciatifs comme les pronoms, les déictiques, les marques temporelles, etc. Sans aucun doute, ces éléments phraséologiques peuvent réaliser partiellement ou totalement des variations sur les plans syntaxique, sémantique, pragmatique et fonctionnel pour s'inscrire dans la situation d'énonciation, mais il est rare que les chercheurs anglophones considèrent l'actualisation des séquences sur le plan énonciatif qui est plutôt une tradition française dans le domaine de l'analyse du discours. Malgré ce fait, nous pouvons toujours entrevoir une corrélation : l'approche anglophone s'enracine dans une tradition de contextualisation ou de linguistique de corpus couplée souvent avec l'approche fonctionnelle et interpersonnelle (Halliday & Martin, 2003; Hyland, 2005; Biber, 2006; Biber & Barbieri, 2007; Halliday & Matthiessen, 2014; Hyland, 2017; etc.). Leurs études visent à expliquer l'intention du locuteur ou de l'auteur derrière l'usage de telles formes de séquences, ainsi que leur fonction textuelle ou pragmatique dans un contexte précis. Cette dimension interpersonnelle qui était la plupart des études sur les unités phraséologiques est examinée dans une perspective d'analyse du discours par les chercheurs francophones (par exemple Bakhtine, 1984; Bres, 2005, 2017; Bres &

Nowakowska, 2008 ; Grossmann & Tutin, 2010 ; Bres & Dufour, 2014 ; etc.). Ces derniers préfèrent s'inscrire dans les dimensions énonciatives et dialogiques du discours afin de détecter la présence de l'auteur ou d'autrui.

Pour conclure, nous pouvons constater que dans le domaine de la phraséologie, les routines dans notre cas sont différentes des notions discutées plus haut tant sur le français que sur l'anglais. Au final, on voit que, la notion de routines sémantico-rhétoriques est plus proche de celle de motifs. Elle est considérée comme un type spécifique de motifs, mais enracinée dans les écrits scientifiques et argumentatifs, ayant une plus grande variabilité sur les plans syntaxiques et fonctionnels. Nous allons maintenant mener une observation plus minutieuse sur les routines sémantico-rhétoriques.

1.2.6. Observations sur les routines sémantico-rhétoriques

Rappelons d'abord notre définition au début de la partie 1.2.1 pour les routines sémantico-rhétoriques : ce sont des expressions récurrentes relativement longues, autour d'un prédicat verbal avec une flexibilité sur les axes syntagmatique et paradigmatic. Elles peuvent remplir des fonctions rhétoriques spécifiques propres aux écrits scientifiques et argumentatifs et s'actualiser sur le plan énonciatif. C'est un type de motifs plus complexe.

Nous proposons d'observer de plus près les paramètres définitoires en prenant en compte les points communs et les différences entre les motifs et les routines. Nous essaierons aussi de répondre aux questions posées au début de la partie 1.2, à savoir les questions sur la nature phrastique, le degré de flexibilité syntaxique et sémantique, les paramètres énonciatifs et fonctionnels ainsi que le niveau d'abstraction de description des routines.

1.2.6.1. Récurrences des routines

Ce paramètre est d'ordre quantitatif. Il s'intéresse à la fréquence des expressions préfabriquées qui doivent apparaître fréquemment dans les corpus. Comme toutes les notions précédemment présentées (des *segments répétés*, *lexical bundles*, *p-frames* jusqu'aux motifs), le premier critère d'identification des routines réside dans le principe de fréquence des co-occurrences des items lexicaux dans un corpus. Dans la définition lexicométrique de Mayaffre (2008, p. 55), la co-occurrence renvoie au concept de « la co-présence ou présence simultanée de deux unités linguistiques (deux mots par exemple ou deux codes grammaticaux) au sein d'un même contexte linguistique ». En général, si l'empan entre deux co-occurents est plus grand, leur cooccurrence va être moins significative (*ibid.*, p. 57). Pour les faits phraséologiques plus

complexes comme le motif, ou les routines dans notre cas, la cooccurrence dépasse la limite de deux unités lexicales, et s'étend aux expressions contenant trois ou plus de trois mots.

Si ce paramètre est fondamental pour définir une routine, la fréquence concrète d'une expression n'est pourtant pas définie précisément. Nous trouvons plusieurs études récentes pouvant servir de référence (Ji & Tutin, 2019, 2020) : elles ont adopté une fréquence d'au moins 3 ou 4 occurrences dans un corpus d'une taille de 5 millions de mots. On pourrait être surpris que la fréquence adoptée soit insignifiante à la première vue. Mais en regardant des exemples cités dans la partie 1.2.1, en particulier l'exemple 1, nous pouvons voir que les expressions visées (en gras) sont longues et réparties dans le texte dans un empan relativement large.

[1] **L'article s'organise donc autour de 3 parties. La première rappelle** l'intérêt de se focaliser sur les relations sémantiques à partir de leurs traces en corpus pour construire des ressources sémantiques comme les terminologies ou les modèles conceptuels. **La deuxième partie fait état de résultats** en linguistique de corpus relatifs à l'étude des relations conceptuelles et qui nous semblent intéressants pour la modélisation des connaissances. **Enfin, dans la troisième partie, nous montrons** comment ces principes linguistiques guident la spécification d'une méthode et d'un outil d'aide au repérage et à la modélisation des relations conceptuelles : le système Caméléon. (Linguistique)

Cette complexité rend mécaniquement la répétition des routines beaucoup moins fréquentes que les collocations ou les marqueurs discursifs. Par exemple, Tran (2014) utilisait une fréquence de 10 pour les séquences polylexicales. L'application d'une fréquence absolue ne doit donc pas prendre les collocations ou les expressions phraséologiques courtes comme références, mais nécessite une adaptation aux routines après une réflexion sur leur complexité syntaxique, sémantique ainsi que la distance éventuellement présente à l'intérieur de leurs constructions. Dans notre étude, nous préférons détecter les routines par la voie des mesures d'association entre les mots qui permettent d'identifier des cooccurrences des mots statistiquement significatives (Kraif, 2016). Nous en préciserons les détails dans la deuxième partie portant sur la méthodologie. En résumé, le premier critère, qui est aussi le plus essentiel, consiste à faire reposer le repérage des candidats des routines sur une certaine significativité statistique dans le corpus.

1.2.6.2. *Nature lexicale, sémantique et phrastique des routines*

Le deuxième paramètre vise à examiner la nature phrastique des routines. Plus exactement, trois aspects sont analysés ici : la nature des composants des routines sémantico-rhétoriques, les limites des routines, et les types de phrases composant les routines.

D'abord, une routine se construit principalement autour d'un prédicat verbal. Ensuite, ce n'est pas une construction purement grammaticale comme les patrons grammaticaux SVO, ou purement lexicale comme les collocations *prendre en charge*, mais une configuration lexico-syntaxique qui intègre à la fois des parties fixes et variables sur les axes syntagmatiques et paradigmatiques. Nous préférons ici le terme « configuration lexico-syntaxique » dans le but de souligner le fait qu'une routine n'est pas toujours une simple expression, mais une configuration qui peut avoir diverses réalisations lexicales et syntaxiques dans le discours. En d'autres termes, plutôt que des phrases fixes, la plupart des routines sont des groupes de phrases qui correspondent à des constructions syntactico-sémantiques abstraites communes. Prenons l'expression de l'exemple 3 :

nous pouvons constater (dans ce texte) qu'

qui est une réalisation concrète de la routine <Pronom auteur + (*pouvoir|vouloir*) + Verbe de constat + complétive *que*|argument nominal>. Cette routine connaît d'autres réalisations lexicales dans le discours comme [*on|nous*] (*peut|pouvons*) [*noter|observer|relever|remarquer|souligner|mentionner|indiquer|signaler...*]

Les routines sont souvent des phrases incomplètes qui nécessitent un syntagme thématique afin de construire une phrase complète. Par exemple, la routine citée plus haut n'est qu'une phrase incomplète, elle accomplit un contenu propositionnel dans notre corpus toujours de la même façon, soit par un syntagme nominal, soit par une proposition subordonnée introduite par le pronom relatif *que*. Si cette partie est obligatoire dans la production du discours, elle présente une plus grande liberté que les parties fixes et variables des routines. Elle manifeste dans une certaine mesure une contrainte sémantique gouvernée par la cohésion du discours scientifique, mais peu contrainte sur le plan lexical. Même si cette partie ne constitue pas la partie essentielle des routines, sa présence est obligatoire pour les analyser et les modéliser.

En outre, une routine est en mesure d'inclure d'autres items lexicaux facultatifs, ce qui permet d'une part d'élargir la frontière d'une routine et d'autre part de lui accorder sans doute une deuxième fonction discursive. Prenons toujours l'exemple de la routine <Pronom auteur + (*pouvoir|vouloir*) + Verbe de constat + complétive *que*/argument nominal>. Elle a principalement une fonction de (co)constat d'un fait scientifique. En nous ouvrant vers des éléments périphériques ou facultatifs comme des adverbes de concession, de déduction, ou des syntagmes de lieu, etc. nous aurons une routine étendue à la structure <Pronom auteur + (*pouvoir|vouloir*) + Verbe de constat + (Adverbe de concession/de déduction/de réexamen/d'addition) + (lieu textuel ou déictique) complétive *que*/argument nominal>. Cette

structure correspond aux exemples 3 et 6 : *nous pouvons constater (dans ce texte) que et on peut (cependant) remarquer (sur le tableau 6a) que*. La participation de l'adverbe *cependant* équipe la routine d'une valeur de concession, l'intègre ainsi dans le mouvement argumentatif du discours scientifique. Les éléments facultatifs ici *dans ce texte, sur le tableau 6a* manifestent l'évidentialité.

Certes, la plupart des routines ont une construction propositionnelle formée autour d'un prédicat et possèdent une variabilité syntagmatique et paradigmaticque, mais il faut noter qu'il se trouve également des cas exceptionnels qui sont des routines de structure prépositionnelle comme *contrairement à [nos attentes|notre hypothèse|notre proposition| ce que nous avons observé]* (1), *pour (le) dire autrement* (2). La routine (1) *contrairement à nos attentes...* peut avoir une certaine variabilité, mais ne se construit pas autour d'un prédicat verbal. Il s'agit davantage ici d'une structure adverbiale. La routine (2) se manifeste à la fois par la structure prépositionnelle et la lexicalisation quasiment complète sur les plans syntaxique et lexical.

En ce qui concerne les types de phrases des routines, ils sont variables et peuvent comporter des types de phrases comme :

a) Phrase déclarative

- principale : *le terme de X renvoie à*

[9] Étymologiquement, **le terme de médiation renvoie à** « milieu, intermédiaire, moyen ». (Anthropologie)

Dans les phrases déclaratives, on trouve aussi la structure impersonnelle : *il est à noter que...*

[10] **Il est à noter que** la présence de l'inflation dans notre équation interdit la prise en compte simultanée d'effets spécifiques temporels. (Économie)

- subordonnée : *comme nous l'avons démontré au cours de cette étude...*

[11] **Comme nous l'avons démontré au cours de cette étude**, l'armée de l'air a élaboré un dispositif symbolique très sophistiqué. (Science informatique)

- clivée ou pseudoclivée : *c'est bien X qu'il faut remarquer...*

[12] D'autre part, aux racines de l'art moderne, **c'est** bien une incessante manipulation sur le national et l'international **qu'il faut remarquer**, rendant impossible toute approche de l'histoire de l'art en termes d'art national comme international. (Histoire)

b) Phrase injonctive :

- parenthétique qui se trouve souvent dans les parenthèses *voir en annexe*.

[13] Dans le long terme (**voir en annexe**), l'effet est le même qu'à moyen terme puisque la qualité des appartements n'est pas influencée par t (équation A4 en annexe) NNN8. (Économie)

- non parenthétique *voir l'article de... ; prenons un exemple...*

[14] Sur les ambassadeurs urbains, **voir l'article de** Pierre Monnet pour les villes allemandes : Villes, ligues, princes et royauté. (Histoire)

[15] Pour s'en convaincre, **prenons un exemple** dans le domaine de la circulation routière. (Psychologie)

c) Phrase interrogative *faut-il le préciser...*

[16] De plus, **faut-il le préciser**, ce sont les melouk (Mohammed VI et le prince Rachid Moulay) qui ont été les signes oniriques annonciateurs de l'acquisition du pouvoir des génies. [Anthropologie]

La modalité interrogative *faut-il préciser, faut-il rappeler* dans les écrits scientifiques ne consistent pas toujours en une vraie question, mais souvent en un commentaire métadiscursif correspondant à une interrogation rhétorique qui n'attend pas de réponse (Grossmann, 2017, p. 48). *Faut-il préciser* permet de créer un effet de connivence entre l'auteur et le lecteur.

La diversité des types de phrases rend le regroupement des phrases en tant que routine plus complexe. La structure de la phrase peut parfois se modifier, en particulier la modalité. Nous reviendrons sur ce point dans le chapitre 4, section 4.5 pour le regroupement des occurrences des routines.

La plus grande différence entre la notion de routines et celle de motif se manifeste sur la relation syntaxique profonde des items d'une routine : le motif chez Quiniou *et al.* (2012) renvoie aux expressions récurrentes dont la structure syntaxique demeure linéaire²³. Contrairement à ces constructions de surface, les routines dépassent la linéarité syntaxique, mais tissent des relations de dépendance syntaxique profonde entre les constituants, comme celles qui sont représentées dans les modèles syntaxiques de dépendance inspirés de Tesnière (1965). Au lieu d'interroger la syntaxe de surface d'une routine, on s'intéresse à sa syntaxe profonde en observant les rôles sémantiques des constituants. Nous pouvons expliquer ce point en reprenant les exemples *nous reprenons la définition de Duschmoll et la définition adoptée*

²³ En fait, dans le projet Phraseorom du notre laboratoire Lidilem à l'Université Grenoble Alpes, les chercheurs mobilisent la méthode des arbres lexico-syntaxiques (*cf.* section 4.3.2) pour extraire des motifs profonds distincts des réalisations de surface de Quiniou *et al.* (2012). Pour plus de détails, voir à ce sujet l'article de Diwersy, S., Gonon, L., Goossens, V., Kraif, O., Novakova, I., Sorba, J & Vidotto, I. (2021). La phraséologie du roman contemporain dans les corpus et les applications de la PhraseoBase. *Corpus*, 22, 1–22. Par conséquent, les routines sémantico-rhétoriques sont plus proches des motifs chez ces auteurs au lieu de chez Quiniou *et al.* (2012) ou Longrée & Mellet (2009, 2013).

ici est celle de Duschmoll dans l'article de Kraif et Tutin (2017). Même si les deux expressions expriment un sens équivalent, leurs formations syntaxiques superficielles ne le sont pas. Néanmoins, si nous les analysons du point de vue sémantique, nous aurons :

- l'agent : l'auteur des écrits scientifiques (*nous*),
- le prédicat d'emprunt (*reprenons, adoptée*),
- l'objet scientifique (*la définition*),
- la source de l'objet (*de Duschmoll*) (Kraif & Tutin, 2017).

La conceptualisation de la relation sémantique des routines par une relation hiérarchique plus profonde permet de regrouper les candidats routines qui ont des propriétés sémantiques identiques malgré une diversité syntaxique et des réalisations lexicales diverses. Mais cela ne veut pas dire que nous allons ignorer les variations concrètes des routines. En fait, notre objectif est d'établir une modélisation linguistique des routines à un niveau assez abstrait, mais en tenant compte de l'analyse des réalisations concrètes de chaque routine, qui occupe également une place centrale afin de révéler les nuances fonctionnelles et énonciatives de chaque réalisation. Par conséquent, dans notre étude, nous n'avons pas repris le découpage en rôle sémantique pour regrouper les occurrences des routines. Nous avons adopté plutôt un découpage sémantico-syntaxique afin de classer les routines d'une façon inductive (nous y revenons dans la partie méthodologie).

Nous pouvons maintenant répondre aux questions posées sur la nature de phrases des routines : en général, les routines sont des propositions incomplètes ou complètes, qui peuvent intégrer ensemble quatre types d'éléments, illustrés plus bas à l'aide des exemples 17 et 18 (Tutin & Ji, 2022, p. 53) :

- des parties fixes, contraintes sur le plan lexical (en gras),
- des parties obligatoires, mais peu contraintes sur le plan lexical, partiellement contraintes sur le plan sémantique (en italique),
- des éléments facultatifs renvoyant à des mots grammaticaux ou syntagmes non essentiels (souligné),
- des éléments périphériques qui permettent d'intégrer les routines dans le fonctionnement discursif et permettent de co-construire avec les routines l'argumentation du discours scientifique (en encadré).

[18] Néanmoins, **il est important de souligner** que cette validation a été réalisée avec des lycéens âgés de 12 à 18 ans. (Sciences de l'éducation)

[19] [...] comme on peut le voir dans le tableau 4, les élèves de 2 de GT des familles [...] (Sciences de l'éducation)

Ces quatre types d'éléments conditionnent fortement la description linguistique sur les frontières des routines, leurs variations syntagmatiques et paradigmatiques, et enfin leurs types phrastiques.

1.2.6.3. *Fonctions rhétoriques des routines*

Le troisième critère concerne les fonctions rhétoriques spécifiques propres au genre discursif des écrits scientifiques. Autrement dit, pour que ces séries d'expressions récurrentes forment une routine, il faut qu'elles assurent certaines fonctions sémantiques et rhétoriques. Les routines ont tendance à s'associer aux « fonctions rhétoriques spécifiques au genre [du discours scientifique], comme la filiation scientifique indiquée par un auteur, la démarcation par rapport à un pair, l'évaluation des activités scientifiques des pairs, l'élaboration de preuves factuelles », etc. (Tutin, 2014). Dans les exemples cités (1-8), la routine²⁴ suivante *L'article s'organise donc autour de 3 parties. La première rappelle, La deuxième partie fait état de, Enfin, dans la troisième partie, nous montrons* sert à annoncer l'organisation du texte. La routine *J'emprunte l'expression à* a une fonction d'emprunt terminologique, c'est-à-dire que l'auteur réutilise les termes ou les notions des pairs comme « un appui terminologique » (Grossmann, Tutin, & Silva, 2009), afin de signaler son « affiliation à un courant ou à une école [ce qui l'aide] à définir [son] identité du chercheur, à clarifier l'usage des concepts qu'il utilise » (*ibid.*). La routine < Pronom auteur + (*pouvoir|vouloir*) + Verbe de constat + complétive *que*|argument nominal > *comme nous pouvons constater que* exerce une fonction de (co) constat, permet de mettre en relief les informations importantes et les constats essentiels. Il existe encore d'autres fonctions rhétoriques comme la supposition (*on peut [supposer|faire l'hypothèse que]*), la reformulation (*[on|nous] peut|pouvons dire [autrement|aussi]*), pour ne citer que celles-ci.

Outre ces fonctions discursives, les routines peuvent remplir des fonctions interpersonnelles comme l'atténuation (*on peut [considérer|penser] que*), l'affirmation (*il est [certain|évident] que*), ou bien des attitudes comme la surprise (*il est frappant de constater que, il n'est pas [surprenant|étonnant] que*), le jugement (*il n'est pas inutile de | il est sans doute pertinent de*), etc.

Il faut également souligner que les routines sont capables d'exercer plusieurs fonctions à la fois, comme la routine *on peut cependant remarquer que* qui a une fonction de constat et

²⁴ Ici, il s'agit plutôt la réalisation concrète d'une routine de l'annonce du plan du texte.

de concession. La relation entre les routines et les fonctions ne constitue pas donc une relation biunivoque.

1.2.6.4. Paramètres énonciatifs des routines

La question de l'actualisation énonciative des routines constitue un dernier point essentiel pour les définir et les identifier. Certains éléments des routines peuvent s'actualiser au plan énonciatif. Il s'agit souvent des protagonistes de la communication dans le processus de mise en discours des écrits scientifiques. Prenons l'exemple d'une routine d'annonce de l'objectif d'étude :

Objectif	Recherche	Viser à	Marqueurs discursifs	Analyser	Thème
<i>L'objectif de</i>	<i>cet article</i>	<i>est</i>		<i>d'analyser</i>	<i>ces pratiques de certification de la qualité</i>
<i>L'objectif de</i>	<i>cet article</i>	<i>est</i>	<i>précisément</i>	<i>d'explorer</i>	<i>la capacité des personnes à susciter des idées chez les concepteurs d'une bibliothèque numérique...</i>
<i>Les objectifs de</i>	<i>cet article</i>	<i>sont</i>	<i>donc</i>	<i>de tester</i>	<i>l'assurance complète et la conjecture...</i>
<i>L'objectif de</i>	<i>ce papier</i>	<i>est</i>	<i>justement</i>	<i>de présenter</i>	<i>explicitement cette propriété dans une économie...</i>
<i>L'objectif de</i>	<i>cet article</i>	<i>étant</i>		<i>de caractériser</i>	<i>les effets engendrés par une pratique de certification [...], nous nous supposons que...</i>
<i>[...], l'objectif de</i>	<i>cet article</i>	<i>est</i>		<i>de rendre compte de</i>	<i>la montée de nouvelles préoccupations concernant le bois en rivière...</i>
<i>L'objectif de</i>	<i>cet article</i>	<i>est</i>		<i>de montrer</i>	<i>concrètement, [...] la diffusion et l'application des méthodes culturelles flamandes...</i>

Tableau I-3 : L'exemple d'une routine d'annonce de l'objectif d'étude, extrait de notre corpus

Le champ 'Recherche' comporte des éléments qui s'actualisent énonciativement dans le discours. La routine, par l'emploi déictique du déterminant démonstratif *cet|ce* incorpore toute la phrase dans le processus d'énonciation. Ce phénomène d'actualisation, régulier dans les routines, dérive de leur lien avec le dialogisme des écrits scientifiques. Comme Bres (2001) l'indique, le processus de la mise en discours s'accompagne souvent de l'actualisation phrastique. Cette dernière, conçue comme un « passage de la langue au discours », se fait généralement par trois catégories typiques d'opérations dans le discours monologique (Bres, 2001, p. 253) : 1) l'actualisation temporelle, spatiale, personnelle via des anaphores ou des déictiques qui fonctionnent en tant que paramètres de référenciation ; 2) l'actualisation modale ; 3) l'actualisation sémiotique qui ne nous intéresse pas ici. En ce qui concerne les routines

sémantico-rhétoriques dans les écrits scientifiques, elles mettent souvent en jeu les participants du « scénario scientifique » (Tutin & Ji, 2022, p. 53). Plus précisément, il s'agit de l'auteur, du lecteur, de la recherche en cours, du discours en cours ou des discours d'autrui, etc. L'emploi de nombreux éléments déictiques permet d'identifier l'aspect énonciatif des routines :

- L'auteur scientifique : *notre étude porte, comme on l'a dit*
- La recherche en cours : *cette étude porte...*
- Le discours en cours : *comme on le voit ci-dessous...*
- L'auteur et le lecteur : *comme on l'a vu*

Cependant, beaucoup de routines ne s'actualisent pas toujours explicitement. Elles peuvent s'opérer dans les parties obligatoires ou les parties facultatives ; c'est souvent le cas pour des routines de structure impersonnelle : *il est [important|fondamental|essentiel|central] de [souligner|noter|constater...]* Dans ce genre de routine, l'actualisation se réalise souvent dans la partie étendue des routines, c'est-à-dire au niveau de leurs compléments, comme dans les exemples 20-21 :

[20] **Il est à noter que** la jeune artiste a su, parallèlement, se faire aider pour des tâches plus administratives par une association ayant pour mission d'aider les créateurs à monter des projets. (Sociologie)

[21] À cet égard, **il est important de noter que** les observations concernant le vent ne répondaient pas à un désir de connaissance scientifique, mais à des considérations pratiques. (Histoire)

Dans ces exemples, *la jeune artiste*, *les observations* sont bien des anaphores qui s'actualisent dans le complément d'objet, et permettent d'inscrire toute la phrase dans le processus énonciatif, parallèlement aux marqueurs discursifs, aux locutions adverbiales, voire aux déictiques (par exemple *ici*). Dans l'exemple 21, l'élément adverbial *à cet égard*, outre l'inscription énonciative, manifeste l'inscription argumentative des routines dans le discours scientifique. Ainsi, l'opération d'actualisation énonciative des routines s'effectue de manière diverse, soit directement par les pronoms personnels, les déictiques et les anaphores déjà inclus dans les routines, soit de manière implicite à travers les parties étendues comme les compléments ou les marqueurs discursifs. Ces phénomènes énonciatifs sont étroitement liés au dialogisme, assez présent dans le discours scientifique, sur lequel nous reviendrons plus en détail au chapitre 2.

En guise de conclusion, les routines peuvent être définies ainsi : ce sont de phrases ou semi-phrases récurrentes construites souvent autour d'un prédicat dans les écrits scientifiques. Elles comportent des parties fixes, variables tant sur l'axe syntaxique que sur l'axe paradigmatique. Par ailleurs, elles comportent des éléments qui s'actualisent sur le plan

énonciatif et elles opèrent des fonctions rhétoriques spécifiques étroitement liées aux actes scientifiques.

Notre objet d'étude est spécifiquement centré sur les routines de guidage du lecteur. Quelle est la relation entre les routines de guidage du lecteur et les routines sémantico-rhétoriques ? Quel est le statut de la recherche sur les routines de guidage du lecteur dans les écrits scientifiques ? Nous allons répondre à ces questions dans la section suivante.

1.3. Routines de guidage du lecteur dans les écrits scientifiques : état de lieux et définition

Les routines de guidage du lecteur constituent pour nous un type particulier de routines sémantico-rhétoriques, mais qui a pour fonction de guider la lecture. Nous pouvons rendre cette relation d'inclusion plus explicite par la figure 1-2 :

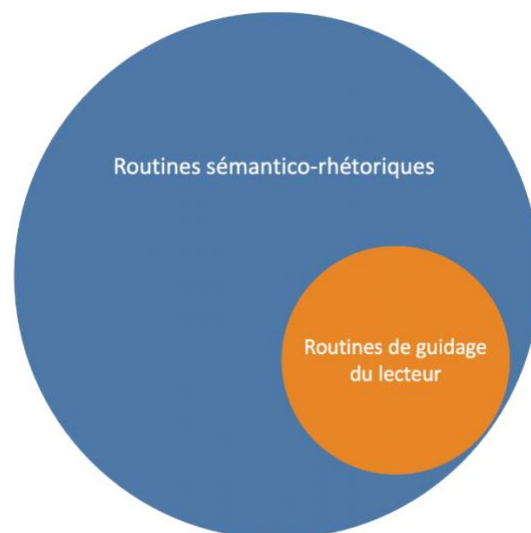


Figure 1-2 : Relation inclusive des routines

Faisant partie des routines sémantico-rhétoriques, les routines de guidage du lecteur possèdent toutes les caractéristiques linguistiques en matière de récurrence, de configurations lexico-syntaxiques complexes, d'actualisation énonciative, mais sont spécifiques en ce qui concerne les fonctions rhétoriques. Rappelons que les routines sémantico-rhétoriques peuvent remplir de fonctions discursives comme la comparaison, l'exemplification, des fonctions de positionnement comme la démarcation, la filiation scientifique, ou des fonctions interpersonnelles pour marquer la surprise, l'évaluation, les sentiments, etc.

L'inclusion des routines de guidage du lecteur dans les routines sémantico-rhétoriques n'est pas inédite. Les routines de guidage ont été évoquées dans plusieurs études sur le lexique

scientifique dans les écrits scientifiques, dans le cadre de multiples projets de recherche²⁵, ainsi que dans les recherches menées au sein du laboratoire Lidilem, à l'Université Grenoble Alpes.

Ces expressions récurrentes favorisant le suivi de la lecture des écrits scientifiques en français ont été évoquées la première fois par Grossmann & Tutin (2010) dans leur étude descriptive sur les verbes de constats (*voir, constater, noter, percevoir, sentir*, etc.). Ces deux auteurs ont souligné la présence des marqueurs de guidage métatextuel dans les écrits scientifiques qui permettent de montrer « la volonté de faire comprendre clairement son intention au lecteur » (*ibid.*). Parmi les marqueurs de guidage, on relève des routines comme *nous examinerons, nous aborderons dans la troisième partie...* Par la suite, Grossmann (2011, 2014a) a identifié à plusieurs reprises des marqueurs ou routines liés au guidage comme les marqueurs *voir, cf., on l'a vu*, ou les phrases en *comme : comme nous l'avons vu dans notre première partie...* Si les premiers travaux cités ici ont noté la présence des marqueurs de guidage dans les écrits scientifiques, sans mettre l'accent spécifiquement sur les routines, les études postérieures menées par Tutin (2014, 2018a, 2018 b) sur les expressions récurrentes dans les écrits scientifiques ont indiqué plus précisément les fonctions de guidage remplies par les routines sémantico-rhétoriques, en dehors des fonctions référentielles et interpersonnelles.

En ce qui concerne les fonctions concrètes qui peuvent guider la lecture, nous pouvons mentionner dans les études citées précédemment, les fonctions liées à la structuration du texte (appelées également fonctions métatextuelles) comme le constat (*on peut constater...*), et le renvoi textuel (*voir la figure X*), etc. Bien que ces études aient fourni des observations intéressantes sur les routines, elles ne sont pas systématiques sur ces éléments phraséologiques dans les écrits scientifiques. Par conséquent, il nous semble important de définir précisément les fonctions de guidages du lecteur. Une définition claire du guidage permet à la fois de cerner la sphère des routines sémantico-rhétoriques à étudier, et d'ancrer l'étude des routines de guidage dans les théories linguistiques plus pertinentes.

Si les routines sémantico-rhétoriques peuvent remplir toutes les fonctions rhétoriques de nature discursive comme la filiation scientifique, la supposition, la reformulation, etc. ainsi que toutes les fonctions interpersonnelles comme l'atténuation, l'affirmation, l'attitude, etc., les routines de guidage du lecteur ont dans l'ensemble une fonction assez pratique : organiser le texte et orienter l'interprétation du discours. Cette fonction de guidage de la lecture suppose

²⁵ Le projet ANR Scientexte entre 2007-2010 : <https://scientext.hypotheses.org/>, et le projet ANR Termith entre 2012-2015 : <https://anr.fr/Projet-ANR-12-CORD-0029>.

que les routines ne cherchent pas à modifier le contenu propositionnel, mais à mieux organiser et expliquer le texte en anticipant les attentes du lecteur afin de rendre le texte plus facile à lire.

En bref, l'étude des routines de guidage du lecteur permet de découvrir les stratégies rhétoriques de l'auteur pour mieux organiser le texte et mieux en expliquer le contenu pour faciliter la lecture tout au long du texte.

1.4. Synthèse

Dans ce chapitre, nous avons vu que le domaine de la phraséologie a connu un développement non seulement dans les objets d'études, mais aussi dans les domaines impliqués comme l'enseignement de la langue, l'analyse du discours, etc. L'extension de la phraséologie permet d'inclure des expressions de natures variées dans son champ d'études, à travers des six critères définitoires sur la nature, le nombre des éléments impliqués dans les unités phraséologiques, la distance et le degré de flexibilité syntaxique entre ces éléments, le nombre d'occurrences ainsi que la compositionnalité sémantique. Ces critères permettent de couvrir la plupart des études portées sur la phraséologie, tant dans l'approche « traditionnelle » comme dans le domaine de lexicologie, que l'approche basée sur l'utilisation ou le corpus comme des études sur les *lexical bundles*, les *p-frames* en anglais, ou les motifs, les routines en français.

Les routines sémantico-rhétoriques s'inscrivent dans la phraséologie étendue, elles constituent un phénomène à définir par rapport aux autres notions dans ce domaine comme les *routines conversationnelles*, les *sentence builders*, les *lexical bundles*, les *p-frames*, les *motifs*, les *routines discursives*. L'observation sur ces notions a montré que les *routines sémantico-rhétoriques* sont proches des *motifs* ou des *routines discursives*, notions développées dans l'analyse du discours en français. Les routines sémantico-rhétoriques renvoient à un type d'expressions qui respectent des quatre critères définitoires : 1) la récurrence ; 2) la composition autour d'un prédicat avec des variabilités syntaxiques et lexicales ; 3) les fonctions rhétoriques spécifiques aux discours étudiés, notamment les discours argumentatifs ; 4) l'inclusion dans le processus d'actualisation énonciative. Les routines sémantico-rhétoriques ainsi définies peuvent inclure l'ensemble d'expressions de natures phrastiques variées. Leurs fonctions sont aussi très variées. On peut trouver des fonctions textuelles, interpersonnelles qui ont été abordées dans les notions phraséologiques en anglais, en particulier les *lexical bundles*.

Les routines de guidage du lecteur constituent un sous-type des routines sémantico-rhétoriques. Elles possèdent toutes les caractéristiques de ces dernières, mais visent à remplir une série de fonctions permettant de mieux organiser le discours et de mieux orienter

l'interprétation du discours. Leur présence peut faciliter la lecture pour le lecteur (ou faciliter le suivi du discours oral pour l'interlocuteur). Elles intègrent donc en parallèle une dimension interactive dans l'approche sociale ou une dimension dialogique dans l'approche énonciative. Notre étude des routines de guidage mettra en évidence ces deux approches fondamentales et complémentaires dans les chapitres 2 et 3.

- l'approche énonciative : les routines de guidage s'inscrivent dans le processus d'énonciation du discours, en particulier le dialogisme, et notamment le dialogisme interlocutif qui traite du fonctionnement interactif entre le discours de l'auteur et celui du lecteur anticipé par l'auteur.
- l'approche fonctionnelle : l'étude des routines de guidage du lecteur doit intégrer un point de vue fonctionnel pour comprendre les fonctions rhétoriques liées au guidage du lecteur.

Chapitre 2. Écrits scientifiques dans l'approche dialogique pour l'étude des routines

Nous allons nous intéresser dans ce chapitre au fonctionnement interactionnel des écrits scientifiques, qui constitue une dimension importante pour étudier les routines de guidage du lecteur. Comme dans de nombreuses formes de communication, l'écrit scientifique, en tant qu'écrit argumentatif, a une fonction dialogique (Bakhtine, 1994). Les écrits scientifiques ne sont désormais pas considérés comme un genre de discours totalement monologique, dans la mesure où l'auteur ne construit pas ses énoncés en ignorant ses lecteurs potentiels. En revanche, pour faciliter la compréhension du texte scientifique destiné aux lecteurs qui travaillent principalement dans le même domaine, l'auteur prend généralement en compte son lecteur et anticipe ses attentes afin de garantir un bon suivi du texte scientifique. L'approche énonciative, notamment dialogique, nous permet de bien comprendre l'étude des routines de guidage dans une approche théorique liée au fonctionnement interactionnel spécifique du genre discursif des écrits scientifiques.

Le présent chapitre examine, dans un premier temps, la définition du discours scientifique à travers la question du genre discursif. L'attention se focalise, dans un deuxième temps, sur les caractéristiques des écrits scientifiques, en particulier des articles de recherche, afin de souligner l'importance du guidage du lecteur. Le dialogisme dans les écrits scientifiques sera développé, dans un dernier temps, pour préciser les caractéristiques énonciatives des routines de guidage du lecteur.

2.1. Le genre des écrits scientifiques

Notre recherche se concentre sur les routines de guidage du lecteur mobilisées par l'auteur tout au long de la rédaction des écrits scientifiques. L'une des questions essentielles à poser porte sur la définition des écrits scientifiques ainsi que sur leurs caractéristiques interactionnelles.

Le terme *discours scientifique* se compose de deux termes *discours* et *scientifique*. Dans l'approche énonciative, le discours, contrairement au concept de langage qui représente un système de signes et de règles relativement stable, s'entend au premier abord comme le résultat de l'utilisation du langage dans une situation de production bien précise. Il englobe donc des énoncés produits et la situation de production (Charaudeau & Maingueneau, 2002, p. 186–187). Plusieurs oppositions doivent être évoquées ici, celles entre *énoncé/texte/discours*, d'une part, et *énonciation/discours*, d'autre part. Guespin (1971, p. 10) signale que l'opposition

énoncé/texte/discours réside dans les différents points de vue pour considérer le texte et l'énoncé : l'énoncé est le « texte [considéré] du point de vue de sa structuration en langage », le discours ajoute à l'énoncé un autre point de vue du « mécanisme discursif qui le conditionne ». L'énonciation diffère du discours par l'inclusion d'un locuteur et d'un auditeur pendant le processus d'énonciation, ce qui forme la deuxième opposition entre *énonciation/discours*.

L'énonciation peut ainsi être vue comme l'action individuelle de la production du discours, et le discours est le résultat obtenu après cette action d'énonciation. Par conséquent, cette utilisation individuelle doit prendre en compte la finalité de communication, son rapport avec le destinataire, les normes discursives au sein d'une communauté discursive pour répondre aux besoins communicatifs. Tous ces éléments peuvent être regroupés dans l'ensemble des conditions de production du discours.

La situation de production est déterminée par le domaine de l'activité humaine où s'inscrit le discours. Elle doit répondre aux conditions spécifiques et aux finalités d'un domaine donné, qui influencent à leur tour le discours sur trois dimensions : le contenu thématique ; le style de langue (choix lexicaux, grammaticaux et phraséologiques) ; la « construction compositionnelle », c'est-à-dire, l'organisation discursive qui est caractérisée par « le type de structuration et de fini d'un tout, type de rapport entre le locuteur, et les autres partenaires de l'échange verbal (rapport à l'auditeur, ou au lecteur, à l'interlocuteur, au discours d'autrui, etc.) » (Bakhtine, 1984, p. 265–269). Ces trois dimensions permettent non seulement la distinction des types discursifs en fonction des domaines d'activités humaines, mais aussi aident à préciser les genres du discours dans un domaine précis²⁶.

Kerbrat-Orecchioni (2009) discute également la question du discours dans son ouvrage *L'énonciation : De la subjectivité dans le langage*, et regroupe ces caractéristiques dans *l'univers du discours* qui comporte :

- 1) la situation de communication, c'est-à-dire des conditions concrètes de la communication - la nature du domaine et des finalités à prendre en compte ;
- 2) les contraintes stylistico-thématiques renvoient aux « caractères thématiques et rhétoriques du discours, c'est-à-dire en gros, les contraintes de 'genre' ». (*ibid.*, p. 17)

²⁶ Ici, nous prenons la distinction de Maingueneau : « type de discours », « genre de discours » (Maingueneau, 1991, p. 213). Le premier représente une catégorisation élémentaire des discours en fonction des secteurs de l'activité discursive, de la cohérence textuelle, comme le discours littéraire, politique, scientifique... Le deuxième sous-catégorise les types de discours en prenant compte « des lieux d'énonciation déterminés » (*ibid.*). Par exemple, pour le discours littéraire, nous trouvons le théâtre, la poésie, le roman, etc.

Nous voyons que Bakhtine et Kerbrat-Orecchioni partagent presque le même concept en discutant de la question du discours, et qu'ils mettent l'accent sur la nécessité et l'importance de préciser les genres du discours dans les études linguistiques du discours, parce qu'un discours non homogène en matière de domaines, de finalités et de textualité poserait de grands problèmes lors de son exploitation et de son analyse, et que le genre discursif détermine « son vocabulaire de construction, ses formes d'organisation, ses contenus attendus, ses modes rédactionnels » (Rastier & Pincemin, 1999, p. 87).

Le type de discours au centre de notre étude est le discours scientifique. Il est donc pertinent de le définir à partir de l'adjectif *scientifique*. Cet adjectif restreint ici la nature du discours vis-à-vis d'autres types de discours tels que le discours politique, littéraire, etc. L'origine de l'adjectif *scientifique* remonte au VI^e siècle où le latin dominait encore l'Europe. Il est inventé par Boèce pour traduire les textes du philosophe Aristote (Benveniste, 1974, p. 250–253), et il a ensuite connu un développement chez Bacon en 1623, dans une acception qui reste celle qui a toujours cours aujourd'hui. Depuis, *scientifique* renvoie à la science moderne qui est basée sur la méthode expérimentale (Malherbe & Pousseur, 1985, p. 115). Il faut noter ici que l'acception du terme *scientifique* peut inclure aujourd'hui toute discipline de la science, non seulement les sciences « dures » ou exactes, mais aussi les sciences humaines ou sociales (Tran, 2014), dans lesquelles s'inscrit notre recherche.

À partir de notre interprétation de la notion de discours ci-dessus, nous pouvons donc préciser que le discours scientifique est un discours qui se produit dans le domaine scientifique, circonscrit par des conditions spécifiques et des finalités précises du domaine scientifique. En d'autres termes, il s'agit d'un « produit [réalisé] dans le cadre de l'activité de recherche à des fins de construction et de diffusion du savoir » (Rinck, 2010, p. 428), qui :

1. Présente un contenu thématique concernant les connaissances scientifiques ;
2. Emploie un langage avec un style propre aux domaines des sciences ;
3. Respecte l'organisation discursive de la représentation de l'activité dans le domaine scientifique.

Sans aucun doute, la première dimension – le contenu thématique, dans notre cas – ne nous concerne pas directement vu que les routines de guidage du lecteur ne participent pas à la construction des connaissances scientifiques propres aux domaines scientifiques. Ces routines agissent plutôt comme un agent favorisant la construction du discours et l'interprétation qui en résulte. Par contre, les deux dernières dimensions sont pour nous d'une grande importance :

d'une part, le style propre au discours scientifique ne se manifeste pas seulement par le choix des lexiques terminologiques, mais aussi par « un sous-langage spécifique, caractérisé par des traits textuels, pragmatiques, syntaxiques et lexicaux remarquables » (Tutin, 2014, p. 27)²⁷, entre autres par des éléments phraséologiques. D'autre part, l'organisation du discours scientifique, déterminée par le type de structuration et la relation auteur-lecteur, régit réciproquement le style du discours scientifique : les attentes du lecteur sur la langue et la façon de structurer le texte conditionnent la production du discours de l'auteur, ce qui se manifeste finalement sur le discours produit.

Ces trois dimensions sont déterminantes dans le concept de genre discursif. Elles permettent de discriminer plus précisément les genres du discours scientifique. Loffler-Laurian (1983, p. 9–12) distingue 6 classes du discours scientifique en fonction de la situation de production, de la finalité visée et de la nature du contenu, en ajoutant l'identité des interlocuteurs, que nous les récapitulons ici dans le tableau 2-1 :

Genres du discours scientifique	Situation de production	Nature du contenu	Finalité visée	Identité des interlocuteurs
Discours scientifique spécialisé ou académique	Revue de recherche ou colloques de recherche	Connaissances issues de recherche propre au domaine des interlocuteurs	Transmission des savoirs, besoin d'être validé par les pairs de l'auteur	Chercheurs dans un domaine donné et partagé.
Discours de semi-vulgarisation scientifique	Revue traitant des sujets appartenant aux différents domaines	Connaissances issues de recherches propres au domaine du destinataire, pas du destinataire	Vulgarisation des savoirs	Destinateur : professionnels spécialistes ; Destinataire : formateurs universitaires.
Discours de vulgarisation scientifique	Presse générale	Connaissances dans un domaine, pas de recherche	Vulgarisation des savoirs basiques	Destinateur : journalistes ; Destinataire : grand public.
Discours de type mémoire, thèse, etc.	Documents relativement longs, traitant un sujet précis	Des connaissances sur un sujet précis dans un domaine scientifique	Présentation de travail de recherche à la fin du master ou du doctorat pour obtenir le diplôme.	Destinateur : étudiants, spécialistes ; Destinataire : jurys, chercheurs spécialistes, enseignants.
Discours scientifique pédagogique	manuels ou ouvrages didactiques	Concerne des connaissances d'initiation ou approfondies rédigées par l'enseignant	Usage des étudiants ou des élèves.	Destinateur : enseignants ; Destinataires : étudiants.

²⁷ Cette citation est le résumé de Tutin (2014) de l'ouvrage de Kocourek (1991) : *La langue française de la technique et de la science*.

Discours scientifique officiel	Formulaires de forme, rapports ou textes fonctionnels	Des informations concernant le déroulement d'expériences	Usage communicatif, administratif ou politique	Destinateur : chercheurs, équipes de chercheurs, administratifs ; Destinataire : instances officielles non spécialisées.
--------------------------------	-------------------------------------------------------	----------------------------------------------------------	------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------

Tableau 2-1 : Typologie du discours scientifique (Loffler-Laurian, 1983, p. 9-12)

Parmi cet ensemble de discours scientifique étendu, seul le discours scientifique spécialisé ou académique nous intéresse. Au sein du discours scientifique spécialisé, il faut aussi distinguer le canal de communication, qu'il soit oral comme dans les textes oraux ou transcrits de communication orale (de colloque scientifique), ou écrits comme dans les articles de revues de recherche. Nous nous concentrons dans ce travail sur le discours scientifique spécialisé écrit, plus précisément, sur les articles de recherche²⁸.

Les auteurs des articles de recherche sont généralement des chercheurs, et de façon remarquable, les destinataires sont également des chercheurs du même domaine scientifique. On observe donc une coïncidence entre le locuteur et l'interlocuteur (Poudat, 2006, p. 51).

La coïncidence locuteur/interlocuteur du discours scientifique académique, et le canal de communication écrit ont un effet sur notre étude des routines de guidage. Le rapport auteur-lecteur occupe une place centrale dans notre étude, parce qu'il caractérise le fonctionnement interactionnel du genre étudié. Le canal de communication via l'écrit (les articles de recherche dans notre cas) va influencer considérablement à la fois les choix linguistiques (dans notre cas, le choix des routines de guidage du lecteur), et la représentation du fonctionnement interactionnel des écrits scientifiques. Cette représentation fournit une ossature théorique permettant d'observer les routines de guidage du lecteur dans une perspective énonciative, en particulier dialogique. Dans cette étude, nous avons utilisé les articles de recherche comme modèle des écrits scientifiques pour l'étude des routines.

2.2. Articles de recherche

Notre objet de recherche porte sur les articles de recherche. Ce genre de discours scientifique a fait l'objet de nombreuses études dans plusieurs perspectives, comme l'*English Academic Discourse* (Swales, 1990 ; Fløttum *et al.*, 2006 ; Hyland, 2001, 2009), le genre discursif (Poudat, 2006 ; Rinck, 2006), les spécificités lexicales discursives (Pecman, 2004 ;

²⁸ Bien que nos observations sur les routines de guidage soient effectuées sur les articles de recherche, nous pensons qu'elles seront valables, au moins en partie, pour des textes plus longs tels que les thèses ou les ouvrages, qui entrent aussi dans le discours spécialisé s'adressant à des pairs.

Biber, 2009 ; Tutin & Grossmann, 2013 ; Tran, 2014) et la comparaison disciplinaire et culturelle (Dahl, 2004 ; Fløttum, Dahl, & Kinn, 2006 ; Hyland & Bondi, 2006 ; Hyland, 2008).

2.2.1. Connaissances scientifiques dans les articles de recherches

Les articles de recherche représentent le genre le plus prééminent du discours scientifique (Hyland, 2009, p. 67). Le prestige des articles de recherche se manifeste à la fois par la création des nouvelles connaissances et le contrôle de la transmission de ces connaissances par des experts (*ibid.*, p. 68). Le contrôle se fait généralement par la régulation des contraintes éditoriales non seulement sur le plan formel, mais aussi sur le plan propositionnel (le contenu). Comme Boure (1998, p. 107, cité par Poudat, 2006, p. 51) le signale :

L'article scientifique obéit à des règles et à des codes particuliers au niveau du contenu (lourd appareillage théorique et méthodologique, présentation de résultats nouveaux ou de synthèses critiques de l'état des connaissances dans un domaine particulier du savoir) et de sa forme (importance du paratexte – notes, références bibliographiques, annexes, tableaux, schémas, recours à un style impersonnel et utilisation d'un vocabulaire spécialisé).

Ces conventions formelles et propositionnelles diffèrent en fonction des domaines qui concernent principalement les trois grands champs des sciences « dures », des sciences sociales et des sciences humaines. Ces trois champs présentent des particularités sur le plan des connaissances, schématiquement mises en évidence par Hyland (2009, p. 63) :

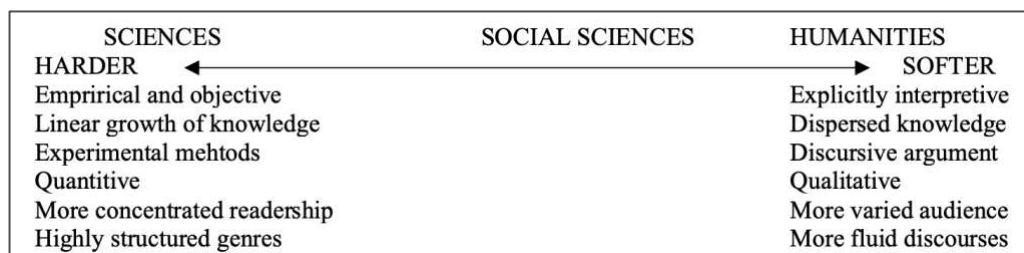


Figure 2-1 : Le continuum des connaissances scientifiques (Hyland, 2009, p. 63)

Les connaissances dans les disciplines des sciences humaines ont tendance à se fonder explicitement dans une démarche interprétative, en s'appuyant sur des analyses qualitatives et sur l'intersubjectivité, celles des disciplines des sciences dures semblent plus empiriques et objectives et s'appuient de préférence sur des observations quantitatives (*ibid.*), tandis que les sciences sociales se situent au milieu, et demandent d'une part une interprétation explicite, d'autre part une démonstration plus objective, recourant parfois à des méthodes mathématiques, comme l'économie (Fløttum, Dahl, & Kinn, 2006, p. 21). Par exemple, dans l'étude de Fløttum *et al.* (2006, p. 57), les articles de recherche en économie et en linguistique sont similaires dans

l'utilisation des éléments métatextuels et de références, mais ils sont très différents des articles de médecine. Les sciences humaines et sociales sont plus herméneutiques avec des interprétations plus subjectives (*ibid.*), et suivent donc des conventions formelles et propositionnelles très différentes des sciences dures, non seulement sur la structure formelle du texte, mais aussi sur l'organisation du contenu plus largement tributaire de l'emploi d'éléments linguistiques à fonction discursive et rhétorique. On peut ainsi faire l'hypothèse que les articles en sciences humaines et sociales mobilisent plus de routines pour organiser le discours. Plus les disciplines sont proches des sciences dures, plus l'organisation discursive sera explicitée, par exemple, l'économie ou la psychologie. Ce point sera discuté dans le dernier chapitre.

2.2.2. Organisation des articles de recherches

En général, pour les articles de recherche dans les sciences dures, on décrit l'organisation des contenus du texte par le schéma Introduction, Méthode, Résultats Et Discussion (IMRED²⁹), un processus qui suit la macrostructure conceptuelle des articles (Swales, 1990, p. 133). Néanmoins, les articles de recherches en SHS ne suivent pas forcément, et même rarement, cette structure prototypique. Leurs structures sont plus flexibles, mais respectent les règles éditoriales suivantes (Poudat, 2006, p. 54) :

- le titre ;
- le(s) nom(s) de l'auteur et leur filiation ;
- le résumé avec les mots-clés ;
- le corps de l'article qui peut comporter l'introduction, le développement, la conclusion.
- les notes de bas de page et la bibliographie.

Ces éléments peuvent être divisés en deux types : les éléments purement textuels comme les résumés et le corps de l'article et les éléments du périphrase comme le titre, les notes de bas de page et la bibliographie. En outre, les articles de recherches peuvent également utiliser des représentations graphiques comme les tableaux et les figures.

On peut voir que, le corps des articles de recherche en SHS, malgré la flexibilité de la structure, comporte des sections identifiables comme l'introduction, le développement ou l'analyse ou la discussion et la conclusion (*ibid.*). On verra dans le dernier chapitre que certaines routines se trouvent souvent dans une partie textuelle spécifique.

²⁹ En anglais, on emploie plutôt le sigle IMRAD (Introduction, Material and methods, Results And Discussion). Ce schéma a été proposé par Bruce(1983, cité par Swales, 1990, p. 133).

À l'intérieur de ces sections textuelles, on peut créer des sous-sections en fonction du besoin réel de la démonstration de la recherche. Le découpage du contenu propositionnel en sections textuelles permet au lecteur de mieux suivre le processus argumentatif et les procédés rhétoriques, qui sont plus centraux pour les articles de recherche en SHS qu'en sciences dures comme on l'a vu plus haut. L'aspect rhétorique des articles de recherche en SHS est fondamental pour la réussite de la tâche de persuasion. Il ne faut pas oublier que « l'objectif rhétorique définitif de l'article de recherche est de convaincre le public destiné au point que cet article devient une partie intégrante du fond scientifique d'un domaine particulier » (Fløttum *et al.*, 2006, p. 30). En outre, la taille des articles de recherche est souvent limitée pour des raisons de publication. Dans ce cas, bien guider le lecteur dans un texte argumentatif condensé devient très important. Cela demande à l'auteur de faire des efforts afin de rendre l'écrit plus lisible pour son lecteur. Ce style *reader-friendly* (Hyland, 2005, p. 44) mettant en évidence l'interaction entre l'auteur et le lecteur, peut être examiné à travers les traces linguistiques de la présence du locuteur-scripteur, en suivant l'approche dialogique, et plus largement l'approche énonciative française.

La présence de l'auteur est primordiale pour l'étude énonciative des écrits scientifiques (Poudat, 2006 ; Rink, 2009 ; Fløttum *et al.*, 2006), mais le destinataire constitue également un élément incontournable pour étudier les écrits scientifiques. Les rôles de l'auteur, inhérents au genre discursif des articles de recherche, se font selon la représentation collective de la communauté scientifique, qui est elle-même préétablie dans l'interaction entre l'auteur et le lecteur.

2.3. Les écrits scientifiques : un lieu de dialogisme

Dans cette partie, nous voudrions nous attarder davantage sur les propriétés énonciatives – et en particulier dialogiques – des écrits scientifiques, qui expliquent l'interaction auteur-lecteur dans une perspective discursive. Nous adoptons ici un concept central – le dialogisme introduit par Bakhtine, et développé ensuite dans la sphère praxématique, qui s'enracine dans le laboratoire Praxiling dans une perspective d'analyse du discours.

Il faut, avant de traiter la notion de dialogisme dans les écrits scientifiques, distinguer deux concepts à la fois associés et distinctifs : la monologalité et la dialogalité, vu que ces deux concepts influencent au premier abord la représentation de la structure externe du discours – soit l'organisation macrotextuelle du discours, et ceci d'une façon importante (Bres, 2005, p. 55), et affecte ensuite la structure interne – l'arrangement microtextuel sur « les niveaux lexico-sémantique et syntaxique de l'énoncé » (Bres & Dufour, 2014, p. 60).

2.3.1. Monologalité vs dialogalité

Le concept de monologalité est souvent traité en lien avec celui de dialogalité. Les deux concepts se rapportent aux modes d'énonciation, modes de structuration et contraintes cognitives du discours étudié (Colletta, 2013, p. 25). La dialogalité renvoie au texte dialogal qui se définit « par l'alternance des locuteurs qui détermine les frontières des différents tours de parole » comme des conversations, des interviews, des débats, des pièces de théâtre (Bres & Dufour, 2014, p. 60), qui peut être observée à travers des « phénomènes de rétroaction immédiate » (Kerbrat-Orecchioni, 2005, p. 17). Le paramètre décisif des propriétés du texte dialogal serait la forme prototypique de va-et-vient conversationnel entre plus d'un interlocuteur qui influence la structure externe du texte, et forme un « dialogue externe » (Bres, 2017). Il s'agit d'un type d'interaction qui représente « la manifestation par excellence de l'interactivité » où la maximalisation de l'impact exercé par la présence physique du destinataire sur l'énoncé est mise en avant (Kerbrat-Orecchioni, 2005, p. 15). La présence d'un interlocuteur réel est donc indispensable pour s'engager dans le processus de l'interaction continue et adapter les comportements mutuels par négociation constante (Kerbrat-Orecchioni, 2005, p. 19). Ainsi, le sens d'un discours dialogal peut être co-construit par son locuteur et l'interlocuteur, à titre d'exemple : une conversation face à face, ou par téléphone dont les interlocuteurs communiquent sur-le-champ. Cette conversation va se manifester, dans la structure externe, par une reprise de parole tour à tour entre les interlocuteurs. Chaque interlocuteur peut couper la conversation pour signaler son incompréhension ou poser des questions qui surgissent au fil du discours.

En revanche, dans le cas du discours monologal, on est surtout confronté à un type de texte qui est défini « comme un objet linguistique qui matérialise un acte de communication en se présentant sous la forme d'un ensemble complet, organisé et cohérent d'énoncés » (Colletta, 2013, p. 25). Il s'agit d'« un tour de parole d'une sorte particulière [dont] les répliques antérieure et ultérieure sont absentes de la structure externe – le texte ne se présente pas sous la forme d'un enchaînement de tours – mais elles n'en affectent pas moins la structure interne » (Bres & Dufour, 2014, p. 60), soit un « dialogue interne » (Bres, 2017). L'absence de va-et-vient et de réplique implique l'absence d'interlocuteur réel sur le champ. Mais cela n'exclut pas la présence d'un interlocuteur virtuel que le locuteur peut imaginer lors de la communication via l'écrit.

À l'instar de Maingueneau (2016, p. 48), on considère que tout discours, monologal ou dialogal, est de nature interactive :

Toute énonciation, même produite sans la présence d'un destinataire, est en fait prise dans une interactivité constitutive (on parle aussi de dialogisme).

La présence d'un destinataire réel ne peut pas constituer un facteur décisif pour définir la nature interactive du discours. Par contre, elle constitue un facteur central pour l'interaction verbale entre les interlocuteurs. Cette interaction reflète la caractéristique fondamentale de l'énonciation : « l'accentuation de la relation discursive au partenaire », pour reprendre l'expression de Benveniste (1974, p. 85). Cette relation discursive est tracée par le « cadre figuratif de l'énonciation » (*ibid.*), c'est-à-dire, l'auteur et le lecteur dans notre cas.

Précisément, accepter le fait que les écrits scientifiques sont interactifs implique d'examiner ces écrits scientifiques à partir de la projection de la perception, de l'intérêt et des attentes du lecteur (Hyland, 2005, p. 11). Néanmoins, il faut préciser les rôles joués par l'auteur et le lecteur dans les écrits scientifiques. L'auteur de l'écrit scientifique assure le rôle d'un protagoniste réellement présent tout au long de son texte, alors que son lecteur, dans une certaine mesure, ne peut pas réagir directement. Mais cela ne signifie pas qu'il ne participe pas à la construction du sens. Bien au contraire, le lecteur se manifeste ici comme un agent important qui décide de la manière de réaliser l'énonciation, les choix lexicaux et les stratégies rhétoriques mises en place. Dans cette optique, l'auteur est obligé d'anticiper sans cesse les attentes du lecteur. Pour y parvenir, il lui est nécessaire d'établir préalablement un destinataire intentionnel, d'anticiper les compétences langagières, culturelles et discursives de ce destinataire, afin de lui permettre de déchiffrer le texte (Maingueneau, 2016, p. 41). Savoir anticiper les compétences du lecteur demande à l'auteur des écrits scientifiques d'avoir une bonne connaissance des compétences langagières et culturelles de la communauté scientifique, dans laquelle le lecteur visé s'inscrit. D'un autre point de vue, l'auteur des écrits scientifiques doit également disposer de ces compétences afin de les mobiliser lors de la rédaction pour que sa production du texte soit validée par la communauté scientifique, et qu'une bonne interprétation du côté lecteur soit assumée.

Nous constatons que les écrits scientifiques, codés via des signes graphiques qui forment des énoncés écrits, avec la présence de l'auteur (destinateur selon le terme de Jakobson, 1963) et l'absence de l'interlocuteur réel (destinataire réel), peuvent être considérés comme des discours monologiques. Cependant, au niveau de leur structure interne, les écrits scientifiques sont parfaitement dialogiques.

2.3.2. Le dialogisme dans les écrits scientifiques

Le dialogisme dans les écrits scientifiques a déjà été constaté par de nombreux chercheurs (Rinck & Tutin, 2007 ; Rinck, & Grossmann, 2009 ; Grossmann & Tutin, 2010). Rappelons ici que le discours scientifique est un genre de discours visant la validation et la transformation des connaissances dans la communauté scientifique. En ce sens, l'auteur des écrits scientifiques cherche également la validation de son écrit par la communauté scientifique. Pour ce faire, l'auteur est censé prendre en considération ses lecteurs potentiels, ce qui motive ainsi le dialogisme³⁰ dans les écrits scientifiques (Grossmann & Tutin, 2010).

Parmi les découvertes de Bakhtine, la plus importante a consisté à identifier que le dialogisme constitue un caractère fondamental du discours.

Intentionnellement ou non, chaque discours entre en dialogue avec les discours antérieurs tenus sur le même objet, ainsi qu'avec les discours à venir, dont il pressent et prévient les réactions (Todorov, 1981, p. 8).

Ce caractère permet de mettre en avant le fait que le discours se produit et se développe entre des sujets, et d'inclure ainsi les pôles participants du discours dans les études linguistiques du discours ; dans notre cas, ce seront l'auteur et le lecteur des articles de recherche. Comme dans tout écrit monologique, les caractéristiques dialogiques mobilisées dans ce genre d'écrits scientifiques se manifestent par le fait que l'auteur des écrits scientifiques doit veiller à l'orientation de son discours vers les discours d'autrui et des siens.

Cette orientation se réalise à travers les trois dimensions qui sous-tendent les relations entre des énoncés (Charaudeau & Maingueneau, 2002, p. 175) ou les différentes rencontres entre des discours (Bres, 2017, p. 3) : la dimension interdiscursive, la dimension interlocutive et la dimension intralocutive³¹.

2.3.2.1. *Dialogisme interdiscursif*

Le dialogisme interdiscursif implique l'interaction entre des énoncés de l'auteur et ceux des autres. Il concerne les échanges avec la communauté discursive à laquelle appartient

³⁰ Ce concept, emprunté à l'œuvre de Bakhtine, jouxte souvent le terme polyphonie dans le champ linguistique. Nous avons choisi le terme dialogisme au lieu du terme polyphonie pour deux raisons : 1) Le terme polyphonie s'occupe de la question de « la superposition de plusieurs voix » dans un discours (Ducrot, 1984, p. 183), par exemple : un roman qui met en scène plusieurs énonciateurs. 2) Le dialogisme concerne « l'orientation de tout discours » (Bres, 2005, 2017) qui concerne l'interaction entre des discours, des interlocuteurs. Il met l'accent sur la dépendance de l'auteur du discours sur les discours antérieurs et sur les anticipations des réponses de ses lecteurs.

³¹ Bakhtine a relevé deux dimensions pour analyser la question du dialogisme : le dialogisme interdiscursif et le dialogisme interlocutif. Bres (2005) a développé la théorie de dialogisme et proposé une troisième dimension pour intégrer le discours sur le discours même de l'auteur : le dialogisme intralocutif.

l'auteur par le renvoi aux discours scientifiques déjà réalisés par ses prédécesseurs autour d'un même objet. Maingueneau (2021, p. 16) souligne que « le discours ne prend sens qu'à l'intérieur d'un immense *interdiscours* ». Cet interdiscours est au fondement du discours, il est « l'ensemble des unités discursives (relevant de discours antérieurs du même genre, de discours contemporains d'autres genres, etc.) avec lesquelles un discours *particulier* entre en relation implicite ou explicite » (Charaudeau & Maingueneau, 2002, p. 324). Ici, le *discours particulier* signifie le discours en cours d'énonciation.

Plus précisément, en émettant ses propres énoncés sur un objet, le sujet du discours ne peut pas s'empêcher d'entrer en interaction avec des énoncés réalisés précédemment par d'autres auteurs sur ce même objet (Bres, 2005, 2017). Dans ce processus, le sujet du discours interagit constamment avec les autres sujets : il présente, avec ses propres énoncés, en traversant « un milieu d'expressions et d'accents étrangers » (Bakhtine, 1978, p. 101), son attitude de consentement, de démarcation ou d'objection sur les points de vue autour d'un même objet ; il exprime ses émotions ou ses évaluations, et donne enfin sa propre compréhension quant à un objet choisi, son image discursive. Comme le dit Bakhtine (1978, p. 102) : « sur toutes ses voies vers l'objet, dans toutes les directions, le discours en rencontre un autre, « étranger », et ne peut éviter une action vive et intense avec lui ».

Ce propos montre l'essence du dialogisme interdiscursif : il s'agit d'un dialogue interne réalisé au travers de la rencontre entre des énoncés du sujet et des énoncés précédemment produits dans le temps. De plus, la dimension interdiscursive illustre également l'aspect social du discours : même si les énoncés véhiculent différents points de vue et émotions du sujet, celui-ci est obligé d'employer un système de langage partagé par la communauté cible, afin de garantir la compréhension de son propre discours.

Les écrits scientifiques sont fondamentalement interdiscursifs. L'auteur des articles en SHS construit son propre discours en se basant sur les discours déjà là dans la discipline. Il s'agit d'une circulation des connaissances scientifiques qui permet à « une réinterprétation permanente de l'ensemble discursif qui constitue cette communauté [scientifique et disciplinaire] » (Boch, Rinck, & Grossmann, 2009, p. 24). On parle également de « mémoire discursive »³² d'après Courtine (1981) qui a développé cette notion à partir des travaux de M. Foucault, une notion utile pour comprendre la dimension interdiscursive dans les articles de recherche :

³² Certains linguistes utilisent le terme « mémoire interdiscursive », comme Adam (2006).

La notion de mémoire discursive concerne l'existence historique de l'énoncé au sein de pratiques discursives réglées par des appareils idéologiques, elle vise ce que Foucault (71, p. 24) relève à propos des textes religieux, juridiques, littéraires, scientifiques, « discours qui sont à l'origine d'un certain nombre d'actes nouveaux, de paroles qui les reprennent, les transforment, ou parlent d'eux, bref, les discours qui indéfiniment, par-delà leur formulation, sont dits, restent dits et sont encore à dire ». (Courtine 1981, p. 52-53)

Les écrits scientifiques dépendent largement de cette mémoire discursive pour traiter une question de recherche. À travers un dialogue interne avec « l'ensemble d'évidences, de croyances, de représentations, d'argumentaires » (Amossy, 2005, p. 67), ils s'inscrivent dans la communauté scientifique et disciplinaire et se démarquent des autres discours traitant la même question de recherche, par leur positionnement. Le système des citations dans les écrits scientifiques est ouvertement interdiscursif. L'auteur des écrits scientifiques peut marquer son positionnement de rejet, de désaccord, de filiation, d'adhésion ou de nouvelle proposition par des routines de citations (*en prenant l'appui sur* quelqu'un... ; *conformément aux résultats obtenus par* quelqu'un ; *nous rejoignons le point de vue de* quelqu'un) (Florez, 2013, p. 73-78).

De plus, il faut respecter les règles linguistiques du discours scientifique (Amossy, 2005, p. 67), reformulées par Grossmann & Tutin (2010), comme un « pré-construit mobilisé au sein d'un champ disciplinaire » pour que ses propres idées soient bien comprises par les destinataires. En d'autres termes, l'auteur des écrits scientifiques ne construit pas son propre discours sans se référer aux fondements théoriques déjà établis ou aux recherches précédemment menées. Au contraire, il doit faire appel aux connaissances disciplinaires et échanger constamment avec la communauté discursive afin de spécifier sa propre recherche.

2.3.2.2. *Dialogisme interlocutif*

Si on dit que le dialogisme interdiscursif attribue les énoncés à un tiers, le dialogisme interlocutif va être attribué à l'interlocuteur (Grossmann & Tutin, 2010). Par conséquent, la dimension interlocutive du dialogisme repose sur une prémisse importante : la prise en compte de l'interlocuteur, même s'il est virtuel. Il est à noter que le discours est avant tout un acte de langage, et un acte de langage dépend non seulement du sujet parlant, mais aussi de l'interlocuteur. Comme le constate Benveniste (1974, p. 241) en parlant du discours, et que nous avons déjà cité :

Toute énonciation supposant un locuteur et un auditeur, et chez le premier l'intention d'influencer l'autre en quelque manière.

Bien sûr, Benveniste n'est pas le seul qui découvre ce caractère intentionnel lors de la mise en œuvre de l'énonciation, Todorov (1981, p. 70), également en examinant les œuvres de Bakhtine, note que :

L'énoncé se construit entre deux personnes socialement organisées, et s'il n'y a pas d'interlocuteur réel, on le présuppose en la personne du représentant normal, pour ainsi dire, du groupe social auquel appartient le locuteur. Le discours est orienté vers l'interlocuteur, orienté vers ce qu'est cet interlocuteur.

Todorov précise donc que « l'énoncé n'est pas l'affaire du seul locuteur, mais le résultat de son interaction avec un auditeur, dont il intègre par avance la réaction » (*ibid.*, 1981, p. 69–70). En ce sens, le sujet et l'interlocuteur constituent deux paramètres centraux dans l'analyse du discours. Le rapport tissé entre le sujet et l'interlocuteur forge un espace intersubjectif où les deux participants se rencontrent pour se comprendre, ce qui rend importante l'intercompréhension entre ces deux protagonistes. Force est de constater que bien que le dialogisme interlocutif pose l'interlocuteur au cœur du discours, il n'exige pas la présence d'un interlocuteur réel. Le sujet peut toujours se faire une image de ses interlocuteurs afin de pouvoir anticiper leurs réactions (Amossy, 2005, p. 68).

Comme un dialogue virtuel, la dimension interlocutive prend en charge le dialogue avec le lecteur par le biais de « la réponse d[u lecteur] [que l'auteur] sollicite et sur laquelle il anticipe » (Brès, 2017, p. 3). Dans les écrits scientifiques, l'auteur fait face à un pair virtuel qui n'est pas présent physiquement, mais plutôt discursivement, en particulier à travers des dispositifs d'orientation de la lecture. Autrement dit, l'auteur des écrits scientifiques doit anticiper sans cesse la compréhension-réponse de l'interlocuteur en faisant une image de ses lecteurs potentiellement adressés (Bres, 2005, p. 53). L'orientation de la lecture s'incarne souvent à travers des fonctions rhétoriques spécifiques, par exemple la bonne volonté de l'auteur pour le guidage du lecteur réalisé via les marqueurs de guidage, l'anticipation des attentes du lecteur, la négociation et la co-construction du sens (Grossmann & Tutin, 2010). Dans notre cas, la dimension dialogique est centrale pour la question des routines de guidage du lecteur. Cette dimension s'incarne souvent à travers des fonctions rhétoriques spécifiques, en particulier celles qui reflètent la volonté de l'auteur de guider le lecteur, par exemple dans des routines de (co)constat. L'expression *nous voyons bien ici que* invite ainsi le lecteur à co-constater avec l'auteur un fait, afin d'engager le lecteur dans l'argumentation. Toutefois, certaines fonctions rhétoriques peuvent appartenir à plusieurs dimensions dialogiques en même temps. Par exemple, la routine de réparation *cela ne veut pas dire que* implique au premier abord la dimension intralocutive, car elle rectifie ou invalide le sens ou le point de vue de

l'argument précédemment produit par l'auteur lui-même. Elle concerne également la dimension interlocutive en répondant à l'anticipation de la mise en cause potentielle chez le lecteur pour éviter un malentendu latent.

2.3.2.3. *Dialogisme intralocutif*

Au-delà de ces deux dimensions du dialogisme, Bres (2005) a développé ensuite, dans la sphère praxématique qui s'enracine dans le laboratoire Praxiling, une autre dimension : le dialogisme intralocutif. Aussi nommée autodialogisme par Authier-Revuz (1995), cette dimension correspond au fait que l'auteur lui-même est tout d'abord son premier lecteur, ses énoncés interagissent constamment avec ses propres énoncés antérieurement dits et ultérieurement produits. Elle concerne donc le discours de l'auteur lui-même, et reflète sa réflexion sur son propre discours, caractérisée par la réflexivité métalinguistique ou métaénonciative. Plus concrètement, l'auteur peut nommer un phénomène ou un fait scientifique par un nouveau terme (*j'appelle X Y*), ou corriger ce qu'il a dit précédemment (*je ne veux pas dire que...*).

2.3.2.4. *Croisement des dimensions dialogiques*

Ces trois dimensions ne sont pas étanches et peuvent apparaître simultanément dans un seul énoncé (Bres, 2017), comme dans le cas des énoncés reformulatifs (*en d'autres termes, c'est-à-dire, pour le dire autrement*) qui peuvent être à la fois intralocutifs et interlocutifs. Ils appartiennent à la dimension intralocutive, par le fait que l'auteur les utilise pour reformuler ses propres énoncés précédemment produits, mais ils entrent aussi dans la dimension interlocutive car la reformulation a pour objectif de clarifier une information éventuellement difficile pour le lecteur, ce qui reflète donc l'anticipation des répliques ou objections du lecteur. Dans ce cas, nous pouvons également parler de fonctionnement dialogique interlocutif-anticipatif (Bres, Nowakowska, & Sarale, 2016) : l'auteur anticipe les réactions potentielles du lecteur en fonction de ses connaissances sur son destinataire, et effectue ensuite soit un acte d'explication par la voie de la reformulation, soit un acte de correction par la réparation d'un énoncé X, afin de répondre par avance aux questions posées ou de dissiper le malentendu éventuellement provoqué à la suite de cet énoncé X.

Dans notre recherche, les routines de guidage du lecteur ont pour objectif la conduite de la lecture à travers une interprétation aussi proche que possible de ce que l'auteur veut exprimer. Elles s'inscrivent davantage dans la dimension interlocutive, parce que leurs emplois sont souvent mobilisés par l'anticipation des attentes du lecteur. Cependant, elles ont également à faire avec les autres deux dimensions. La dimension intralocutive peut être considérée comme

interlocutive anticipative dans la mesure où l'auteur adapte son propre discours afin de manifester la clarté de ses idées, positionnements, attitudes ou sentiments. La dimension interdiscursive permet à l'auteur de renvoyer le lecteur aux autres études pertinentes qui soutiennent son raisonnement, ce qui manifeste donc aussi la prise en compte du lecteur.

En résumé, les conceptions bakhtiniennes et praxémiques autour du dialogisme du discours intègrent dans le discours des protagonistes qui ne sont pas pris en compte dans les modèles linguistiques classiques, et reposent sur l'idée que le discours existe et se développe entre différents sujets, impliquant l'interaction entre les discours de ses sujets. Les écrits scientifiques ne peuvent pas échapper à ce principe.

Les études sur le dialogisme se réalisent par le biais d'observation sur les traits linguistiques, en particulier par les pronoms personnels dans les articles de recherche (Fløttum *et al.*, 2006 ; Rinck, 2006). Ces études permettent de comprendre les rôles de l'auteur et du lecteur, qui sont centraux pour notre étude et nous intéressent dans la section suivante.

2.3.3. Les rôles de l'auteur et l'enrôlement du lecteur dans les articles de recherches

Les rôles de l'auteur des articles de recherche sont tributaires de la représentation socio-discursive de la communauté scientifique. En parallèle, le lecteur des articles de recherche est absent physiquement, mais peut être présent discursivement.

2.3.3.1. Les rôles de l'auteur dans les articles de recherche

L'auteur peut remplir plusieurs rôles dans les articles de recherches, comme le montre l'étude menée par Fløttum *et al.*, (2006) sur l'emploi des pronoms personnels à la première personne (*I, we, je, nous...*) dans les articles de recherche en linguistique, en médecine et en économie dans les langues anglaise, française et norvégienne. L'étude a montré que l'auteur des articles de recherches peut exercer quatre rôles rhétoriques : l'énonciateur du discours, le chercheur, l'argumentateur et l'évaluateur. Ces quatre rôles peuvent être identifiés à partir des verbes³³ combinés avec les pronoms personnels revoyant à l'auteur, mais aussi par le biais des éléments métatextuels qui les entourent³⁴ (*ibid.*, p. 83).

³³ Fløttum (2003b, 2003a) ont proposé des classifications des verbes avec des critères lexico-sémantique et sémantico-pragmatique.

³⁴ Dans l'optique de l'identification des rôles d'auteur, la présence d'un verbe est plus décisive que les éléments métatextuels pour identifier le rôle des pronoms auteurs (à la première singulier). Nous reprenons l'exemple de Fløttum *et al.* (2006, p. 90) : *In this paper I argue that ...* Le verbe *argue* est un verbe d'argumentation, malgré la présence du marqueur *in this paper*, le rôle d'argumentateur est plus fort que le rôle d'énonciateur du

La mission de l'énonciateur du discours consiste à présenter explicitement les activités de représentation verbale ou graphique ou à structurer le texte (*ibid.*, p. 85). Ce rôle d'énonciateur s'accomplit par l'emploi des verbes discursifs destinés à mettre en scène la voix de l'auteur et la présentation de son travail, comme *synthétiser, discuter, présenter, décrire*, etc., ou des verbes permettant de structurer le texte, tels que *se focaliser, retourner, conclure*. Concernant les éléments métatextuels, il s'agit des marqueurs permettant de structurer le texte comme *supra, article, ci-dessus, maintenant* (la liste n'est pas exhaustive) (*ibid.*, p.163).

Le rôle de chercheur vise à présenter les activités scientifiques effectuées pour la recherche. Il se réalise avec la présence des verbes de recherche qui sont liés à l'acte ou à l'activité du processus de recherche, comme *analyser, examiner, tester, utiliser, comparer, explorer* (*ibid.*, p. 84). Ce rôle est le plus présent dans les articles de recherche, avec un taux de 46 %³⁵ dans la distribution des rôles d'auteur (*ibid.*, p. 91). Cela est dû au haut degré de crédibilité et d'objectivité réclamé par les articles de recherche. Dans ce rôle de chercheur, l'auteur scientifique trouve nécessaire de rapprocher son *éthos scriptural*³⁶ de « l'éthos de la science » (Merton, 1942 repris par Amossy, 2015, p. 189) par le biais des dispositions linguistiques comme l'impersonnalité, autrement dit, l'effacement énonciatif, afin d'obtenir l'adhésion à la communauté scientifique composée de ses lecteurs cibles ainsi que leur reconnaissance en tant qu'auteur scientifique fiable.

Le rôle d'argumentateur concerne le positionnement et l'argumentation de l'auteur. Il est identifiable par des verbes de positionnement, comme, *argumenter, proposer, rejeter...* Le rôle d'évaluateur est lié à l'aspect émotionnel et évaluatif de l'auteur. Ce rôle peut être identifié par les verbes permettant d'évaluer (*feel, find something* en anglais) ou par les constructions permettant de communiquer des émotions (*be content to, be struck by*).

L'étude de Fløttum *et al.* (2006) montre que le rôle de chercheur est le plus présent, suivi par celui de l'énonciateur du discours. Cela correspond aux particularités des articles de recherche. D'une part, les articles de recherche visent principalement à la démonstration des recherches effectuées dans le monde matériel. En ce sens, l'auteur est a priori un chercheur.

discours. Mais dans l'exemple *in this section, I argue that...* l'auteur peut jouer deux rôles en même temps : l'argumentateur et l'énonciateur du discours.

³⁵ Fløttum *et al.* (2006, p. 91) recensent la distribution des six rôles d'auteur du pronom personnel à la première personne singulière dans 450 articles de recherche : l'auteur du discours : 26%, l'auteur du discours & le chercheur : 5%, le chercheur : 46%, l'argumentateur : 2%, l'auteur du discours & l'argumentateur : 2%, l'évaluateur : 4% et le reste : 3%.

³⁶ Cet *éthos scriptural*³⁶ renvoie à la représentation de soi reconstituée à partir du discours produit par l'auteur, soumis aux conditions socio-discursives visant des finalités bien précises requises par le discours produit, c'est-à-dire, aux contraintes du genre discursif (Amossy, 2015, p. 113).

D'autre part, l'auteur, en tant que protagoniste qui rédige l'écrit scientifique, doit gérer la rédaction pour communiquer sa recherche. Du point de vue communicationnel, le rôle de l'énonciateur du discours est essentiel pour que le lecteur puisse suivre le texte. Bien sûr, Fløttum *et al.* (2006) signalent la co-présence des rôles de l'énonciateur dans un seul énoncé, l'auteur peut être à la fois l'énonciateur du discours et le chercheur, ou l'énonciateur du discours et l'argumentateur, comme dans l'exemple *in this section, I argue that...* l'auteur joue le rôle d'énonciateur du discours et celui d'argumentateur (*ibid.*, p. 90).

Dans notre étude, nous supposons que l'auteur joue principalement le rôle d'énonciateur. Cela est lié aux fonctions de guidage qui visent à présenter d'une façon plus explicite le discours au lecteur. En parallèle, nous trouvons que l'approche consistant à identifier les rôles de l'auteur par les verbes associés peut être discutable. Les verbes de recherche citée plus haut (*analyser, examiner...*) et les verbes de positionnement (*proposer*) se trouvent régulièrement dans les routines de guidage, mais ils permettent de réaliser des fonctions structurant le plan du texte, comme nous le verrons dans les chapitres 5 et 8. Par ailleurs, d'autres indicateurs intéressants peuvent être mis en évidence pour identifier les rôles de l'auteur, comme les marqueurs métatextuels, discursifs et argumentatifs, des déictiques et les adjectifs évaluatifs.

2.3.3.2. *L'enrôlement du lecteur dans les articles de recherche*

En ce qui concerne le lecteur, bien qu'il ne soit pas présent lors de l'énonciation, il peut être enrôlé dans le discours par plusieurs moyens linguistiques. Par exemple, en utilisant les pronoms personnels *nous* ou *on* inclusifs dans les articles en français, l'auteur peut enrôler le lecteur dans le processus d'argumentation, comme co-témoin de la démonstration scientifique de l'auteur. Fløttum *et al.* (2006, p. 117) ont identifié 6 valeurs du pronom personnel *on* dans les articles de recherche en français :

Valeurs du pronom ON	du renvoie à	correspondant à
ON 1	auteur(s)	<i>je / nous</i>
ON 2	auteur(s) + lecteur(s)	<i>je / nous + vous</i>
ON 3	auteur(s) + communauté discursive limitée	<i>je / nous + vous (nos collègues)</i>
ON 4	auteur(s) + communauté discursive non limitée	<i>je / nous + vous (tout le monde)</i>
ON 5	lecteur(s)	<i>vous (le(s) lecteur(s))</i>
ON 6	autres	<i>il(s) / elle(s) / d'autre(s) chercheur(s)</i>

Tableau 2-2 : Les valeurs de ON dans les articles de recherche en français (Fløttum *et al.*, 2006, p. 117)

Parmi les 6 valeurs, le lecteur est présent discursivement dans les cas 2, 3, 4 et 5.

En effet, l'auteur peut jouer plusieurs rôles, mais c'est plutôt le rôle de l'énonciateur du discours qui permet d'assurer l'interaction entre l'auteur et le lecteur, et d'inviter aussi le lecteur dans le partage de la démonstration d'une question scientifique.

Du point de vue énonciatif, le rôle d'énonciateur garantit l'interaction entre l'auteur et le lecteur qui se manifeste au niveau discursif, c'est-à-dire qu'elle peut être projetée dans le fonctionnement interactionnel du discours scientifique, plus précisément, dans la perspective dialogique.

2.4. Guidage du lecteur dans les écrits scientifiques : procédés extralinguistiques et linguistiques

Bien que notre objet d'étude concerne les routines de guidage du lecteur, il faut préciser que le guidage ne s'effectue pas seulement à travers des moyens linguistiques (grammaticaux, lexicaux, discursifs). Nous pouvons recourir à de multiples dispositifs typographiques et aménagements schématiques (Péry-Woodley, 2000 ; Jacques, 2017). Guider la lecture consiste d'abord à construire le discours en répondant aux contraintes d'organisation textuelle du genre discursif cible, en fonction de l'interdiscours existant. De nombreuses mesures d'ordre matériel, par exemple mettant en relief la typographie du discours, sont utilisées afin de planifier visuellement le texte à l'aide des éléments non verbaux de mise en texte. Citons à titre d'exemple : l'ajout de figures ou graphiques explicatifs, de schémas procéduraux ou récapitulatifs, de tableaux, l'énumération par chiffres des informations (Péry-Woodley 2000), la gestion des espaces et des ponctuations, l'agencement de la structuration hiérarchique à l'aide d'un titrage structuré (Jacques, 2014, 2017) ou d'un découpage adapté (Fløttum, Dahl, & Kinn, 2006). Il faut noter que cette composition matérielle du texte est largement régie par le genre discursif : une recette de cuisine ne s'organise pas comme une pièce de théâtre, bien évidemment. De surcroît, dans les discours scientifiques, le texte écrit des sciences « dures » respecte souvent le schéma IMRAD pour construire son architecture (*cf.* section 2.2.2), tandis que le texte écrit des sciences humaines et sociales est moins contraint par cette composition et il recourt davantage aux indices péritextuels comme des titres de sections pour découper le texte et assurer le bon suivi du texte à travers des fonctions référentielles ou thématiques (Jacques, 2005).

Hyland (2005) fait les mêmes remarques sur l'étude du métadiscours – un concept discursif centré sur les moyens permettant de faciliter le suivi du discours et de créer la bonne relation intersubjective entre auteur et lecteur, que nous allons développer dans le troisième chapitre. Pour des discours écrits, l'univers des ponctuations et les éléments sémiotiques comme

le surlignage, l'italique, les retraits, etc. (Hyland, 2005, p. 28) exercent aussi un impact sur la lisibilité et l'interprétation du discours.

À côté de ces moyens non linguistiques, l'utilisation des marques linguistiques contribue aussi à guider la lecture. Il s'agit ici de se concentrer sur le niveau proprement discursif, et de mobiliser des éléments lexicaux ou énonciatifs pour organiser le contenu propositionnel du discours. Adam, en exploitant l'organisation du texte dans sa théorie sur *la linguistique textuelle*, inventorie les ressources linguistiques permettant de structurer le texte discursivement, ce qui concourt à faciliter la lecture : les anaphores et co-références, des connecteurs argumentatifs de causalité (*en effet, étant donné que*), de concession (*certes, bien que, malgré, quoique*), de conclusion (*donc, alors, par conséquent*), de réfutation (*mais, pourtant, néanmoins*), les organisateurs textuels d'énumération (*premièrement, ensuite*), d'exemplification (*par exemple, notamment*), temporels (*d'abord, puis*) ou spatiaux (*au-dessous, plus loin*) (Adam, 2005, p. 118–128).

2.5. Synthèse

Dans ce chapitre, nous avons discuté du fonctionnement interactif des écrits scientifiques, qui influence l'analyse des routines de guidage du lecteur, car ces dernières s'actualisent souvent dans le processus de l'énonciation. Nous nous sommes particulièrement intéressée à l'interaction entre l'auteur et le lecteur. Cette interaction se manifeste sur le niveau énonciatif par le concept de dialogisme, une théorie riche permettant d'interroger le fonctionnement interactif à travers l'énoncé et l'énonciation. L'emploi des routines de guidage du lecteur est mobilisé par le dialogisme dans les écrits scientifiques, en particulier le dialogisme interlocutif.

Le dialogisme propose un modèle adapté pour analyser les routines de guidage du lecteur dans l'approche énonciative, mais il ne propose pas de modèles lexicaux ou phrastiques concrets pour étudier les fonctions de guidage, d'autant plus que les études dialogiques portent souvent sur des traces linguistiques comme les pronoms personnels, les déictiques, des marqueurs relativement courts. Il nous est nécessaire de convoquer des ressources théoriques complémentaires, afin de mener une étude systématique sur ce point. C'est pour cette raison que nous recourons par la suite à l'approche fonctionnelle, telle qu'elle a été envisagée pour l'analyse du discours scientifique en anglais.

Chapitre 3. Le métadiscours pour l'étude des fonctions de guidage

Dans le chapitre précédent, nous avons discuté des caractéristiques énonciatives des écrits scientifiques, ainsi que de l'importance de mobiliser des éléments linguistiques qui orientent le lecteur comme les routines pour la bonne réception et interprétation du contenu scientifique. Dans cette partie, nous introduisons une approche qui nous permet de proposer une typologie fonctionnelle des routines de guidage du lecteur dans la suite de notre recherche.

Avant d'entrer dans les détails, nous proposons d'abord de parcourir quelques ressources exposant des classifications des éléments phraséologiques selon les fonctions. Cet état de l'art va montrer que les classifications sont très variées, mais qu'elles ne permettent pas de construire un modèle pour étudier systématiquement les fonctions de guidage. On verra que toutes les fonctions analysées dans ces modèles incluent des fonctions susceptibles de faciliter la lecture.

3.1. Classifications des expressions phraséologiques dans les écrits scientifiques

La classification des éléments phraséologiques selon leurs fonctions intéresse de nombreux linguistes. Notre travail n'est donc pas original dans le domaine de la phraséologie scientifique, mais nouveau pour les routines de guidage. Existe-t-il déjà des classifications incluant des fonctions permettant de guider la lecture ? Si oui, sont-elles utiles pour délimiter les fonctions de guidage ? Pour répondre à ces questions, nous allons d'abord rappeler certaines études ou ressources élaborées pour le discours scientifique. Ces études s'inscrivent souvent dans une perspective didactique et proposent des classements des expressions phraséologiques en mettant en lumière leurs fonctions.

3.1.1. Le Louvain *English Academic Purposes Dictionary*

Le *Louvain English Academic Purposes dictionary*³⁷ (*LEAD*) (Granger & Paquot, 2010) est un dictionnaire en ligne qui propose des ressources phraséologiques en anglais pour aider les apprenants dans leurs tâches rédactionnelles des textes scientifiques. Ces ressources recensent des collocations et des propositions récurrentes associées à différentes fonctions et proposent la consultation via des accès onomasiologiques et sémasiologiques. Les fonctions retenues sont rhétoriques, comme le montre l'accès onomasiologique du dictionnaire qui regroupe 18 fonctions rhétoriques dans la figure 3-1 :

³⁷ Le site : <https://leaddico.uclouvain.be/>.

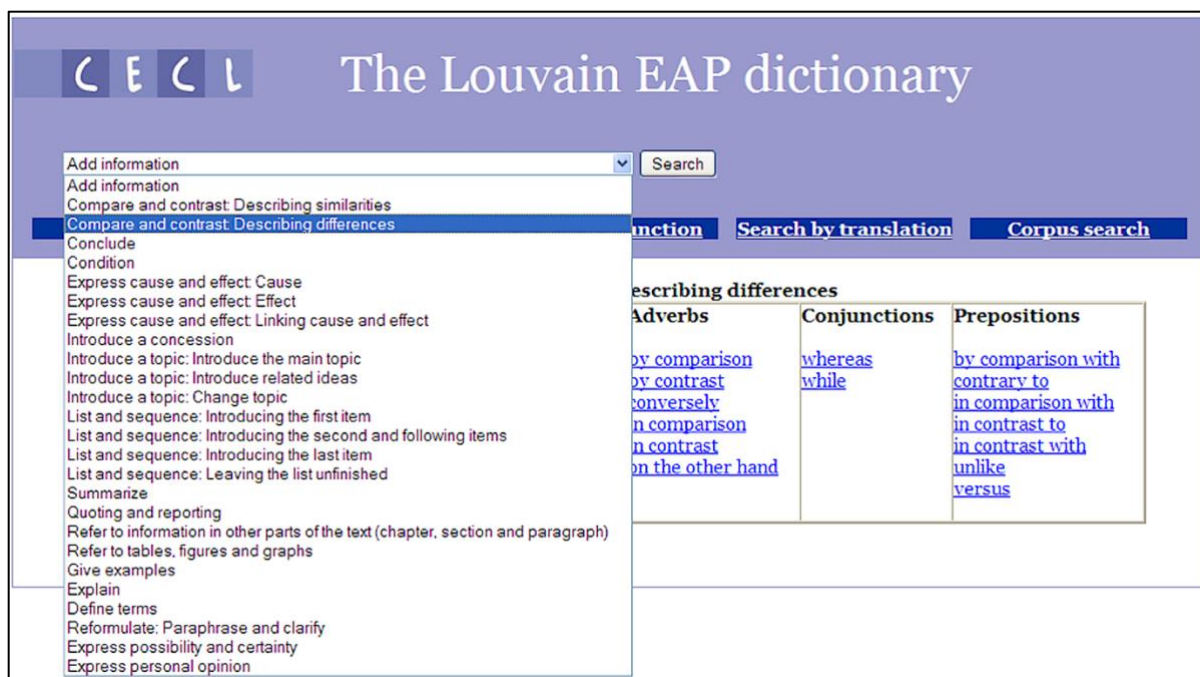


Figure 3-1 : L'interface du dictionnaire LEAP (Granger & Paquot, 2010, p. 323)

Certaines fonctions rhétoriques ont une visée d'accompagnement du lecteur, comme les fonctions permettant l'organisation du discours (<add information>, <conclude>, <introduce a topic>, <liste and sequence>, <summanize>, <refer to information in other parts od the text>, <refer to tableau, figure and graphs>). D'autres fonctions s'inscrivent plutôt dans l'argumentation et le positionnement, comme <expresse cause and effet>, <express personal opinion>. Dans ce dictionnaire, les expressions phraséologiques stockées comportent des phrases, mais la plupart des éléments recensés restent des collocations ou locutions.

3.1.2. Séquences discursives (Tran, 2014)

Tran (2014), qui a proposé une catégorisation pour les séquences discursives comme *en résumé*, *à condition que*, *dans l'intention de...*, a identifié deux grandes familles :

- celles qui exercent des fonctions métadiscursives comme la reformulation (*en d'autres termes, c'est-à-dire*), l'exemplification (*par exemple, tel(el) que*), l'addition (*ainsi que, de plus*), la structuration (*d'une part, d'autre part*), la topicalisation (*en termes de, en ce qui concerne*), l'énumération (*tout d'abord, par la suite*), la conclusion (*en conclusion, pour conclure*), la comparaison (*à l'instar de, en différence de*) ;
- celles qui remplissent des fonctions argumentatives à savoir la causalité (*en raison de, du fait que*), la conséquence (*par conséquent, il en résulte*), la concession (*en revanche, dans une certaine mesure*), l'opposition (*par contre, à l'opposé à*), la condition (*à condition que, en cas de*), la finalité (*de manière à, de sorte que*).

Cette typologie des séquences discursives partage des similitudes avec les fonctions rhétoriques du *Louvain English Academic Purposes Dictionary*. Plus précisément, la liste fonctionnelle de ce dernier recouvre une sphère plus large en introduisant des fonctions de positionnement (<express personal opinion>, <express possibility and certainty>. Cette différence est due aux objectifs visés par ces deux études. Pour le *LEAD*, l'objectif est de fournir un outil d'aide à la rédaction des écrits scientifiques. Dans cette optique, les fonctions principales dans les écrits scientifiques doivent être inventoriées afin de proposer une ressource phraséologique complète. Cependant, le travail de Tran (2014) ne se concentre que sur les séquences polylexicales à fonction discursive qui constituent l'objet d'étude de sa recherche.

3.1.3. Marqueurs discursifs polylexicaux (Siepmann, 2007)

Une autre étude intéressante est celle de Siepmann (2007) sur des marqueurs discursifs polylexicaux dans le discours scientifique en français. Ces marqueurs comportent au moins deux mots et renvoient souvent aux phrases préfabriquées d'une fréquence moyenne³⁸. Son étude classe les marqueurs avec les critères des fonctions discursives et interpersonnelles, proposant donc 22 fonctions, comme nous le voyons dans le tableau 3-1 :

Comparaison/contraste	<i>Même remarque sur/Il en va autrement de</i>
Concession	<i>en admettant que.../encore faut-il que (...)</i>
Exemplification	<i>Prenons un seul exemple. /Pour prendre un exemple adjectival, ...</i>
Explication	<i>On peut en donner l'explication suivante.</i>
Définition	<i>Le X est un Y si l'on a... ; Il est d'usage en/dans X de Vinf.</i>
Énumération	<i>Signalons (enfin)... /Pour commencer...</i>
Résumé/terminaison	<i>Une dernière remarque sur.../Nous terminons en indiquant...</i>
Inférence	<i>Il ressort de ce qui précède que.../Il découle de ces constatations que.../</i>
Causalité	<i>Deux raisons à cela/Pourquoi ? Pour deux raisons...</i>
Annonce	<i>Je me propose de passer en revue... /Nous verrons que...</i>
Introduction thématique	<i>Abordons maintenant... /Première remarque :...</i>
Exclusion	<i>Excluons toutefois le cas de (...)</i>
Digression	<i>Notons au passage que (...)/Soit dit en passant/</i>
Interrogation/réponse	<i>Le tout est de savoir si (...)</i>
Mise en relief	<i>Il faut être bien conscient que.../Précisons que...</i>
Information	<i>Pour information, retenons que (...)/Trois points à dégager...</i>
Clarification	<i>Soyons bien clair...</i>
Suggestion	<i>Il est sûr que.../Il va sans dire que.../L'expérience montre que.../</i>
Hypothèse	<i>On peut supposer que.../On peut penser que.../Supposons que...</i>
Restriction	<i>Un autre problème concerne...</i>
Renvoi et attribution	<i>[nom de personne] affirme que.../en nous inspirant de [nom de personne]</i>
Reformulation/récapitulation	<i>Pour le dire en d'autres termes... /j'ai cité précédemment...</i>

Tableau 3-1 : Fonctions des marqueurs discursifs polylexicaux chez Siepmann (2007)

³⁸ Les expressions phraséologiques retenues dans cette étude doivent être capables de fonctionner en tant qu'unités lexicales singulières, comme *même remarque sur, enfin et surtout...* La fréquence moyenne est définie par rapport aux marqueurs monolexicaux (*cependant, mais*). Elle s'applique aux expressions (=>2 mots). Par rapport aux marqueurs monolexicaux de fréquence haute comme *néanmoins* (1354 occurrences dans le Monde 1995), *pourtant* (7490 occurrences), les marqueurs de fréquence moyenne ont des fréquences relativement basses comme *il en va de même pour/de* (76 occurrences), ou *plus grave* (60 occurrences) (Siepmann, 2005, p. 51).

Ces fonctions sont élaborées en se basant sur des intuitions sémantico-pragmatiques de l'auteur et sont très proches des classements des marqueurs métadiscursifs que l'on verra dans la section 3.2.3. Cette étude est intéressante, car on trouve dans les exemples donnés, des phrases qui peuvent être considérées comme des routines. Ces phrases sont de différents types : des phrases complètes (*on peut en donner l'explication suivante*), des phrases incomplètes qui représentent la majorité des occurrences (*je me propose de passer en revue...*), des phrases injonctives (*excluons toutefois le cas de, notons au passage que*), des phrases impersonnelles (*il faut être bien conscient que, il est sûr que*), des phrases infinitives prépositionnelles (*pour prendre un exemple adjectival, pour le dire autrement*), des phrases nominales (*une remarque sur*), les propositions gératives (*en admettant que*), des phrases interrogatives (*Pourquoi ? Pour deux raisons*). La diversité des natures phrastiques montrée dans les exemples nous sera utile pour catégoriser les routines. En parallèle, la catégorisation met au jour au moins les traits discursifs fonctionnels du discours français.

3.1.4. *Academic Formulas List* (Simpson-Vlach & Ellis, 2010)

À partir d'un corpus multidisciplinaire d'écrits académiques et de présentations orales qui contiennent respectivement à peu près 2,1 millions de mots, l'*Academic Formulas List* de Simpson-Vlach & Ellis (2010) inventorie les expressions entre 3 et 5 mots apparaissant plus de 10 fois par million de mots. Par exemple, les expressions plus fréquentes comportent des locutions ou des phrases ouvertes comme *on the other hand, it has been, it is possible*. Basées sur l'approche des *lexical bundles* de Biber (pour plus de détails, voir section 1.1.3.2), les expressions recensées sont classées à l'aune de trois grandes fonctions, à savoir la référence, le positionnement, ainsi que l'organisation du discours, avec une liste de fonctions plus détaillée. Nous détaillons ici les fonctions d'organisation du discours qui sont récapitulées dans le tableau 3-2 (Simpson-Vlach & Ellis, 2010, p. 501–502) :

Organisation du discours	Exemples dans l'écrit	Exemples à l'oral
Métadiscours et références textuelles	<i>In the next section, in this paper, show in table</i>	<i>I was gonna say, come back to, what I'm saying</i>
Introduction du thème et focalisation	<i>For example, what are the,</i>	<i>I have a question, if you look (at), what I want to,</i>
Élaboration du thème	<i>In more details, such as those, as a result of the, for this reason</i>	<i>Any question about, i mean if, you know what I'm, as a consequence</i>
Marqueurs discursifs	<i>Even though the, in conjunction with</i>	<i>And if you, thank you very much, by the way</i>

Tableau 3-2 : Les sous-classes des fonctions d'organisation du discours

Cette classification regroupe, outre les expressions servant à introduire, à focaliser et à élaborer le thème du discours, des expressions métadiscursives textuelles et des marqueurs discursifs qui servent à organiser le discours. Encore une fois, l'organisation du discours est essentielle dans les fonctions des séquences polylexicales, comme on l'a vu dans les deux études précédemment discutées.

3.1.5. Academic Phrasebank (Davis & Morley, 2018)

Élaboré par l'Université de Manchester, l'*Academic Phrasebank* (Davis & Morley, 2018) est aussi un outil d'aide à la rédaction en ligne en anglais scientifique comme le *LEAD*. Il présente un ensemble d'expressions phrastiques souvent utilisées dans les articles de recherches. Inspiré des analyses du genre du discours scientifique de Swales (1990), il propose deux entrées, l'une par la liste des fonctions générales de l'écrit scientifique, l'autre par la liste des fonctions fréquentes retrouvées dans les parties textuelles (l'introduction, la conclusion, etc.) de l'écrit scientifique³⁹.

La première liste rassemble 11 fonctions qui sont indépendantes des parties textuelles. Nous trouvons par exemple, des expressions pour classer ou lister des informations (*X can be classified into Xi and Xii*) ; pour donner des exemples (*this is exemplified in the work undertaken by*) ; pour exprimer la prudence (*A possible explanation for this might be that...*), etc.

La deuxième liste classe des fonctions particulièrement récurrentes dans certaines parties textuelles. Ces fonctions peuvent se subdiviser en de nombreuses sous-fonctions. À titre d'exemple, pour introduire le travail (fonctions principales), nous pouvons utiliser des expressions servant à souligner l'importance du thème de recherche (*X is an important aspect of*) ; à montrer la structure du texte (*the first section of this paper will examine...*), etc. Dans la partie méthodologique, les expressions visant à décrire les méthodes (*criteria for selecting the subjects were as follows*) sont dominantes. Il en va de même pour les expressions qui présentent des résultats (*this result is somewhat counter-intuitive*) ou qui discutent des découvertes (*what is surprising is that*) dans la partie analyse ou de discussion. Nous rassemblons les deux listes ainsi que leurs fonctions principales dans le tableau 3-3 :

³⁹ Se référer au site de l'Academic Phrasebank : <https://www.phrasebank.manchester.ac.uk/>.

Functions of Academic phrases in English	
General language functions	Functions in different sections
(1) Being cautious	(1) Introducing work
(2) Being critical	(2) Referring to sources
(3) Classifying and listing	(3) Describing methods
(4) Compare and contrast	(4) Reporting results
(5) Defining terms	(5) Discussing findings
(6) Describing trends	(6) Writing conclusions
(7) Describing quantities	
(8) Explaining causality	
(9) Giving examples	
(10) Signalling transition	
(11) Writing about the past	

Tableau 3-3 : Les fonctions principales intégrées par l'Academic Phrasebank

Ces fonctions, qualifiées de « communicatives », assurent le bon déroulement de la communication du travail de recherche. Elles visent à fournir des ressources phraséologiques concernant la réflexion sur le contenu (l'argumentation) et sur l'organisation de l'écrit.

Ces descriptions sont intégrées dans des outils linguistiques pour faciliter la rédaction des articles académiques en anglais, ce qui invite à réfléchir également à la constitution des ressources phraséologiques de langue française dans une même finalité.

3.1.6. Classement fonctionnel des routines sémantico-rhétoriques (Chevalier, 2019)

Nous voudrions enfin aborder une autre ressource fonctionnelle sur des routines sémantico-rhétoriques proposée par Chevalier (2019)⁴⁰. Adoptant la même définition que nous pour les routines sémantico-rhétoriques (*cf.* Chapitre 1), elle identifie 12 fonctions dans les écrits scientifiques réparties dans les sections textuelles que nous récapitulons ici dans le tableau 3-4 :

⁴⁰ Il s'agit aussi une étude intéressant pour modéliser les routines, nous y revenons dans la partie méthodologique.

Sections textuelles	Fonctions rhétoriques	Exemples des routines
Introduction	1. Annoncer une problématique ou un sujet	<i>cet article porte sur/traite de</i>
	2. Préciser les objectifs	<i>cet article a pour objectif/cherche à/visé à ; l'objectif de cette recherche est de...</i>
	3. Formuler les questions	<i>cet article pose les questions suivantes/ce sont les questions aux- quelles tente de répondre le présent article</i>
Cadre conceptuel	4. Dresser l'état des connaissances issu de la recherche	<i>X (date) soutient que,/pour X, selon X, comme X le souligne, P (Auteur, date)</i>
	5. Déclarer ses affiliations	<i>Nous avons retenu les éléments du modèle de X (Auteur, date)/nous nous appuyons sur...</i>
Méthodologie	6. Présenter sa démarche méthodologique	<i>pour atteindre cet objectif... cette question nous amène à...</i>
Résultat	7. Présenter les données	<i>comme indiqué dans le tableau 5, la figure 3 montre...</i>
	8. Établir les constats	<i>nous avons trouvé que... les données montrent que... il appert que...</i>
Discussion	9. Comparer les résultats à d'autres recherches	<i>nos résultats concordent avec, contrairement aux études précédentes...</i>
Discussion ou conclusion	10. Préciser les limites de la recherche	<i>Cette étude comporte certaines limites, il convient de souligner les limites de cette recherche.</i>
	11. Proposer des pistes de recherche futures	<i>la recherche fait émerger des pistes de recherche, il pourrait être intéressant de prolonger les investigations en ce sens</i>
	12. Formuler des recommandations	<i>certaines recommandations peuvent être faites, [compte tenu de ces résultats/constats], nous recommandons que... il est essentiel de faire x [pour atteindre tels résultats]</i>

Tableau 3-4 : 12 Fonctions identifiées pour les routines sémantico-rhétoriques dans les écrits scientifiques

L'élaboration des fonctions recensées dans cette étude est de proche de l'*Academic Phrasebank* : l'idée est toujours de trouver des fonctions liées aux parties textuelles des écrits scientifiques. Chevalier (2019) souligne les fonctions les plus saillantes dans les démarches rédactionnelles qui se retrouvent dans toutes les langues. On observe donc par exemple <annoncer une problématique ou sujet> peut rapprocher à la fonction <introducing work> classée dans l'*Academic Phrasebank*.

Il ressort de cette revue de la littérature que les fonctions observées et retenues pour des expressions récurrentes dans le discours scientifique sont riches et touchent plusieurs aspects inhérents du discours scientifique, rhétoriques, argumentatifs et discursifs. Ces ressources sont intéressantes, mais ne correspondent pas pleinement à notre objectif de recherche. Par conséquent, il nous faut trouver un modèle fonctionnel qui permet de prendre en compte de façon plus systématique les fonctions de guidage. Nous proposons ainsi de nous intéresser à la notion de métadiscours et aux classements associés dans ce cadre.

3.2. Métadiscours

La notion de métadiscours est étroitement liée à la nature dialogique du discours à travers l'anticipation des attentes du destinataire, dans l'intention de mettre en évidence la prise en compte de ce dernier. Plus concrètement, par le biais de l'emploi des ressources linguistiques spécifiques, le métadiscours permet de mieux orienter l'interprétation du discours. Hyland a résumé ce lien entre le métadiscours et le dialogisme du discours :

Le métadiscours est donc un lien important entre un texte et son contexte, car il indique les attentes des lecteurs concernant certaines formes d'interaction et d'engagement. Il met en évidence le rôle dialogique du discours en révélant la compréhension qu'a un auteur d'un public donné à travers la manière dont il s'adresse à ses lecteurs et à leurs besoins⁴¹. (Hyland, 2005, p. 13)

Du point de vue théorique, le concept de métadiscours met en évidence l'importance de la mobilisation des attentes du récepteur du discours, un point essentiel qui fait écho au dialogisme développé dans le deuxième chapitre. En parallèle, ce concept permet, du point de vue pratique, des catégorisations assez fines et opérationnelles pour des marqueurs métadiscursifs. Ce sont ces deux avantages qui motivent notre décision de recourir à la notion de métadiscours.

La notion de métadiscours est polysémique. Les linguistes ancrés dans la perspective énonciative, par exemple J. Rey-Debove (1997) ou Authier-Revuz (1995), emploient le terme de « métadiscours » pour les notions de métalangage et de métaénonciation : la notion est définie soit en tant que réalisation de la métalangue naturelle dans le discours, c'est-à-dire, la mise en œuvre métalinguistique du langage (Rey-Debove, 1997, p. 20), soit comme moyen linguistique pour commenter l'énonciation (Authier-Revuz, 1995, p. 18), renvoyant à la réflexivité du langage (nous y revenons dans le Chapitre 6). Cependant, ces études n'ont pas proposé de catégorisations pour les marqueurs métadiscursifs. C'est pourquoi nous nous inspirons ici plus directement des modèles linguistiques du métadiscours développés et utilisés largement dans les travaux en anglais, notamment dans le cadre de l'*English Academic Writing*.

La notion de métadiscours est beaucoup étudiée dans le monde anglophone dans le domaine de l'*English for Academic Purposes* (EAP), notamment pour l'écrit des textes académiques. Maintes recherches sur le métadiscours ont déjà montré que l'utilisation du métadiscours peut faciliter la communication, consolider les arguments, élever la lisibilité et

⁴¹ Notre traduction, la citation originale est : Metadiscourse is therefore an important link between a text and its context as it points to the expectations readers have for certain forms of interactions and engagement. It highlights the dialogic role of discourse by revealing a writer's understanding of an audience through the ways that he or she addresses readers and their needs (Hyland, 2005, p. 13).

construire une relation intersubjective entre l'auteur et le lecteur (Hyland, 2005, p. 5). Les marqueurs métadiscursifs exercent ainsi une grande influence sur la perception du texte chez des destinataires d'un discours (*ibid.*, p. 11). Il nous paraît pertinent de mobiliser cette notion, car elle met particulièrement en valeur la prise en compte du lecteur, c'est-à-dire qu'elle est orientée vers l'anticipation du public cible, pour laquelle le guidage du lecteur est fondamental. Selon l'approche dialogique (*cf.* Chapitre 2), le discours est le dialogue entre l'auteur et le lecteur. L'ajout du métadiscours ne modifie pas de façon drastique le fond du dialogue, mais le rend plus facile à suivre. Dans l'exemple 1 : *Cela veut dire que* constitue un marqueur métadiscursif qui ne change pas radicalement le sens de la phrase, mais aide à reconsidérer l'énoncé précédent *il voit... la réalité du langage intérieur* en introduisant une autre orientation argumentative. Le lecteur aura ainsi une compréhension plus claire des idées de l'auteur.

[1] La démonstration de Vygotskij est convaincante ; *il voit dans le fait que le sujet utilise les significations du mot d'une manière fonctionnelle le fondement et la réalité du langage intérieur. Cela veut dire qu'on ne peut pas localiser la pensée dans le foyer intérieur du sujet, car c'est en utilisant le langage que le sujet donne naissance à la pensée.* (Linguistique)

Dans les écrits scientifiques, l'emploi des marqueurs métadiscursifs permet à l'auteur de s'exprimer avec efficacité et d'établir un lien avec le lecteur. Le métadiscours est un concept qui permet non seulement de comprendre comment l'auteur guide le lecteur et interagit avec lui, mais aussi de fournir en particulier des ressources fécondes pour le classement des marqueurs lexicaux ou phrastiques qui pourraient être qualifiés de métadiscursifs. Un sous-ensemble de ces marqueurs inclut clairement des marqueurs visant à guider la lecture.

Dans cette section, nous aborderons en premier lieu la genèse de la notion de métadiscours et enchaînerons par quelques définitions. Ensuite, nous entrerons dans les détails de certains des modèles de métadiscours les plus utilisés, afin de mettre en lumière les différentes façons de classer les ressources linguistiques en tant que marqueurs métadiscursifs, notamment les marqueurs de guidage. Une proposition mettant en évidence la pertinence du métadiscours pour notre étude sera présentée à la fin de la section.

3.2.1. Genèse de la notion de métadiscours

La notion de métadiscours est essentiellement utilisée pour désigner dans un texte des éléments linguistiques permettant de comprendre l'utilisation du langage et les moyens que le locuteur ou de l'auteur emploie pour guider l'interprétation du destinataire (Hyland, 2005, p. 3).

Initialement forgé par Harris en 1959 dans son article pour discuter de la structure transformationnelle de la langue anglaise (Harris, 1959), le terme de *métadiscours* est utilisé

pour désigner des mots qui n'ont pas de sens concret, mais qui correspondent à des remarques sur le contenu principal du discours (*we have found that...*). Ces éléments linguistiques n'ont pas pour objectif de communiquer des informations essentielles dans le texte, mais servent à assurer la clarté de la communication de ces informations. Par la suite, d'autres linguistes comme Meyer (1975) et Lautamatti (1978) soulignent également l'existence d'éléments linguistiques de second niveau qui aident à la structuration ou à l'organisation du texte. Lautamatti (1978), dans une perspective d'observation de la progression du topique, met en évidence des éléments non thématiques à côté d'éléments thématiques. L'étude du métadiscours reste cependant marginale à cette période et se présente comme de petits fragments disséminés dans les études linguistiques portant sur des objets d'étude divers distincts de celle-ci. C'est à partir des années 80 que le métadiscours est devenu une notion centrale pour l'analyse de discours de type argumentatif, en particulier le discours scientifique (par exemple Crismore, 1983, 1984, 1989 ; Intaraprawat & Steffensen, 1996). Il a été théorisé de façon plus aboutie ensuite par des linguistes fonctionnels.

Williams (1981, p. 211–212) est le premier à utiliser le terme *métadiscours* dans son livre *Style : ten lessons in clarity and grace*. Il l'a identifié, dans une perspective de didactique de l'écrit, comme un ensemble de phénomènes qui aident à construire le discours, et le définit comme « l'écrit sur l'écrit, tout ce qui ne renvoie pas au contenu traité »⁴² (*ibid.*, p. 209–210). Le métadiscours est considéré par ce linguiste comme un style d'écriture, dont l'utilisation n'a pas pour objectif de se référer au contenu des idées de l'auteur, mais de se référer à l'auteur lui-même, au lecteur, et à l'énonciation de l'auteur (*ibid.*, p. 65, 9e édition). Il précise que l'utilisation ou non d'éléments métadiscursifs peut produire un effet important sur la perception de la relation entre l'auteur et le lecteur. Son emploi à bon escient favorise la clarté du texte, tandis qu'une trop grande redondance peut au contraire nuire à sa clarté. Dans le modèle de Williams (1981, p. 121–123), le métadiscours réalise deux fonctions principales dans le discours écrit : celle de guider la lecture avec des marqueurs de transition (*and, but, therefore*), de structuration (*first, second*), et celle d'exprimer le degré de certitude comme les marqueurs d'atténuation ou *hedges* en anglais avec des mots comme *usually, often, a certain number of, may, appear, suggest, very, obviously, key, central*, etc. Le développement du métadiscours chez Williams inspire par la suite des études plus systématiques sur le métadiscours, en particulier des marqueurs spécifiques au métadiscours. Bien entendu, l'approfondissement de

⁴² Traduction personnelle de la définition en anglais : “writing about the writing, whatever does not refer to the subject matter being addressed” (Williams, 1981, p. 209-210).

cette notion engendre également des débats sur sa définition, son identification dans les corpus, et le classement des fonctions dites métadiscursives.

3.2.2. Définition du métadiscours

Les routines de guidage se classent sous le chapeau des marqueurs métadiscursifs. Il est donc indispensable de comprendre d'abord la définition du métadiscours. Néanmoins, il existe plusieurs définitions qui tantôt se ressemblent, tantôt s'éloignent les unes des autres. En fonction de leur point de départ théorique et de leur façon de classer les marqueurs métadiscursifs, on peut les diviser en deux grands courants : le métadiscours du modèle interpersonnel ou interactif, qu'on nomme parfois modèle « léger » et le métadiscours du modèle réflexif, qualifié de modèle « lourd »⁴³ (Ädel & Mauranen, 2010). Ces deux courants ont développé leurs définitions selon des approches différentes. L'approche de la linguistique systématique fonctionnelle (Halliday & Matthiessen, 2014) (SFG⁴⁴) guide la réflexion des modèles interactifs/interpersonnels, en particulier leurs classements fonctionnels des marqueurs métadiscursifs. Les modèles interactifs/interpersonnels élaborent un univers théorique en séparant le niveau métadiscursif du niveau propositionnel du langage, mettant en évidence les aspects textuels et interpersonnels du discours. Les modèles réflexifs s'inspirent davantage du schéma des fonctions du langage de Jakobson, questionnant l'aspect réflexif du discours.

3.2.2.1. Métadiscours des modèles interactifs et interpersonnels

Le métadiscours des modèles interpersonnels et interactifs représente un modèle très répandu dans le monde anglophone pour étudier le discours de type argumentatif. Nous pouvons citer des chercheurs dans l'approche de la linguistique systématique fonctionnelle de Halliday & Matthiessen, (2014), comme Vande Kopple (1985), Crismore (1983, 1989), Crismore *et al.* (1993) et Hyland (2005). Le fait que l'on attribue l'adjectif *interactif* ou *interpersonnel* à la dénomination de leurs modèles s'explique par l'importance de la relation interpersonnelle entre les participants dans le métadiscours. Les modèles conçus dans cette perspective sont construits sur une prémisse majeure : la distinction entre le niveau propositionnel et le niveau métadiscursif du discours. Rappelons que les études de la première époque, succinctement décrites dans la première partie de cette section, ont signalé la différence entre les informations propositionnelles et le second niveau d'ordre d'organisation du texte et d'éléments assurant la clarté du texte, sans plus de développement. La division entre le niveau propositionnel et le

⁴³ Nous avons repris ici les dénominations de Ädel & Mauranen (2010) et traduit 'thin' par « léger », 'thick' par « lourd ».

⁴⁴ SFG = Systemic Functional Grammar, il est traduit en français comme « linguistique systématique fonctionnelle ».

niveau métadiscursif du discours constitue l'idée centrale des études sur le métadiscours Lautamatti (1978) et chez Williams (1981). Elle influence ensuite la conceptualisation de la notion de métadiscours chez des chercheurs comme Vandekopple, Crismore ou Hyland, cités plus haut.

Vandekopple (1985) et Crismore (1983, 1989) développent leurs définitions du métadiscours en se basant sur la division nette entre le niveau propositionnel et le niveau métadiscursif. Dans la définition du premier chercheur, le métadiscours est envisagé comme un niveau indispensable à côté du niveau propositionnel, qui « n'ajoute pas de contenu propositionnel du texte, mais qui a pour fonction d'aider le lecteur ou l'auditeur à organiser, clarifier, à interpréter, à évaluer les informations déjà transmises et à y réagir » (Vandekopple 1985, p. 83). Quant à Crismore, le chercheur qui a publié le plus d'articles sur le métadiscours dans les années 80 (Aguilar, 2008, p. 66), il a également développé une définition relativement claire soulignant cette division :

Matériel linguistique présent dans les textes, écrits ou oraux, qui ne contribue en rien au contenu propositionnel, mais qui est destiné à aider l'auditeur ou le lecteur à organiser, interpréter et évaluer l'information donnée⁴⁵. (Crismore *et al.*, 1993, p. 40)

Dans un article plus récent, Vandekopple (1997) a réaffirmé le niveau métadiscursif en tant que niveau complémentaire du niveau propositionnel, et a montré que c'est à ce niveau que s'opèrent les influences du public destinataire en matière d'organisation, d'interprétation, d'évaluation et d'attitude vis-à-vis du niveau propositionnel.

Outre la distinction nette entre le niveau propositionnel et le niveau métadiscursif du discours, ces deux linguistes mettent l'accent, dans leurs définitions, sur la présence de l'auteur et l'anticipation des attentes du lecteur. Par exemple, la définition de Vandekopple montre que le métadiscours peut être considéré comme l'ensemble des ressources linguistiques qui font émerger la présence de l'auteur dans le discours (Hyland, 2005, p. 18). Plus précisément, son modèle décrit des stratégies rhétoriques que l'auteur doit mobiliser pour construire un discours lisible et convaincant : des stratégies pour aider le lecteur à comprendre l'organisation textuelle ; d'autres pour signaler la présence de l'auteur à travers les expressions des sentiments ou des interactions avec le destinataire. Dans un travail plus ancien sur les marqueurs métadiscursifs dans les textes des sciences sociales et non scolaires, Crismore (1983) donne une définition qui

⁴⁵ Notre traduction. La citation originale est : Linguistic material in texts, written or spoken, which does not add anything to the propositional content but that is intended to help the listener or reader organize, interpret and evaluate the information given (1993, p 40).

met en lumière l'importance de la présence de l'auteur et de guider la lecture en anticipant les attentes du lecteur :

Discours d'un auteur sur un discours donné ; il s'agit de l'intrusion de l'auteur dans le discours, de manière explicite ou non, pour orienter le lecteur plutôt que pour l'informer. Le métadiscours est constitué des directives données aux lecteurs afin qu'ils comprennent ce qui est dit et signifié dans le discours primaire et qu'ils sachent comment 'interpréter' l'auteur⁴⁶. (Crismore, 1983)

Nous pouvons voir que Vande Kopple et Crismore interprètent de manière semblable la définition du métadiscours basée sur la séparation des niveaux de sens inspirés tous deux par la linguistique systémique fonctionnelle.

Ce point de vue n'est pas partagé par Hyland (2005), qui considère qu'il est erroné de séparer les niveaux. Il considère que le sens du texte se réalise par l'action simultanée des niveaux au lieu d'agir indépendamment. Par conséquent, il ne faut pas traiter le niveau métadiscursif comme complémentaire du niveau propositionnel, mais les voir comme consubstantiels et agissant de concert (*ibid.*, p. 23). De plus, il met l'accent sur la nature interpersonnelle du métadiscours et propose la définition suivante :

Le métadiscours est le terme générique renvoyant aux expressions autoréflexives pour négocier le sens d'un texte, aider l'auteur (ou le locuteur) à exprimer un point de vue et à s'engager avec les lecteurs en tant que membres d'une communauté donnée.⁴⁷ (Hyland, 2005, p. 37).

Dans sa définition, Hyland souligne trois principes du métadiscours et donne des détails que nous pouvons synthétiser ici (Hyland, 2005, p. 38–48) : 1) la distinction entre l'aspect propositionnel et l'aspect métadiscursif ; 2) l'interaction entre l'auteur et le lecteur ; 3) la distinction entre références externes et internes du discours.

- 1) Le métadiscours est distinct de l'aspect propositionnel du discours. Au lieu de distinguer « les niveaux de sens », Hyland préfère l'expression de Sinclair « plans du discours » pour expliquer cette distinction entre l'aspect propositionnel et l'aspect métadiscursif. Cette théorie est un modèle destiné aux récepteurs du texte. Elle explique comment la langue nous permet d'organiser des matériels informatifs dans le discours et de négocier

⁴⁶ Notre traduction. La citation originale est : An author's discoursing about the discourse; it is the author's intrusion into the discourse, either explicitly or non-explicitly, to direct the reader rather than inform. Metadiscourse is the directives given to readers so they will understand what is said and meant in the primary discourse and know how to "take" the author (Crismore, 1983).

⁴⁷ Notre traduction. La citation originale est : Metadiscourse is the cover term for the self-reflective expressions used to negotiate meaning in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community (Hyland, 2005, p. 37).

la relation avec le destinataire afin de garantir la compréhension du texte par ce dernier. Le discours envisagé dans ce modèle contient donc deux dimensions, une dimension interpersonnelle liée aux aspects informatifs, expressifs, sentimentaux et interactionnels qui concerne la nécessité de négocier les contenus avec le destinataire, et une autre dimension textuelle qui permet de transformer les informations du monde réel en discours. Ces deux dimensions se situent entre le monde réel et le langage, nommés également « le plan autonome » et « le plan interactif ». C'est sur le premier que le langage est utilisé pour organiser des expériences dans le monde réel, et que l'auteur joue le rôle d'informateur, selon la distinction des rôles de l'auteur et du lecteur dans le modèle de Hunston (2000, p. 183, cité par Hyland, 2005, p. 40). Entre le plan autonome et le monde réel se situe le plan interactif, dans lequel se réalise l'aspect métadiscursif. L'auteur y joue le rôle de constructeur du texte. En résumé, l'aspect propositionnel reflète les choses dans le monde réel et les informations principales mises en discours. Le métadiscours reflète les choses dans le discours : comment l'auteur organise et exprime des informations pour ses lecteurs intentionnels. C'est avec l'utilisation du métadiscours que l'auteur peut produire des contenus propositionnels de façon plus cohérente, claire et persuasive pour son public cible. Ces deux aspects sont inséparables.

- 2) Le métadiscours renvoie aux aspects du texte qui incarnent l'interaction entre l'auteur et le lecteur. Son modèle met en évidence la nature interpersonnelle du métadiscours. Il insiste sur le fait que tout élément métadiscursif doit être interpersonnel, car il prend en compte les connaissances du lecteur, ses attentes en matière d'organisation du texte et de traitement des informations. Avec toutes ces anticipations, l'auteur met en place des stratégies rhétoriques qui lui permettent de réaliser l'organisation textuelle et le traitement des informations. Cette remarque a un grand impact sur la classification des marqueurs métadiscursifs dans le discours chez Hyland et lui fait prendre la décision de ne pas isoler l'aspect textuel de l'aspect interpersonnel, étant donné que l'aspect textuel constitue l'une des caractéristiques interpersonnelles du métadiscours.
- 3) Hyland attire l'attention sur la distinction entre les relations externe et interne du discours. La relation externe renvoie à celle qui existe entre le monde réel où nous vivons et le discours. La relation interne relie les discours (ce qui rejoint l'idée du dialogisme). Le métadiscours doit se référer à la relation interne du discours, non à la relation externe vers le monde réel. Par exemple, des connecteurs de séquence peuvent décrire l'ordre des expériences dans le monde réel et mettre en ordre des informations du discours. Ils sont qualifiés de métadiscursifs pour la deuxième fonction

d'organisation discursive⁴⁸. La relation interne du discours reflète également la réflexivité du discours qui montre que le métadiscours fonctionne en tant qu'outil pour mener des réflexions sur le discours lui-même.

Ces trois principes ont un impact sur le classement fonctionnel du métadiscours chez Hyland. Bien que les définitions des modèles interactifs/interpersonnels du métadiscours comportent des différences, elles mettent en évidence toute l'importance de l'interaction entre l'auteur et le lecteur. L'idée centrale est que le discours est vu comme un acte social, et que le métadiscours est un moyen linguistique et rhétorique pour assurer une bonne interaction entre l'auteur et le lecteur à travers cet acte social.

3.2.2.2. *Métadiscours du modèle réflexif*

Un autre courant du métadiscours est celui de chercheurs comme Ädel (2006), Ädel & Mauranen (2010). Dans son travail de comparaison des marqueurs métadiscursifs dans les textes écrits de type argumentatif écrits par des apprenants avancés de l'anglais d'origine suédoise et des natifs anglophones anglais et américains, Ädel (2006) propose un nouveau modèle réflexif du métadiscours ainsi qu'une définition plus abstraite. Elle conceptualise le modèle réflexif dans un contexte de recherches menées sur la réflexivité du langage et s'inspire de certaines fonctions du langage issues du schéma de Jakobson. D'une part, la réflexivité est une propriété inhérente au langage –l'être humain peut utiliser la langue non seulement pour décrire le monde extérieur, mais aussi pour commenter son propre discours. Le métadiscours constitue une partie des éléments linguistiques qui permettent de réaliser cette réflexivité sur le discours lui-même. Dans cette optique, l'auteur est amené à prendre conscience du texte actuel, du langage utilisé, de lui-même et du destinataire réel ou intentionnel du texte actuel (Ädel, 2006). D'autre part, la linguiste juge qu'il n'est pas très pertinent de faire la distinction entre le niveau propositionnel et le niveau métadiscursif, comme on le fait dans les modèles interactif/interpersonnel, et que le recours au schéma des fonctions langagières de Jakobson permet d'éviter cette distinction.

L'idée principale d'Ädel (2006) est d'analyser le phénomène du métadiscours à partir des trois fonctions –métalinguistique, expressive et conative –inspirées du schéma fonctionnel du langage de Jakobson (1963). Ces trois fonctions se concentrent respectivement sur le

⁴⁸ En fait, la distinction entre le niveau du monde réel et celui du discours est difficile dans les articles de recherches en SHS, car elle se manifeste souvent implicitement.

texte/code, l’auteur et le lecteur du texte actuel. Pour l’expliquer de façon plus claire, elle propose un triangle réflexif pour illustrer son modèle.

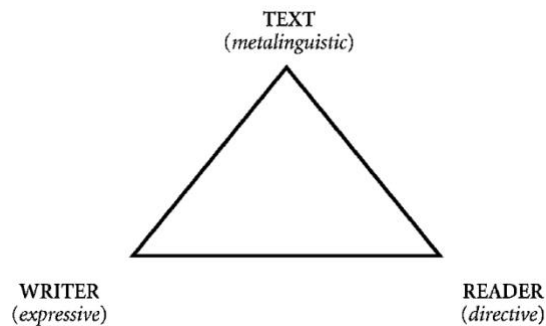


Figure 3-2 : Triangle réflexif du métadiscours (Ädel, 2006, p. 18)

Dans ce triangle réflexif, le texte/code fait appel au matériel linguistique qui se concentre sur le texte, sur l’écriture ou le discours qui s’y déroule. En d’autres termes, le métadiscours peut remplir des fonctions métalinguistiques dans le discours. Ce matériel linguistique peut concerner plusieurs plans du texte, allant des mots utilisés aux parties textuelles, voire au texte entier (Ädel, 2006, p.18–19). La fonction expressive concerne l’auteur et renvoie au matériel linguistique qui implique l’auteur du texte actuel. La fonction conative implique le lecteur, représenté ici par le matériel linguistique référant au lecteur cible souvent imaginaire (*ibid.*, p. 19). Par rapport aux modèles interactifs/interpersonnels qui mettent en évidence la propriété de l’interpersonnalité du métadiscours, le modèle réflexif ne néglige pas cette propriété, mais met en valeur surtout la propriété de réflexivité du métadiscours, manifestée notamment par sa fonction métalinguistique.

En tenant compte des trois constituants du triangle du modèle réflexif, le métadiscours est défini de la façon suivante :

Le métadiscours est un texte sur le texte en cours, ou le commentaire explicite de l’auteur sur son propre discours en cours. Il témoigne d’une conscience du texte actuel ou de son utilisation de la langue en tant que telle, ainsi que de l’auteur et du lecteur actuels en tant qu’auteur et lecteur⁴⁹ (Ädel, 2006, p. 20).

Du point de vue fonctionnel, Ädel (2006) distingue généralement deux grands types de métadiscours : le métadiscours métatextuel et le métadiscours qui renvoie à l’interaction entre l’auteur et le lecteur. Le métadiscours métatextuel regroupe tous les éléments concernant la structuration du texte et l’usage de la langue du discours. Sa fonction réside davantage dans le

⁴⁹ Notre traduction, la citation originale est : Metadiscourse is a text about the evolving text, or the writer’s explicit commentary on her own ongoing discourse. It displays an awareness of the current text or its language use *per se* and of the current writer and reader *qua* writer and reader. (Ädel, 2006, p. 20).

guidage du lecteur tout au long du discours. Le deuxième type de métadiscours concerne l'interaction que l'auteur voudrait maintenir avec son lecteur cible. Partant d'une image préétablie du lecteur intentionnel, la mobilisation de ce type de métadiscours peut permettre à l'auteur d'atteindre deux objectifs principaux : établir une relation avec son lecteur et la maintenir d'une part, exprimer sa propre attitude et influencer son lecteur d'autre part.

Au-delà de ces deux types, ce modèle du métadiscours peut être observé à travers les modes impersonnels et personnels. De façon plus précise, il s'agit ici, d'une classification basée sur des critères purement grammaticaux. Le mode impersonnel désigne le métadiscours qui n'implique pas la référence directe à l'auteur et au lecteur du discours présent, manifesté immédiatement par la présence des pronoms personnels déictiques *I, me, my, we, our, you, your*, etc. C'est souvent au discours ou au texte en cours qu'il se réfère. Nous trouvons souvent ce mode mobilisant des éléments métatextuels, par exemple : *as shown above, in this essay, ... will be discussed in the following, see page 16* (p. 20), etc. Le métadiscours faisant référence directe à l'auteur ou au lecteur du discours présent en utilisant des pronoms personnels, ou des substantifs métalinguistiques comme *writer, reader, audience*, etc. appartient au mode personnel. Ce mode se trouve davantage dans le deuxième type de métadiscours – interaction auteur-lecteur, comme *you may be thinking, I would like to discuss the topic which, as we have seen*.

En partant du triangle réflexif, Ädel (2006, p. 21) retrace les intersections entre les pôles <texte>, <auteur>, <lecteur> et les définit comme les différentes configurations du métadiscours dans la figure 3-3 :

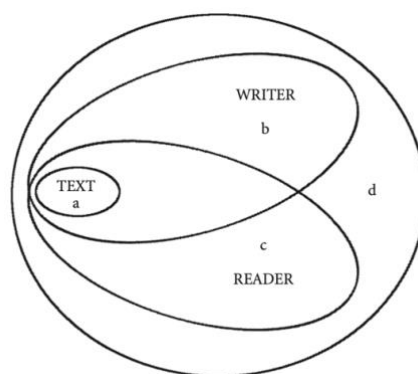


Figure 3-3 : Configuration du métadiscours chez Ädel (2006, p. 21)

La dimension 'Text' désigne le métadiscours axé sur le texte, autrement dit le métadiscours métatextuel. Les autres dimensions renvoient à l'interaction entre l'auteur et le lecteur, qui est représentée par le métadiscours interpersonnel. En ce qui concerne les différents

usages des pronoms personnels, la dimension 'Writer' se manifeste par l'emploi des pronoms qui renvoient à l'auteur du discours. La dimension 'Reader' se manifeste par l'emploi des pronoms explicitement dédiés au lecteur (*you, your, etc.*), et la dimension 'd' constitue un univers au-delà des trois dimensions, orientées vers les participants, c'est-à-dire l'auteur et le lecteur en même temps (*we, our*) (Ädel, 2006, p. 21).

Le métadiscours développé par Ädel met en évidence la réflexivité de l'auteur non seulement sur son propre texte en cours, mais aussi sur le lecteur intentionnel. Tout élément métadiscursif doit permettre de renvoyer constamment à la situation de la production du discours ainsi qu'aux protagonistes du discours.

3.2.2.3. *Définitions du métadiscours de ses différents modèles*

Dans les deux parties précédentes, nous avons observé deux courants du métadiscours dans le monde de la recherche anglophone. Malgré les définitions différentes, nous identifions les points essentiels de la notion du métadiscours.

Premièrement, le concept de métadiscours insiste sur l'établissement d'une relation sociale entre l'auteur et le lecteur à travers le discours. Le lecteur exerce une forte influence sur les stratégies rhétoriques que l'auteur doit mettre en œuvre pour garantir la bonne réception du discours. Les définitions présentées plus haut, indiquent que le métadiscours sert à la compréhension du discours, tant en ce qui concerne l'organisation textuelle que sur le contenu propositionnel et les attitudes ou sentiments qu'il véhicule dans le discours. En résumé, quelle que soit l'approche où se développent les notions de métadiscours, la nécessité pour l'auteur de mobiliser certains éléments linguistiques afin de répondre aux attentes du lecteur constitue le pilier théorique qui motive l'utilisation des marqueurs métadiscursifs.

Deuxièmement, le métadiscours reflète la réflexivité de l'énonciation de l'auteur. Proche de la réflexion d'Authier-Revuz, elle concerne le discours en train de se produire. L'auteur est obligé de faire des discours sur son propre discours, afin de vérifier à tout moment l'intelligibilité de son texte. Hyland (2005, p. 45–47) parle de la relation interne des références du discours, tandis que Ädel préfère le terme « réflexivité » pour démontrer le fait que le métadiscours est le discours qui commente le discours avec le discours, renvoyant donc à la réflexivité énonciative du discours.

Troisièmement, à côté des fonctions interpersonnelles servant à tisser la relation entre l'auteur et le lecteur, une fonction principale du métadiscours consiste bien à guider la lecture. Cette fonction s'instancie à travers des fonctions textuelles chez Vande Kopple, Crismore, à

travers des fonctions interactives chez Hyland, et à travers des fonctions métatextuelles chez Ädel.

Dans notre étude, nous préférons une définition du métadiscours prenant en compte les trois remarques présentées ci-dessus. Le métadiscours constitue un moyen linguistique pour assurer l'interaction entre l'auteur et le lecteur sur le plan discursif. Il concerne donc le discours en train de se produire et permet à l'auteur de bien orienter l'interprétation de son discours à travers sa capacité de réflexivité énonciative. À côté des fonctions interpersonnelles servant à exprimer les sentiments, à tisser les relations sociales entre l'auteur et le lecteur, le métadiscours comporte l'ensemble des fonctions facilitant la lecture.

Nous allons maintenant examiner de plus près les fonctions identifiées par ces modèles du métadiscours, et isoler celles qui pourraient remplir une fonction de guidage du lecteur dans le discours.

3.2.3. Classements fonctionnels des marqueurs métadiscursifs

Il faut noter que non seulement la définition du métadiscours fait l'objet de débat, mais aussi les typologies qui en découlent. Nous rappelons que les modèles interactifs/interpersonnels et réflexifs ont proposé des définitions différentes, mais la mission essentielle du métadiscours est de faciliter la lecture et de tisser une relation interpersonnelle entre l'auteur et le lecteur. Pour atteindre cet objectif, des marqueurs linguistiques ont été classés fonctionnellement. Nous nous référons ici aux classements issus des modèles déjà évoqués, dont certaines classes qui montrent une utilité pour guider la lecture seront discutées de façon plus détaillée.

3.2.3.1. Classements fonctionnels des marqueurs métadiscursifs dans les modèles interactifs/interpersonnels

Vande Kopple (1985), Crismore *et al.* (1993) ont classé les marqueurs métadiscursifs en textuels et interpersonnels. Cette division prend son origine dans la théorie de la linguistique systématique fonctionnelle de Halliday, pour laquelle, il s'agit d'interpréter le langage d'un point de vue fonctionnel. Halliday a identifié trois métafonctions essentielles qui agissent simultanément pour construire un discours, à savoir les métafonctions idéationnelle, interpersonnelle et textuelle (Halliday & Matthiessen, 2014) :

- La métafonction idéationnelle (ou bien référentielle) indique la représentation du processus et des choses qui existent dans le monde subjectif et objectif. On parle même de la dimension propositionnelle qui vise à véhiculer des informations référentielles.

C'est aussi le niveau propositionnel du discours que nous avons déjà évoqué plus haut. Ce sont des éléments linguistiques informationnels ou représentatifs qui expriment les expériences du monde extérieur ou du monde intérieur de notre propre perception (Vande Kopple, 1985, p. 86).

- La métafonction interpersonnelle a trait à l'utilisation concrète du langage pour tisser les relations sociales et personnelles, reflétant ainsi la relation interpersonnelle. Cette fonction consiste soit à exprimer la personnalité ou les sentiments de l'auteur, soit à réaliser des interactions sociales avec les participants dans une situation de communication. Elle met en relief l'aspect social du discours (Halliday & Matthiessen, 2014, p. 29–30).
- La métafonction textuelle transforme des tronçons de discours parlé ou écrit en un texte cohérent et unifié. Elle réalise le passage à un sens articulé qui est différent d'une liste des phrases aléatoires. Elle unifie les deux premières métafonctions dans le texte. On peut même dire que c'est bien cette métafonction qui unifie les pratiques langagières et sociales en discours (Halliday & Matthiessen, 2014, p. 29–30).

Dans la mesure où ces linguistes pensent que le métadiscours ne touche pas le niveau propositionnel qui correspond à la dimension idéationnelle dans l'approche SFG (*Systemic Functional Grammar*), ils s'orientent naturellement vers les deux autres dimensions : textuelles et interpersonnelles. C'est sur ces deux niveaux que le métadiscours joue son rôle afin de rendre le discours plus accessible au lecteur. Leurs classifications fonctionnelles sont illustrées dans le tableau 3-5.

Les typologies du métadiscours élaborées par Vandekopple (1985), Crismore et al. (1993)	
Vandekopple (1985)	Crismore et al. (1993)
<p><u>Métadiscours textuel :</u></p> <p>Connecteurs de texte montrent la connexion entre les parties textuelles, comportent :</p> <ul style="list-style-type: none"> • Marqueurs d'ordre (<i>first, next, in the second place</i>) • Marqueurs de rappel (<i>as I mentioned in Chapter 2</i>) • Marqueurs de topique (<i>with regard to</i>) <p>Gloses : Explication, clarification, définition (<i>in other word, for example</i>)</p> <p>Marqueurs de Validité expriment l'engagement de l'auteur par rapport à la probabilité ou la vérité d'une déclaration, incluent :</p> <ul style="list-style-type: none"> • Marqueurs d'atténuation (<i>might</i>) • Marqueurs emphatiques (<i>clearly, undoubtedly</i>) • Marqueurs d'attribut (<i>according to Einstein</i>) <p>Narrateurs informent la source d'une information : (<i>according to Smith, the Prime Minister announced that</i>)</p>	<p><u>Métadiscours textuel :</u></p> <p>Marqueurs textuels</p> <ul style="list-style-type: none"> • Connecteurs logiques montrent la connexion entre des idées (<i>therefore, so, in addition</i>) • Marqueurs d'ordre indiquent l'ordre du texte (<i>first, next, finally, 1, 2, 3</i>) • Marqueurs de rappel réfèrent à la partie précédente (<i>as we saw in Chapter one</i>) <p>Marqueurs interprétatifs</p> <ul style="list-style-type: none"> • Gloses expliquent le contenu du texte (<i>for example, that is</i>) • Marqueurs illocutoires nomment l'acte réalisé (<i>to conclude, in sum, I predict</i>) • Marqueurs d'annonce annoncent le contenu à venir (<i>in the next section</i>)
<p><u>Métadiscours interpersonnel :</u></p> <p>Marqueurs illocutoires expliquent les activités discursives de l'auteur (<i>to conclude, to sum up, we predict</i>)</p> <p>Marqueurs d'attitude expriment l'attitude de l'auteur vis-à-vis du contenu présenté (<i>unfortunately, interestingly, I wish that</i>)</p> <p>Commentaires s'adressent directement au lecteur (<i>You will agree that</i>)</p>	<p><u>Métadiscours interpersonnel :</u></p> <p>Marqueurs d'atténuation montrent l'incertitude de l'auteur vis-à-vis à l'authenticité d'une assertion (<i>might, possible</i>)</p> <p>Marqueurs de certitude expriment la certitude absolue d'une assertion (<i>certainly, know, shows</i>)</p> <p>Marqueurs d'attribut donnent la source d'une information (<i>Smith claims that</i>)</p> <p>Marqueurs d'attitude affichent les états affectifs de l'auteur (<i>I hope/agree, surprisingly</i>)</p> <p>Commentaires établissent une relation avec le lecteur (<i>you may not agree that</i>)</p>

Tableau 3-5 : Les typologies lexicales du métadiscours de Vandekopple (1985) et de Crismore et al. (1993)

Vandekopple (1985, p. 87) estime que le métadiscours textuel montre « comment nous lions et relient les propositions individuelles afin qu'elles forment un texte cohérent et cohésif et comment les éléments individuels de ces propositions ont un sens en conjonction avec d'autres éléments du texte ». Le métadiscours interpersonnel « peut nous aider à exprimer nos personnalités et nos réactions au contenu propositionnel de nos textes et à caractériser l'interaction que nous aimerions avoir avec nos lecteurs à propos de ce contenu » (*ibid.*). Par conséquent, les catégories classées dans la dimension textuelle regroupent des marqueurs de fonctions visant à aider à mieux organiser le texte, à rendre le texte plus compréhensible pour le lecteur, et à expliciter la source des informations communiquées ou les actes discursifs réalisés par l'auteur. La dimension interpersonnelle couvre des marqueurs dont les fonctions

favorisent le maintien de l'interaction avec le lecteur ou l'expression des sentiments ou commentaires.

Ces deux typologies présentent des similitudes : les marqueurs d'ordre, de rappel, des gloses sont tous classés dans la catégorie textuelle qui vise à structurer le texte. Pour nous, ces marqueurs ont bien pour fonction de guider la lecture : les marqueurs d'ordre permettent de comprendre l'organisation des matériaux propositionnels du discours, ce qui rendra le texte plus agréable à suivre. Les marqueurs de rappel renvoient le lecteur à une partie textuelle antérieure ou à une question déjà traitée précédemment, ce qui permet de conserver la continuité du développement du texte. Les gloses ont plutôt une fonction métalinguistique qui vise à reformuler, clarifier, exemplifier, expliquer ou définir. Les marqueurs d'attitude et les commentaires sont tous étiquetés interpersonnels, grâce à leur fonction d'expression des attitudes de l'auteur et d'adresse directe au lecteur.

Les points de divergences sont cependant plus nombreux ici. Il faut signaler que Crismore et *al.* (1993) ont élaboré leur typologie en réorganisant les catégories de Vande Kopple. Par exemple, les marqueurs de validité, d'atténuation, emphatiques et d'attribut ont été séparés et déplacés dans la dimension interpersonnelle, alors que Vande Kopple les classe dans la dimension textuelle ; les marqueurs d'attribut reclassés absorbent les narrateurs ; les marqueurs illocutoires sont classés dans la dimension textuelle au lieu de la dimension interpersonnelle de Vande Kopple. Les connecteurs temporels ont été exclus, mais les connecteurs logiques sont maintenus, ce qui est justifié par l'argument de l'auteur selon lequel le métadiscours ne concerne pas l'aspect temporel des événements réalisés dans le monde réel, mais concerne la connexion logique dans le texte (Crismore et *al.*, 1993, p. 46). D'ailleurs, Crismore et *al.* (1993, p. 46–47) divisent la dimension textuelle en deux sous-classes : marqueurs textuels et interprétatifs. La première classe aide à organiser le texte tandis que la deuxième est prévue pour aider le lecteur à interpréter, à mieux comprendre le sens du texte et les stratégies rhétoriques de l'auteur (*ibid.*, p 46–47). Hyland (2005, p. 33) ne trouve pas nécessaire cette distinction, en soulignant que la fonction d'interprétation est inhérente au métadiscours textuel qui contribue à la bonne interprétation par la réalisation de la cohérence du texte. Certes, il est curieux de voir que les marqueurs d'annonce qui indiquent une partie textuelle à venir sont considérés comme interprétatifs, tandis que les marqueurs de rappel d'une partie textuelle antérieure font partie des marqueurs textuels. Pour nous, les marqueurs d'annonce ou de rappel ont tous une fonction de structuration et appartiennent donc au niveau textuel.

Le modèle de Hyland (2005) ne fait plus de différence entre les niveaux textuels et interpersonnels : il met en évidence la nature interpersonnelle du métadiscours. Sa catégorisation divise le métadiscours en niveaux interactionnel et interactif (*cf.* le tableau 3-6), une distinction empruntée aux deux aspects de l'interaction de Thompson (2001, p. 61, cité par Hyland, 2005, p. 44).

Typologie du métadiscours de Hyland (2005)		
Classes	Fonctions	Exemples
<p><u>interactif :</u></p> <p>Marqueurs de transition</p> <p>Marqueurs d'encadrement</p> <p>Marqueurs endophoriques</p> <p>Marqueurs d'évidentialité</p> <p>Gloses</p>	<p>Guider la lecture</p> <p>Connexion logique</p> <p>Se référer aux actes, séquences ou étapes du discours</p> <p>Référence intratextuelle</p> <p>Référence intertextuelle</p> <p>Explication et clarification</p>	<p><i>in addition; but; thus</i></p> <p><i>finally; my purpose is...</i></p> <p><i>noted above; in section 2</i></p> <p><i>according to X ...</i></p> <p><i>such as; in other word...</i></p>
<p><u>Interactionnel :</u></p> <p>Marqueurs d'atténuation</p> <p>Marqueurs de certitude</p> <p>Marqueurs d'attitude</p> <p>Marqueurs d'automenton</p> <p>Marqueurs d'engagement</p>	<p>Impliquer le lecteur dans le texte</p> <p>Refus d'engagement, dialogue ouvert</p> <p>Exprimer la certitude</p> <p>Exprimer l'attitude</p> <p>Références à l'auteur</p> <p>Établir une relation avec le lecteur</p>	<p><i>might; possible ; perhaps</i></p> <p><i>in fact; it is clear that...</i></p> <p><i>I agree; un fortunately...</i></p> <p><i>I; we; my; me, our ...</i></p> <p><i>consider; you cans see that...</i></p>

Tableau 3-6 : Typologie du métadiscours de Hyland (2005, p. 49)

La dimension interactionnelle implique les moyens que l'auteur utilise pour commenter le contenu, exprimer ses attitudes ou ses sentiments comme les marqueurs d'engagement *you can see that* (Hyland, 2005, p. 49), classés par ses prédécesseurs dans la classe interpersonnelle de commentaires. En dépit des noms différents, nous observons que ces deux classes exercent des fonctions identiques. La nouvelle classe des marqueurs d'automenton est envisagée pour recouvrir la référence à la présence de l'auteur dans le texte, dévoilée par les pronoms de première personne et les adjectifs possessifs (*I, me, my, mine, we, our, ours*). Ce choix permet de découvrir les stratégies de positionnement et l'identité scripturale de l'auteur (*ibid.*, p. 53). En fait, ce point nous paraît discutable, car on peut trouver d'autres traces de la posture de l'auteur (adjectif évaluatif par exemple).

La dimension interactive reflète la prise en compte du lecteur par l'auteur et indique les moyens linguistiques que l'auteur utilise pour guider la lecture (*ibid.*), et dans cette dimension se classent des marqueurs semblables à la catégorie textuelle citée plus haut. Comme Crismore, Hyland inclut également des marqueurs de transition⁵⁰ ayant pour fonction la connexion

⁵⁰ Ces marqueurs peuvent renvoyer non seulement à la relation interne au texte entre des arguments, mais aussi aux rapports entre des activités dans le monde extérieur comme ajouter des activités, comparer les

logique du texte. Ce sont des conjonctions ou des locutions ou des propositions adverbiales qui visent à aider le lecteur à interpréter la connexion pragmatique entre les étapes argumentatives à travers trois sous-catégories (Hyland, 2005, p. 50) :

- 1) Addition : ajoute des éléments aux arguments dans le texte ; cette fonction se réalise souvent par des conjonctions comme *and, furthermore, moreover...*
- 2) Comparaison : établit un lien entre deux arguments afin de faire ressortir des ressemblances (*equally, in the same way*) ou des différences (*in contrast, however*).
- 3) Conséquence : formule une conclusion (*in conclusion*) ou indique une déduction (*thus, therefore*), ou un contrargument (*nevertheless, anyway*) (*ibid.*).

Les marqueurs de transition dans les études sur l'anglais font partie des marqueurs discursifs qui permettent d'organiser le discours actuel, mais ils assurent davantage la cohérence logique. C'est pour cette raison que les exemples donnés sont tous des marqueurs logiques (*and, anyway, however...*). Nous reviendrons sur ce point dans la troisième partie où nous développerons notre propre typologie des routines de guidage du lecteur.

Les marqueurs d'encadrement visent à signaler les frontières textuelles et à ordonner des arguments, afin de construire la structure schématique du texte. Quatre fonctions sont retenues (Hyland, 2005, p. 51) :

- 1) Structuration : enchaîne les parties textuelles du discours ou ordonne des arguments internes du discours. Les marqueurs lexicaux représentatifs sont souvent liés au lexique des parties textuelles ou des séquences : *in chapter x, in part x, this section, first, firstly, second, secondly, to start with...*
- 2) Étiquetage (des étapes discursives) : marque explicitement des étapes discursives, p. ex. avec *to summarize, in sum, by way of introduction...*
- 3) Anticipation (annonce des objectifs) : dénote l'objectif principal du discours, p.ex. avec des expressions comme *I argue here, the paper proposes...*
- 4) Changement thématique : indique la transition thématique, avec des marqueurs tels que : *well, all right, now, let us return to...*

Les marqueurs endophoriques sont des expressions renvoyant le lecteur à une autre partie du texte. Comme les marqueurs de structuration, ils utilisent le lexique des parties textuelles comme *chapitre, figure, section*, etc. comme dans *refer to the next section, as noted*

événements, expliquer pourquoi et comment les choses se produisent (Hyland, 2005, p. 51). Nous nous intéressons ici seulement à ses fonctions discursives et logiques pour organiser les matériaux discursifs.

above... Avec ces marqueurs, l'auteur peut appuyer son argumentation en se référant soit au contenu déjà communiqué, soit à ce qui va être communiqué bientôt, ce qui permet de faciliter la compréhension et d'aider le lecteur à récupérer le sens exact exprimé par l'auteur.

Les marqueurs d'évidentialité consistent en un moyen métalinguistique qui permet de représenter les idées provenant d'autres sources. Cette fonction est en rapport avec l'attribution à la source d'un énoncé ou d'un point de vue qui pourraient renforcer l'argument de l'auteur. Nous pouvons aussi dire qu'il s'agit d'un renvoi intertextuel rappelant souvent le nom des chercheurs cités comme *according to x, to cite/quote x (ibid.)*.

Les gloses reflètent l'anticipation de l'auteur concernant les connaissances du lecteur par l'auteur. Leur fonction tend à fournir des informations supplémentaires en reformulant, expliquant ou développant ce qui a été dit pour assurer que le lecteur est capable de comprendre la pensée de l'auteur. Les marqueurs les plus courants sont des expressions de reformulation (*in other word*), d'exemplification (*for example*), d'explication (*that is to say, that means*), de définition (*defined as*), etc. (Hyland, 2005, p. 52).

L'observation pointue des marqueurs de guidage (marqueurs interactifs) ainsi que de leurs fonctions métadiscursives montre que l'auteur doit organiser son discours en répondant aux attentes du lecteur : le texte doit se construire conventionnellement selon un texte modèle avec une structure claire et une direction prévisible, et les arguments doivent s'organiser de façon appropriée et convaincante (Hyland, 2005, p. 54).

La littérature autour de ces trois typologies du métadiscours de modèle interactif/interpersonnel montre que les marqueurs de guidage concernent plus l'organisation et l'interprétation du texte que l'expression des émotions, des sentiments ou des points de vue.

3.2.3.2. Classements fonctionnels des marqueurs métadiscursifs dans le modèle réflexif

Dans le modèle réflexif, le métadiscours est fractionné en *Métatexte* et *Interaction auteur-lecteur* qui sont décrits ensuite par les configurations personnelles et impersonnelles (voir tableau 3-7). La catégorie *Métatexte* concerne les actes discursifs de l'auteur en lien avec le l'organisation, le rappel, l'annonce, la conclusion, etc. *L'interaction auteur-lecteur* se focalise sur la relation entre l'auteur et le lecteur, soit pour anticiper les attentes du lecteur, soit pour engager le lecteur dans le discours (Ädel, 2006, p. 36–37). La modalité impersonnelle fait référence au texte actuel et à la langue utilisée. La modalité personnelle fait référence soit à l'auteur du texte en train de se dérouler (les pronoms de première personne), soit au lecteur

imaginaire du texte actuel (pronoms de deuxième personne), soit à l'auteur et lecteur en même temps (les pronoms de première personne inclusifs).

Modèle réflexif du métadiscours chez Ädel (2006)			
Fonctions	Modalités	Références	Exemples
Métatexte	Impersonnel	Références au texte/code	<i>firstly, secondly, thirdly, ... the above mentioned list ... in other words ... the question is ... a definition of ... involves ... this essay will primarily deal with...</i>
		Références à l'auteur	<i>as I have shown ... my conclusion is that ... by... I mean ... as a writer, I would like to argue that...</i>
	Personnel	Références au lecteur	<i>as you have seen ... you might want to read the last section first...</i>
		Références aux participants	<i>as we have seen ... in our discussion above if we take ... as example...</i>
Interaction auteur-lecteur	personnel	Références aux participants	<i>I know you think that ... correct me if I am wrong, but...</i>
		Références au lecteur	<i>now, dear reader, you probably ... does this sound ... to you?</i>

Tableau 3-7 : Modèle réflexif du métadiscours d'Ädel (2006, p38)

Ce qui nous paraît surprenant ici, c'est qu'Ädel construise la frontière entre les valeurs personnelles et impersonnelles uniquement par la présence ou l'absence des pronoms personnels dans les marqueurs. Nous pensons que l'absence du pronom personnel n'implique pas le manque d'interaction ou d'engagement, comme dans les phrases : *as mentioned above*, *voir l'annexe 2*, etc.

Il est clair, selon cette typologie, que l'auteure a exclu des marqueurs de sentiments ou de commentaires que nous trouvons principalement retenus par les typologies précédentes, qu'elle juge appartenir plutôt au champ de positionnement (*stance* en anglais) (*ibid.*, p. 38–39).

Par rapport aux modèles interactifs qui proposent au préalable des listes de mots ou marqueurs isolés, associés à telle ou telle fonction en tant que fonctions de base applicables au discours étudié, le modèle réflexif ne fournit pas de liste des fonctions présumées, et seule une étude en contexte permet d'identifier les fonctions métadiscursives. Il s'agit là de la plus grande différence entre ces deux courants du métadiscours, qui est essentiellement une différence de nature méthodologique. Les modèles interactifs/interpersonnels s'appuient sur une approche plus quantitative en ce qui concerne l'identification du métadiscours dans les corpus à partir de la liste de marqueurs (Ädel & Mauranen, 2010). Quant au modèle réflexif, il prend souvent les pronoms personnels et certains noms (par exemple, *writer/reader*) comme indices pour détecter les marqueurs de métadiscours sans liste prédéfinie, notamment pour le

mode personnel, dans une approche qualitative orientée par les données (*ibid.*). L'extraction des marqueurs métadiscursifs personnels à l'aide des techniques textométriques nécessite un travail de tri et d'analyse qualitative qui prenne en compte leurs fonctions métadiscursives, car ces indices sont ambigus et ne revêtent pas toujours une fonction métadiscursive.

À titre d'illustration, nous nous intéressons ici aux fonctions identifiées dans une étude qui applique ce modèle réflexif à un corpus de textes écrits académiques en anglais. Bien que l'objectif de cette étude vise à comparer l'utilisation des marqueurs métadiscursifs entre les natifs et non natifs de l'anglais, elle propose une vue d'ensemble des fonctions métadiscursives dans l'écrit académique rédigé par des étudiants universitaires. Deux corpus ont été choisis, un sous-corpus d'articles rédigés par des apprenants suédois issu du corpus *International Corpus of Learner English* (ICLE), élaboré par Sylviane Granger à l'Université de Louvain (Belgique), et un sous-corpus d'articles écrits par des natifs de l'anglais issu du corpus LOCNESS –*the Louvain Corpus of Native English Essays*. Les résultats permettent d'identifier 16 fonctions dans le mode personnel et 9 fonctions dans le mode impersonnel, que nous récapitulons dans le tableau 3-8 :

Typologie des fonctions métadiscursives dans les écrits académiques en anglais (Ädel, 2006)			
Mode	Classes	Sous classes et fonctions	Exemples
Personnel	Métatexte	Sur le code : Définition Commentaire énonciatif Sur la structure du texte Introduction du thème Focus sur le thème Conclusion Exemplification Rappel Ajout Argumentation Contextualisation	<i>What do we mean by _ then?</i> <i>What I am saying is _</i> <i>In the course of this essay, we shall</i> <i>Now I come to the next idea which I</i> <i>In conclusion, I would say that _</i> <i>As an example of _, we can look at _</i> <i>As I mentioned earlier, _</i> <i>I would like to add that _</i> <i>The _ which I argue for is _</i> <i>I have chosen this subject because</i>
	L'interaction entre l'auteur et le lecteur	Anticipation de la réaction du public Clarification Alignement des perspectives Imagination des scénarios Hypothèse sur le lecteur Attirer l'attention du lecteur	<i>I do realise that all this may sound _</i> <i>I am not saying _, I am merely pointing out that _</i> <i>If we [consider/compare] _, we [can/will] [understand/see] _</i> <i>If you consider _, you can perhaps imagine _</i> <i>You have probably heard people say that _</i> <i>I hope that now the reader has understood _</i>
<u>Impersonnel</u>		Références au texte ou au code Marqueurs phoriques Étiquettes discursives Énonciation/définition Exemplification Conclusion Introduction du thème Argumentation Addition Glosses	<i>essay, paragraph, text, word</i> <i>above, again, already, back to, before below, earlier, following, here, later, next, now, over to</i> <i>as it were, call, express, mention, say, spoken, ask, question</i> <i>example, instance, such as</i> <i>conclude, sum</i> <i>aim, intend, present</i> <i>stress, underline</i> <i>add</i> <i>i.e./i e/ie, mean, namely</i>

Tableau 3-8 : Fonctions discursives du métadiscours dans les écrits académiques en anglais (Ädel, 2006)

Examinons d'abord les sous-classes impersonnelles. Les éléments métadiscursifs classés sont plutôt des items lexicaux simples. La classe 'Références au texte ou au code' renvoie explicitement soit au texte entier, soit aux paragraphes (*essay, text, paragraph*), aux mots ou expressions utilisés dans le texte (*word, sentence, phrase*) (Ädel, 2006, p. 108). Elle concerne donc tous les niveaux du texte. Les marqueurs phoriques correspondant à la classe 'Marqueurs endophoriques' de Hyland (2005), à la classe 'Marqueurs d'annonce' et à la classe 'Marqueurs de rappel' chez Crismore *et al.* (1993). Ils regroupent des mots qui réfèrent aux parties textuelles

à tout moment (Ädel, 2006, p. 101), permettant l'annonce des thèmes à suivre ou la navigation dans le texte et l'argumentation. La sous-classe 'Gloses' est commune à toutes les catégorisations que nous avons parcourues. Elle permet à l'auteur de préciser l'interprétation correcte d'un élément pour son lecteur (*ibid.*, p. 124). Les étiquettes discursives mettent en discours les actes discursifs que l'auteur a réalisés comme la définition, la conclusion, l'exemplification, etc. Ces fonctions sont retrouvées dans le mode personnel pour le métadiscours métatextuel, avec en plus, les trois classes 'Focus sur le thème', 'Argumentation' et 'Contextualisation'. Cette dernière décrit la situation d'énonciation qui « présente des traces de la production du texte, le texte étant explicitement désigné comme un artefact »⁵¹ (Ädel, 2006, p. 64). La classe 'Focus' rappelle le thème déjà traité et le souligne encore une fois. La classe 'Argumentation' met en évidence l'acte discursif argumentatif au lieu d'exprimer simplement un point de vue (*ibid.*, p. 71).

Le métadiscours interpersonnel se divise en deux orientations selon le triangle réflexif : orientation vers l'auteur (nommée "expressive" dans le triangle) et orientation vers le lecteur (nommée "directive" dans le triangle). La première orientation répond aux besoins expressifs de l'auteur pour anticiper la réaction du lecteur (*I do realise that _ may sound*), la clarification de ce qu'il vient de dire (*I am not saying*), et pour inviter le lecteur à partager son point de vue (*as we can see*) (Ädel, 2006, p. 76). La deuxième orientation concerne le renvoi direct au lecteur en créant des scénarios adéquats (*if you consider; imagine, dear reader*) dans la mesure où le lecteur peut voir les choses dans une perspective spécifique (*ibid.*, p. 76), en l'influençant par un appel émotionnel (*ibid.*, p. 78), en formulant des jugements ou hypothèses sur le lecteur imaginaire, par exemple sur sa connaissance, ses attitudes éventuelles par rapport au thème (*You have probably heard*) (*ibid.*, p. 77).

Il est à noter ici que les expressions classées dans le mode personnel incluent souvent des mots à fonction métalinguistique comme *text, paragraph, word, definition, aim, say, etc.* vus en tant que marqueurs métadiscursifs dans le mode impersonnel. De plus, ces mots partagent des traits sémantiques communs. En effet, nous pouvons dire que des items métalinguistiques tels que *above, below, here* indiquent le lieu textuel via une métaphore spatiale, *back to, before, earlier, following, now* sont plutôt issus d'une métaphore temporelle, tandis que *text, paragraph, word, definition* font partie des mots renvoyant à la matérialité ou aux parties du message. Nous pouvons ainsi conclure que les expressions plus longues

⁵¹ Notre traduction pour "xhibits traces of the production of the text, with the text explicitly referred to as an artefact".

pourraient s'obtenir à partir des mots simples qui les construisent. Cela permet de concevoir une méthodologie d'extraction pour des expressions longues, comme nous le verrons au Chapitre 4.

En résumé, le modèle réflexif ne s'instancie que dans un corpus donné, dans la mesure où les fonctions précises y voient le jour. Par exemple, dans une autre étude datée de 2010, Ädel récapitule 23 fonctions du métadiscours personnel dans le discours scientifique entre l'écrit et l'oral, dont certaines fonctions comme la réparation (*maybe I should've said* à l'oral), la reformulation (*if you want*, à l'écrit), la gestion terminologique (*by this we mean that* à l'oral) qui n'apparaît pas dans l'étude précédente. Cette étude peut cependant enrichir notre connaissance des fonctions métadiscursives dans les écrits scientifiques.

En fait, les fonctions discutées dans ces modèles représentent en général des fonctions primaires des marqueurs ou expressions illustrés. Dans de nombreux cas, il est possible pour un marqueur ou une expression de remplir plusieurs fonctions, et vice-versa. La multifonctionnalité des marqueurs métadiscursifs dépend souvent du contexte (Crismore *et al.*, 1993, p. 43 ; Ädel, 2006, p. 33). Par exemple, Ädel (2006, p. 59) pense que la phrase *we will discuss _ in the next section* peut avoir deux fonctions : référence au texte (*in the next section*), et annonce du thème (*we will discuss _*)⁵².

3.2.3.3. Les modèles métadiscursifs et leurs apports à notre étude

La littérature autour du métadiscours nous permet de comprendre que les typologies, tout comme les notions qui lui sont associées, ne sont pas des questions tranchées. Néanmoins, l'examen des principales typologies fournit une entrée féconde sur les fonctions remplies par le métadiscours, en particulier dans le discours scientifique.

Sans entrer dans le détail des classes s'adressant directement au lecteur, les fonctions métatextuelles et les fonctions visant à préciser les codes langagiers utilisés mettent en avant l'importance du guidage du lecteur en termes de structuration du texte et de meilleure orientation de l'interprétation. Les outils métadiscursifs constituent en l'occurrence un moyen linguistique efficace dans les écrits scientifiques, pour mieux aider le lecteur à interpréter le matériel discursif, ou pour en guider la lecture.

⁵² Ce type de phrases (phrase + complément de lieu textuel) est enregistrée comme deux marqueurs discursifs dans le modèle d'Ädel, mais Crismore *et al.* le considèrent comme une expression métadiscursive unique. Dans notre étude, nous recensons aussi ce type de phrase comme une occurrence et privilégions la fonction d'annonce comme la fonction primaire.

Certes, des éléments linguistiques classés dans ces modèles sont plutôt des éléments lexicaux, mais nous remarquons aussi des expressions plus longues, parmi lesquelles, certaines correspondent aux routines qui sont l'objet de notre étude. À titre d'exemple : une expression recensée dans la fonction d'introduction du thème (Ädel, 2006, p. 67) peut être considérée comme une routine dans notre définition.

In this essay	I We	will am/are going to would like to shall	try to attempt to	discuss mention present show write analyze	TOPIC
---------------	---------	---------------------------------------------------	----------------------	-----------------------------------------------------------	-------

Tableau 3-8 : Une expression d'introduction du thème (Ädel, 2006, p. 67)

Par ailleurs, nous avons remarqué la présence régulière des mots à usage métalinguistique partageant des traits sémantiques identiques dans les marqueurs ou expressions cités pour les catégories du métadiscours dans le modèle réflexif chez Ädel. Cette propriété pourrait également s'appliquer aux marqueurs classés dans les modèles interactifs/interpersonnels, notamment celui de Hyland (2005) qui propose une liste de marqueurs métadiscursifs associés aux fonctions.

Comme les fonctions dans les modèles de Vande Kopple (1985) et de Crismore et *al.* (1993) sont similaires, à celles du modèle de Hyland (2005), nous basons donc notre étude sur le classement de Hyland (2005) dans le cadre interpersonnel et les classements d'Ädel (2006, 2010), dans le cadre réflexif, afin de tirer parti de leur complémentarité. Ces classes fonctionnelles représentent deux apports essentiels pour notre étude :

- un apport théorique permettant d'éclaircir les fonctions qui pourraient jouer le rôle de guide de lecture dans les écrits scientifiques, ce qui aiderait à l'élaboration d'une typologie pour les routines de guidage ainsi que leurs définitions.
- et un autre apport d'ordre méthodologique, précisant de manière plus concrète la démarche d'identification et de tri des routines dans le corpus, ainsi que leurs analyses.

3.3. Synthèse

Dans ce chapitre, nous avons parcouru quelques classements des ressources phraséologiques en anglais et en français du discours scientifique. Ces classements comportent des fonctions servant à organiser le texte, à faciliter la lecture, mais n'abordent pas des fonctions de guidage d'une façon systématique. Nous introduisons donc la notion de métadiscours pour étudier systématiquement les fonctions éventuellement favorisant le bon suivi du texte pour le lecteur.

À partir de l'observation des définitions et des fonctions du métadiscours, nous pouvons souligner que cette approche constitue un moyen linguistique privilégié par l'auteur pour guider la lecture. Les éléments linguistiques classés dans ces modèles sont des éléments lexicaux isolés, mais aussi des expressions plus longues. Sans retenir un seul modèle, nous nous référons principalement au modèle de Hyland (2005) et à ceux d'Ädel (2006, 2010), en particulier pour les fonctions liées aux guidages de la lecture de Hyland et pour les fonctions métatextuelles et métalinguistiques d'Ädel. Nous les adaptons à notre corpus pour élaborer une typologie des routines de guidage du lecteur.

Conclusion de la partie I

Dans cette partie théorique, nous avons expliqué les trois approches complémentaires pour l'étude des routines de guidage du lecteur : l'approche phraséologique, l'approche dialogique et l'approche fonctionnelle. Notre travail s'inscrit d'abord dans la lignée des études phraséologiques. L'observation des caractéristiques linguistiques des routines sémantico-rhétoriques et des notions connexes (*routines conversationnelles*, *lexical bundles*, *segments répétés*, *motifs*, etc.) montre que les routines sémantico-rhétoriques possèdent des caractéristiques différentes des autres notions. Ce sont des expressions lexico-syntaxiques récurrentes avec des variations lexicales ou syntaxiques qui exercent des fonctions rhétoriques spécifiques dans le discours argumentatif. Les routines de guidage faisant partie des routines sémantico-rhétoriques partagent ces caractéristiques linguistiques, mais sont plus spécifiques en exerçant des fonctions d'accompagnement du lecteur.

Ensuite, nous avons introduit l'approche dialogique pour inscrire l'étude des routines dans le fonctionnement interactionnel des écrits scientifiques. Néanmoins, l'approche dialogique ne permet pas une étude systématique pour identifier des fonctions précises assurant le guidage. Pour ce faire, des études sur le métadiscours en anglais dans l'approche fonctionnelle ont été retenues, dont deux modèles sont adoptés, afin d'examiner les fonctions candidates de guidage.

L'observation des fonctions visant à guider la lecture dans les modèles métadiscursifs nous montre que les fonctions de guidage du lecteur facilitent la lecture par une meilleure organisation du discours, une meilleure explication du signe linguistique utilisé, du sens d'une phrase et une meilleure sélection des informations saillantes pour le lecteur. Elles peuvent se réaliser en général sur deux plans :

1) sur le plan de la structuration textuelle du discours scientifique : mieux organiser la structure du texte en ordonnant l'enchaînement propositionnel, en expliquant la structure du texte et l'objet/l'objectif du texte (*dans la partie suivante, nous montrons...*), en permettant au lecteur de mieux naviguer dans le texte (*voir l'annexe x*) ;

2) sur le plan propositionnel du discours : mieux orienter l'interprétation du matériel propositionnel chez le lecteur en rendant les informations plus claires ou en évitant des malentendus, par exemple, la définition (*le terme X est défini comme...*), la reformulation (*cela veut dire que...*), la mise en relief des informations essentielles (*il est important de remarquer...*), etc.

Notre étude vise à la fois l'élaboration d'une typologie fonctionnelle et la modélisation des routines de guidage. En outre, les modèles du métadiscours peuvent servir de boussole pour l'extraction, l'identification et le tri des routines de guidage dans le corpus, qui fait l'objet de la deuxième partie.

Partie II. Recueil, traitement des routines de guidage du lecteur dans le corpus TermITH et élaboration d'un modèle d'analyse

Dans la première partie, nous avons décortiqué les caractéristiques syntaxiques, sémantiques, énonciatives et fonctionnelles des routines de guidage du lecteur à partir de trois perspectives, phraséologique, dialogique et fonctionnelle. Nous pouvons voir que ces phénomènes phraséologiques dans les écrits scientifiques constituent un type d'expression d'un empan syntaxique relativement large, avec des réalisations lexicales variables, comme ce que montre la routine de (co)constat suivante :

il est [important|essentiel|central|fondamental] de

[noter|constater|relever|voir|mentionner|montrer] + Syntagme nominal|complétive que

Cette partie dédiée à la méthodologie s'intéresse à deux questions principales : quel corpus utiliser et comment identifier ces routines de guidage du lecteur dans ce corpus malgré leur complexité ? Par ailleurs, nous nous demanderons quel modèle d'analyse adopter afin de prendre en compte tous les aspects linguistiques des routines.

De nos jours, l'étude linguistique des phénomènes langagiers a recours aux corpus à grande échelle, quelle que soit la langue. Le domaine de la phraséologie scientifique tire également profit du développement récent du domaine de la linguistique de corpus et du traitement automatique des langues (voir une synthèse dans l'article de Bendinelli, 2017). La phraséologie scientifique permet aux linguistes de mettre à leur disposition des ressources langagières authentiques très riches, parmi lesquelles on peut citer le site Scienquest (Tutin & Grossmann, 2013), le site du Louvain EAP dictionary (Granger & Paquot, 2010). La linguistique de corpus les aide à vérifier leur intuition, pour obtenir des résultats plus objectifs avec une rapidité remarquable. Néanmoins, le travail des linguistes est indispensable pour vérifier les données brutes issues directement de l'extraction automatique à l'aide des outils textométriques, car la simple observation ne peut pas remplacer l'interprétation et la vérification des linguistes (Sánchez Cárdenas, 2010, cité par Yan, 2017, p. 62). Par conséquent, la linguistique de corpus sera utilisée par nous comme une aide méthodologique pour extraire

automatiquement les expressions candidates. Des tris manuels seront ensuite opérés par une analyse qualitative mise en œuvre après l'obtention des données brutes.

Le corpus constitue le socle sur lequel se fondent les observations en linguistique de corpus. Sa constitution et son annotation déterminent la réussite d'une recherche linguistique. Un corpus doit être organisé pour répondre à une question de recherche. Par exemple, pour décrire le lexique d'une langue (il s'agit d'un ensemble ouvert que l'on ne peut fixer de manière exhaustive), il faut un vaste corpus qui couvre de nombreuses situations de communications, c'est-à-dire de nombreux genres discursifs, afin de recenser autant d'unités lexicales que possible. Par contre, si on s'intéresse à une question linguistique sur un genre donné ou sur la traduction ou la comparaison entre deux langues, il vaut mieux constituer des corpus spécifiques pour le premier cas et des corpus parallèles ou comparables pour le deuxième cas. L'utilisation des méthodes textométriques doit également s'adapter à l'objectif de recherche. En résumé, l'exploitation des corpus pour extraire les objets linguistiques à étudier n'est pas toujours une tâche aisée. Elle implique la prise en considération de multiples facteurs : la constitution du corpus, la mise en place d'une méthodologie fiable de recueil des données, l'application des critères de sélections manuelles assurant l'exactitude des données à retenir. Une fois obtenues les données nettoyées, l'analyse des routines nécessite un cadrage analytique.

Cette partie ne comporte qu'un chapitre, le chapitre 4, qui aborde d'abord le corpus utilisé ainsi que sa construction et les annotations syntaxiques et sémantiques. Le matériel, les méthodes d'extraction utilisées et le nettoyage manuel des données extraites seront détaillés par la suite. Nous terminerons par la conception d'un modèle d'analyse pour les routines de guidage du lecteur en nous référant à quelques modèles pertinents empruntés aux multiples domaines concernés, à savoir le traitement automatique des langues (TAL), la didactique des langues et la linguistique.

Chapitre 4. Extraction, identification et modélisation des routines de guidage du lecteur

Identifier les routines de guidage du lecteur dans un corpus oblige à mener une réflexion sur plusieurs phases de la recherche méthodologique :

1. Quels corpus utiliserons-nous ? Quels en sont la pertinence et les avantages ?
2. Comment établir une ou des méthodes afin d'extraire efficacement des candidats de routines ? Avec quels matériels et outils ?
3. Comment contrôler l'exactitude des données ? Avec quels critères quantitatifs et qualitatifs lors de la sélection ?
4. Comment classer les routines et quel modèle d'analyse adopter pour leur description ?

4.1. Linguistique de corpus pour l'étude phraséologique

Nous voulons d'abord situer notre méthodologie dans la linguistique de corpus, qui est « une discipline empirique qui consiste à observer et analyser de manière quantitative des échantillons de langage rassemblés dans un format informatisé » (Zufferey, 2020, p. 5). Le développement de l'informatique permet aujourd'hui d'exploiter des corpus de grandes tailles afin de mener des recherches sur « l'utilisation réelle du langage dans les textes naturels » (Biber, Conrad, & Reppen, 1998, p. 1, notre traduction). Les outils textométriques profitent également des progrès de la science informatique, rendant la récolte des données plus aisée.

La linguistique de corpus mobilise d'abord la méthodologie empirique. Au lieu de se fonder sur les jugements des linguistes, elle se base sur « l'observation de données extérieures » (Zufferey, 2020, p. 6). L'utilisation des corpus organisés de façon informatisée peut vérifier les intuitions des linguistes sur l'utilisation réelle du langage et éviter le hiatus entre les deux. Comme Sinclair l'a remarqué :

Le contraste mis en évidence entre les impressions sur les spécificités de la langue observées par les individus, et les preuves réunies objectivement à partir de textes est vaste et systématique. Ce contraste conduit à supposer que l'intuition humaine sur la langue est très spécifique et ne reflète pas ce qui se passe réellement lorsque ces mêmes personnes utilisent effectivement la langue⁵³ (Sinclair, 1991, p. 4).

⁵³ Notre traduction, la citation originale est : The contrast exposed between the impressions on language detail noted by people, and the evidence compiled objectively from texts is huge and systematic. It leads one to suppose that human intuition about language is highly specific, and not at all a good guide to what actually happens when the same people actually use the language (Sinclair, 1991, p. 4).

Le recours aux corpus permet d'éviter des erreurs d'analyse causées par l'intuition individuelle des linguistes sur l'usage d'un phénomène langagier.

L'application des corpus dans l'étude linguistique concerne aujourd'hui de nombreux champs linguistiques, incluant des domaines comme la phonétique, la morphologie, la syntaxe, jusqu'à la sociolinguistique ou la diachronie, etc. (Zufferey, 2020, chapitre 2). L'étude du lexique et l'analyse du discours en sont même des applications privilégiées. L'application de la linguistique de corpus à l'étude lexicologique autorise non seulement le recensement du lexique général ou spécifique, mais permet aussi la découverte des patrons préfabriqués comme les collocations, les séquences discursives, les expressions préconstruites, et plus généralement les objets d'étude de la phraséologie. Par exemple, dans le champ lexicologique, l'emploi des corpus permet de compiler des dictionnaires pour les apprenants de langues. Nous pouvons citer ici les dictionnaires commerciaux Antidote⁵⁴, qui recensent 128 000 mots triés des ressources modernes et anciennes à l'aide de la technologie informatique, et proposent une grande richesse de données lexicales aux apprenants de français. Dans le domaine phraséologique, les études sur des expressions récurrentes comme les *lexical bundles*, les motifs etc. présentées dans le premier chapitre sont aussi les fruits de l'exploitation des corpus.

Ainsi, cette discipline combine des techniques analytiques quantitatives et qualitatives, qui favorisent les analyses relevant les singularités statistiquement saillantes et les régularités linguistiques en matière syntaxique, sémantique et fonctionnelle.

Notre étude s'intéresse à la phraséologie dans un genre bien précis – les écrits scientifiques, en particulier les articles de recherches. Ce domaine de la linguistique est indissociable de la linguistique de corpus (Pecman, 2004 ; Nesi & Basturkmen, 2006 ; Biber & Barbieri, 2007 b ; Biber, 2009 ; Simpson-Vlach & Ellis, 2010 ; Ken Hyland, 2012 ; Tutin, 2014 ; Kraif & Tutin, 2017, 2020 ; Novakova & Sorba, 2017 ; Jacques & Tutin, 2018 ; Diwersy *et al.*, 2021 ; la liste n'est pas exhaustive). Ces études illustrent clairement la pertinence de mettre en pratique la linguistique de corpus dans le domaine de la phraséologie.

Utiliser un corpus adéquat est la condition essentielle de la réussite d'une recherche sur la phraséologie scientifique. Nous présentons maintenant le corpus utilisé dans la présente étude.

⁵⁴ Antidote est un logiciel commercial visant l'aide à la rédaction. Il intègre des dictionnaires de définitions, des locutions, des synonymes, des antonymes, des occurrences, des champs lexicaux et de la conjugaison, etc. Ici, nous parlons de la version 10, voir le site : <https://www.antidote.info/fr/antidote-10/documentation/guide-utilisation/introduction/les-dictionnaires>.

4.2. Corpus TermITH-transdisciplinaire : constitution, annotation et intérêt

4.2.1. Composition du corpus

Pour notre étude, nous avons utilisé le corpus TermITH-transdisciplinaire (pour la simplicité, nous l'appellerons désormais TermITH). Ce corpus a été construit d'abord par Tran (2014) dans le cadre de sa thèse sur les séquences discursives (comme *en résumé, de sorte que...*), dans la continuité du projet Scientext,⁵⁵ mais avec une perspective didactique. Le corpus a ensuite été repris par Hatier (2016), avec pour objectif l'identification et l'analyse du Lexique Scientifique Transdisciplinaire (LST désormais, nous y revenons dans la section 4.3.3), pour le projet TermITH⁵⁶. Il est passé d'un corpus initial de 300 articles pour 2 680 000 mots, au corpus final de 500 articles pour environ 5 millions de mots. Ce corpus s'appuie sur une répartition équilibrée en 10 disciplines de Sciences Humaines et Sociales (SHS) : linguistique, psychologie, sciences de l'éducation, économie, sciences politiques, anthropologie, histoire, géographie, sciences de l'information et de la communication (SCI), et sociologie (Jacques & Tutin, 2018). Nous donnons ici dans le tableau 4-1 les chiffres détaillés :

Disciplines	Nombres d'articles	Nombres de mots
Anthropologie	50	493 988
Économie	50	417 944
Géographie	50	400 533
Histoire	50	773 170
Linguistique	50	425 952
Psychologie	50	417 846
Sciences de l'éducation	50	417 069
Sciences de l'information	50	399 007
Sciences politiques	50	548 222
Sociologie	50	540 630
Total	500	4 834 361

Tableau 4-1 : Répartition en disciplines du corpus TermITH (reproduction de Hatier, 2016, p. 58)

Ce recueil de textes n'est pas le fruit du hasard, il respecte certains critères qui ont présidé à sa constitution. En reprenant la définition suivante du corpus chez Rastier (2005, p. 30),

⁵⁵ Ce projet intitulé « Corpus et outils pour étudier le positionnement et le raisonnement de l'auteur dans les écrits scientifiques », est dirigé par deux professeurs du laboratoire Lidilem – Francis Grosmmann et Agnès Tutin. Il a réalisé la constitution de quatre sous-corpus en linguistique, psychologie, économie et sciences du langage.

⁵⁶ Le projet TermITH, « Terminologie et Indexation de Textes intégraux en sciences Humaines et sociales » vise à élaborer une méthode automatique d'indexation afin de désambiguïser les termes propres aux sciences humaines, qui comportent des termes de spécialités disciplinaires et ceux transdisciplinaires scientifiques. Voir son site avec ce lien transdisciplinaire <https://anr.fr/Projet-ANR-12-CORD-0029>. Les études menées se concentrent sur le lexique et la phraséologie scientifique transdisciplinaire, comme la thèse de Yan (2017) sur les verbes, la thèse de Hatier (2016), et notre étude sur les routines.

Un corpus est un regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés : (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'applications.

Tran (2014) a établi 4 critères pour la sélection des textes, afin d'assurer l'homogénéité du corpus (voir les détails dans Tran, 2014, p. 54–58), repris et complétés ensuite par Hatier (2016, p.56–57), que nous récapitulons ici :

1. Critère du sous-genre des articles de recherches. Ces articles proviennent d'abord des revues de classe A et B selon le classement de l'AERES⁵⁷. Ce choix a pour objectif d'assurer la qualité des articles, qui peut avoir un impact sur la qualité du corpus.
2. Critère synchronique. Les articles se situent entre 1997 et 2013 qui regroupent ceux du projet Scientext entre 1997 et 2006, ceux entre 2007 et 2011 élaborés dans le cadre de la thèse de Tran (2014), et ceux entre 2005 et 2013 ajoutés par Hatier (2016).
3. Critère de compétence des auteurs. Les articles sont écrits en langue française par des auteurs francophones experts, validés par le comité de lecture composé par des auteurs qui maîtrisent la rédaction en français, ainsi que l'écriture académique.
4. Critère thématique. Les articles abordent des sujets centraux des disciplines. Les sujets marginaux sont écartés afin de maintenir l'équilibre entre des domaines au sein d'une discipline.
5. Critère de variété des scripteurs. Les articles sont rédigés par des auteurs différents. Chaque auteur offre seulement un article, ce qui évite l'idiosyncrasie.

La prise en compte de ces 5 principes lors de la constitution du corpus TermITH présente des avantages significatifs pour notre étude. D'abord, les articles de recherche constituent le sous-genre le plus formel pour étudier les écrits scientifiques. Un corpus comportant seulement ce sous-genre permet de garantir l'homogénéité en matière de textualité discursive et d'énonciation. Le statut emblématique des articles de recherche pour le discours scientifique permet de relever l'usage habituel des expressions préfabriquées. Par ailleurs, limiter la

⁵⁷ L'Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES) avait pour objectif « l'évaluation des établissements d'enseignement supérieur et de recherche, des organismes et des unités de recherche, des formations et diplômes d'enseignement supérieur » (cité du site <https://www.rue-af.com/exposant/institutionnels/aere/>, consulté le 24 mai, 2022). Elle a proposé une liste des revues en Sciences Humaines et Sociales référencées publiée en juillet 2008, qui classe des rangs de revues allant de A jusqu'à D correspondant « à l'étendue du rayonnement des revues » (*AERES Liste revuesSHS vagueD2009-2012 1 .pdf*, consulté le 24 mai, 2022). Cette agence a été remplacée en 2013 par l'Hcéres (Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur)

sélection des articles à une période récente de 16 ans permet de fournir un corpus dont l'utilisation de la langue académique reste assez stable, ce qui garantit la perspective synchronique de l'étude linguistique, même didactique sur la phraséologie scientifique transdisciplinaire. Enfin, l'élimination des auteurs non francophones et non experts, la restriction du nombre d'articles d'un même auteur, ainsi que la généralité du thème abordé dans les articles sélectionnés régulent les biais culturels, stylistiques et domaniaux.

La sélection des ressources, bien que fondamentale, n'est que la première phase. L'annotation des textes à l'aide d'outils automatiques est également très importante pour élaborer les outils textométriques qui permettent d'extraire automatiquement les objets d'étude ciblés. Ce corpus, baptisé avec le nom du projet TermITH, a été analysé syntaxiquement en dépendances.

4.2.2. Annotation effectuée sur le corpus

Nous allons présenter maintenant le détail de l'annotation du corpus TermITH, qui a été réalisée en suivant deux étapes (Tran, 2014, p. 58–61 ; Hatier, 2016, p. 58–59) :

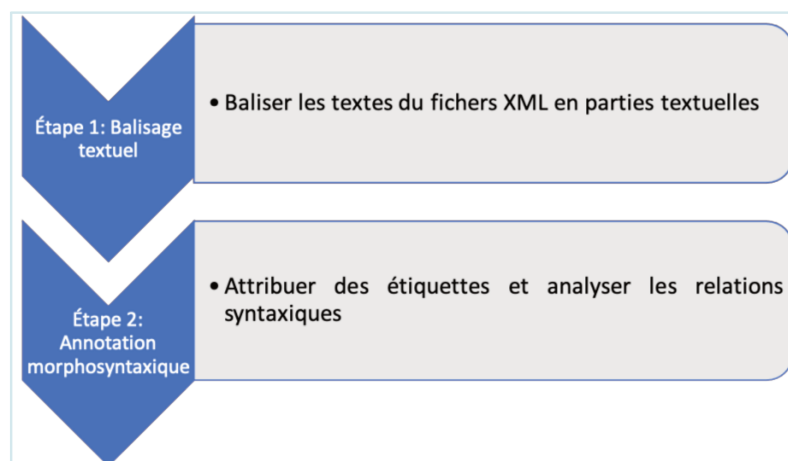


Figure 4-1 : Les deux étapes de constitution du corpus TermITH

Le balisage consiste à transformer les articles au format XML en suivant les prescriptions de la TEI-Lite. Cette opération consiste notamment à segmenter le document afin d'en dégager la structure hiérarchique, en identifiant les parties textuelles comme le résumé, les mots-clés, l'introduction, la conclusion, les notes, les annexes, etc.

Par la suite, le corpus a reçu une analyse morphosyntaxique et une analyse syntaxique à l'aide de l'analyseur en dépendance XIP – Xerox Incremental Parser (Aït-Mokhtar, Chanod, & Roux, 2002, cité par Hatier, 2016, p. 63). Hatier (2016) a choisi cet analyseur grâce à sa fonctionnalité performante autorisant l'intégration « de[s] grammaires adaptées, de[s] nouvelles

dépendances » (*ibid.*) par l'utilisateur dans l'intention de satisfaire ses besoins spécifiques, à savoir le recensement du lexique scientifique transdisciplinaire. Il s'agit donc d'une procédure automatique qui recourt à la fois à l'analyse proposée par XIP en matière d'annotation morphosyntaxique du corpus et à un post-traitement syntaxique en configurant de nouvelles règles pour raffiner des relations syntaxiques entre des tokens (Hatier, 2016, p. 64–66).

XIP fournit d'abord une analyse syntaxique profonde qui permet d'annoter les relations syntaxiques profondes. Ainsi, une liste des relations syntaxiques entre les tokens est fournie. Elle donne des informations sur les types de relations, les identifiants du gouverneur et du dépendant de la relation en suivant le principe des grammaires de dépendance. Hatier (2016, p. 66) donne la phrase *l'analyse de la phrase comprend une liste de tokens et une liste de relations* comme exemple pour l'analyse syntaxique de XIP :

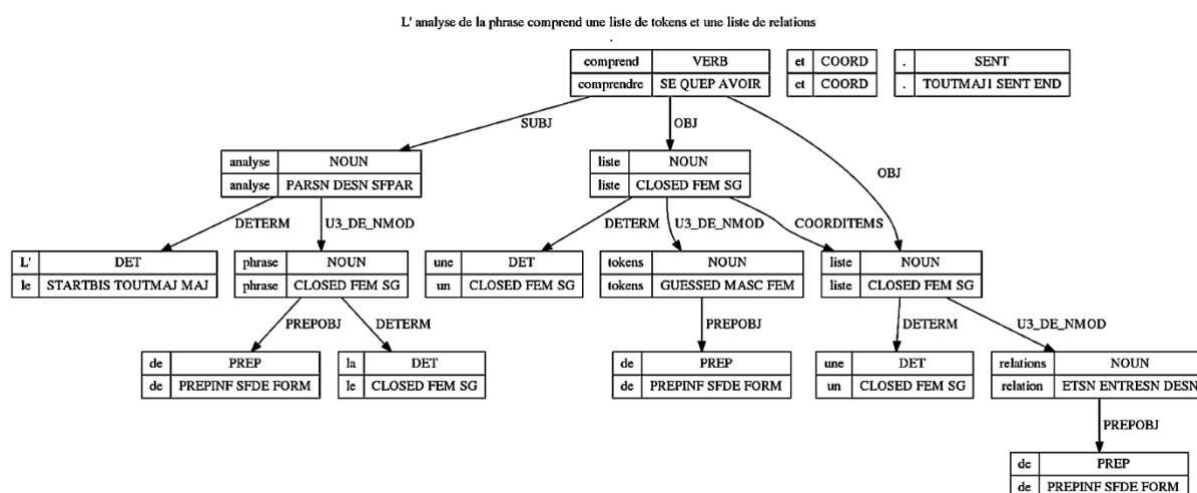


Figure 4-2 : Exemple d'analyse par XIP, repris de Hatier (2016, p. 66)

Après cette analyse automatique, un post-traitement d'optimisation pour raffiner l'analyse syntaxique a été réalisé⁵⁸. Plusieurs types d'améliorations ont été mis en place en observant en contexte des résultats issus de l'analyse automatique :

- L'ajout des relations de dépendance entre des tokens, indiqués par le préfixe *u3_* par exemple : la relation de dépendance adjectivale entre *le point de vue* et *culturel* et la relation d'objet profond entre *reconnu* et *enjeu* dans la phrase *parallèlement, le paysage constitue un enjeu reconnu d'un point de vue économique et culturel*.
- La réécriture des relations de dépendance pour distinguer les constructions

⁵⁸ L'écriture des règles a été réalisée manuellement par Hatier (2016), et le traitement est automatique.

différentes. Par exemple, la relation modifieur de nom (nmod) peut être affinée par les prépositions utilisées dans les contextes réels comme entre *manifestation* et *réforme*, leurs relations de dépendance peuvent être modifiées comme *pour_nmod* et *contre_nmod* pour les deux phrases *la manifestation pour la réforme aura lieu demain* et *la manifestation contre la réforme aura lieu demain*⁵⁹.

- Les normalisations et regroupements des relations de dépendance qui cherchent à adapter des étiquettes existantes au besoin réel par l'attribution aux tokens de nouvelles étiquettes, par le rejet des certaines relations, par la prise en compte des relations indirectes telles que dans les phrases avec les verbes supports et les auxiliaires, ou avec les pronoms relatifs.

Le corpus ainsi traité fournit donc des informations linguistiques suffisantes sur les aspects lexicaux et syntaxiques des parties du discours dans les articles de recherches en français. Plus précisément, les tokens peuvent recevoir des étiquettes telles que les catégories grammaticales (nom, verbe, adjectif, adverbe, etc.), les étiquettes syntaxiques (sujet, objet...) qui indiquent les relations de dépendance syntaxique entre eux. Voici un exemple pour la phrase *Sur un plan théorique, nous allons mobiliser les outils du néo-institutionnalisme* (Sciences politiques) :

id	forme	lemme	catégorie	Relations
0	Sur	sur	PREP	
2	un	un	DET	
4	plan	plan	NOUN	PREPOBJ->0 DETERM->2 U3_ADJMOD->6
6	théorique	théorique	ADJ	RIGHT_CON- TEXT->8
8	,	,	PUNCT	
10	nous	nous	PRON	
12	allons	aller	VERB	SUBJ->10 U3_SUR_VMOD ->4 OBJ->14
14	mobiliser	mobiliser	VERB	OBJ->18 SUBJ->10
16	les	le	DET	
18	outils	outil	NOUN	U3_DE_NMOD- ->29 DETERM->16
20	du	de	PREP	RIGHT_CON- TEXT->22
22	néo-	néo-	PREFIX	RIGHT_CON- TEXT->24
24	institutionna- lisme	institutionna- lisme	NOUN	RIGHT_CON- TEXT->26
26	.	.	SENT	

⁵⁹ Les prép. n'apparaissent qu'en tant que dépendant du nom modifieur. Si l'on s'intéresse à la relation binaire entre nom et complément de nom, la prép. n'apparaît donc pas : d'où l'intérêt de la faire remonter dans l'étiquette de relation.

L'utilisateur est donc capable de « recontextualiser les occurrences et les cooccurrences au travers de ces propriétés » (Hatier, 2016, p. 70). Cette capacité a permis également une meilleure représentation du LST retenu, en prenant en compte leurs différentes acceptions dans les écrits scientifiques, comme nous allons le préciser ultérieurement dans la section 4.3.2.

Le corpus TermITH annoté soigneusement avec ce post-traitement répond à nos besoins concernant l'extraction des routines pour deux raisons (ou deux avantages du corpus) :

1. Les structures des routines tissent souvent entre elles des relations syntaxiques plus complexes que les collocations (*formuler une hypothèse*) ou les séquences polylexicales (*à titre d'exemples*). Leur extraction et leur analyse nécessitent donc d'examiner les relations profondes qui dominent les composantes d'une routine.
2. Nous exploitons le LST élaboré à partir de ce corpus pour mettre en place une méthode semi-automatique (que nous détaillerons en section 4.4.3) afin d'extraire des candidats de routines.

Il faut noter par ailleurs que l'annotation syntaxique n'est pas sans défaut, car le traitement automatique réalisé par des outils d'annotation comme XIP comporte des erreurs tant au niveau de la segmentation, de la lemmatisation que des catégories et des relations syntaxiques (Hatier, 2016, p. 70). Le post-traitement mis en œuvre n'est pas non plus une solution miracle. Par conséquent, l'entreprise d'une analyse qualitative en contexte s'avère indispensable, afin de remédier à ces défauts techniques produits au cours de l'exploitation automatique.

Malgré ces inconvénients, l'utilisation d'un corpus arboré occupe une place primordiale dans la linguistique de corpus et le traitement automatique des langues (désormais TAL), qui permet aux ingénieurs des langues de créer des outils textométriques pour l'exploration comme le SKETCH ENGINE, le plateforme Scienquest⁶⁰, ou le Lexicoscope, que nous utilisons dans le présent travail.

4.3. Outil d'exploitation et la méthode des Arbres Lexico-syntaxiques Récurrents

Les routines sont des groupes d'expressions récurrentes présentant des traits syntaxiques et sémantiques homogènes. Leur extraction nécessite un outil textométrique performant et des

⁶⁰ Le plateforme Scienquest : <https://corpora.aiakide.net/scientext20/>.

méthodes adaptées. Dans cette partie, nous allons présenter d’abord l’outil d’exploitation et la méthode d’extraction intégrée dans cet outil.

4.3.1. Outil d’exploitation – Le Lexicoscope

Nous utilisons le Lexicoscope⁶¹, un outil d’exploration du corpus arboré en ligne élaboré par Olivier Kraif au LIDILEM, qui vise à « permettre un accès rapide aux cooccurents syntaxiques des unités étudiées, sans toutefois requérir la maîtrise d’un formalisme complexe pour la formulation des requêtes » (Kraif, 2016, p. 92). Plus simplement, c’est un outil qui ne se limite pas aux spécialistes de linguistiques de corpus, mais est aussi destiné à un public plus large de linguistes et de chercheurs pour l’étude des profils combinatoires du lexique. Il propose à la fois d’effectuer des requêtes simples par mot-clé, et des requêtes avancées en utilisant le langage de requêtes TQL et un multipivots. La requête simple consiste en une requête basique. Il suffit de saisir le mot ou l’expression à chercher pour obtenir des résultats. La requête TQL est plus avancée, et demande donc la maîtrise du langage de requête. La requête multipivots propose la recherche d’une catégorie grammaticale ou d’une position syntaxique précise pour un mot-clé (pivot de recherche). Prenons l’exemple du mot *définition* :

Types de requêtes	L’expression à saisir pour la requête <i>définition</i>
Requête simple	<i>définition</i>
Requête avancée (TQL)	<l=définition,c=NOUN,# 1>
Requête multipivots	Pour trouver les prépositions utilisées avec le mot <i>définition</i> , on saisit d’abord le mot, choisit NOUN comme catégorie et la relation PREPOBJ. On obtiendra alors les prépositions <i>par, dans, sur, selon</i> .

Tableau 4-2 : Requêtes proposées par le Lexicoscope

À part les trois modes de requêtes, le Lexicoscope combine encore trois fonctionnalités complémentaires que nous pouvons exploiter lors du recueil et du traitement des données :

- 1 Les lexicogrammes
- 2 Les concordances
- 3 La recherche d’expressions complexes

4.3.1.1. Lexicogrammes

Les lexicogrammes implémentés dans le Lexicoscope consistent en des tableaux de cooccurrences qui listent les collocatifs du pivot ainsi que leurs fréquences. Les collocatifs sont les mots qui cooccurrent fréquemment avec le pivot (mot-clé pour la requête). Ils sont extraits

⁶¹ Notons qu’il s’agit de Lexicoscope en version 1 : <http://phraseotext.univ-grenoble-alpes.fr/lexicoscope/lexicoscope.php>. Depuis le décembre 2020, la version 2 de Lexicoscope est mise en ligne pour fournir des requêtes optimisées et plus stabilisées.

à l'aide du calcul d'une mesure d'association statistique (ici le rapport de vraisemblance, noté *log.likelihood*) qui fait émerger les relations syntaxiques significatives entre le(s) pivot(s) et ses (leurs) collocatifs. Voici un exemple de requête du nom *définition* avec une partie du lexicogramme qui montre les dix associations syntaxiques les plus significatives dans le corpus TermITH :

I1	I2	f.deprels	f	f1	f2	N	f.disp	am.log.likelihood	r.log.likelihood
définition_NOUN	\$LE_DET	DETERM RIGHT_CONTEXT SEQNP	537	5928	394496	9736620	10	286,5016	1
définition_NOUN	\$UN_DET	DETERM U3_DEEPATTRSUJ ~U3_DEEPATTRSUJ	149	5928	122088	9736620	10	58,8984	6
définition_NOUN	par_PREP	PREPOBJ	87	5928	26656	9736620	10	151,6611	2
définition_NOUN	dans_PREP	PREPOBJ	78	5928	40999	9736620	10	72,2117	5
définition_NOUN	ce_DET	DETERM	60	5928	40292	9736620	9	36,6352	8
définition_NOUN	sur_PREP	PREPOBJ	40	5928	25488	9736620	10	26,9114	10
définition_NOUN	proposer_VERB	NMOD ~OBJ ~U3_SUR_VMOD ~U3_DEEPOBJ ~SUBJ ~DEEPOBJ ~U3_DE_VMOD ~U3_OBJ	38	5928	8565	9736620	8	85,6813	4
définition_NOUN	donner_VERB	NMOD ~OBJ ~U3_OBJ ~U3_DEEPOBJ ~VMOD ~U3_DE_VMOD U3_DEEPATTRSUJ	34	5928	10697	9736620	9	57,6011	7
définition_NOUN	situation_NOUN	U3_DE_NMOD U3_NMOD	21	5928	11255	9736620	6	18,7931	16
définition_NOUN	problème_NOUN	~U3_DE_NMOD U3_DE_NMOD ~NMOD	16	5928	6998	9736620	5	18,9058	15

Figure 4-4 : Tableaux de cooccurrences du pivot « définition »

Nous pouvons identifier des cooccurrences du pivot *définition* selon ce tableau. Les collocatifs les plus représentatifs regroupent quatre types d'éléments : les déterminants, subsumés par les classes \$LE_DET (déterminants définis), \$UN_DET (déterminants indéfinis), \$CE_DET (déterminants démonstratifs), les trois prépositions *par*, *dans*, *sur*, les noms (*situation*, *problème*) modifiant le pivot, et les verbes *proposer* et *donner*. Par ailleurs, les relations syntaxiques entre ces collocatifs et le pivot ont été enregistrées dans la colonne *f.deprels*. Pour le verbe *proposer*, 8 relations de dépendance ont été recensées.

4.3.1.2. Concordance

La concordance du Lexicoscope donne une représentation alignée des phrases extraites autour du pivot de recherche. En situant le pivot au centre, et ses contextes gauche et droit des deux côtés, la concordance du Lexicoscope fournit « le moyen d'entrer verticalement dans le corpus, afin d'observer les mots au croisement des deux axes perpendiculaires qui fondent

l'observation linguistique : l'axe syntagmatique et l'axe paradigmatique » (Kraif, 2019, p. 67–68). Si on reprend le pivot *définition* pour illustrer les concordances avec le collocatif *proposer*, nous obtenons les fragments textuels de la figure 4-5.

Num [▲]	Identifiant [◆]	Contexte gauche [◆]	Pivot [◆]	Contexte droit [◆]
1	s3548	Les sports à risque seront ainsi envisagés sous l' angle de la	définition	du risque proposée par Erving Goffman [1974 : 123] : « Dans la mesure où le jeu permet de remporter un prix , c' est une occasion , dans la mesure où il expose l' enjeu que l'on mise , c' est un risque .
2	s9277	Poursuivant la réflexion entamée par ces productions , il est proposé dans ce numéro une mise au point sur la	définition	des caractéristiques de l' image photographique qui intéressent l' anthropologue : la nature de la photographie , les fondements épistémologiques d' une méthode , les rapports entre écriture et image , son utilisation comme support à l' entretien , les choix de prises de vue ou la collecte de clichés sur le terrain .
3	s11276	Ainsi parle -t-on dans le plus grand désordre de pictogramme , d' idéogramme ou de logogramme , chaque auteur proposant une	définition	différente pour les deux premiers termes .
4	s10201	Il s' agissait de questionner la	définition	du paysage proposée par ces acteurs , leurs représentations de leur territoire forestier , ainsi que leurs pratiques de gestion .
5	s10294	La	définition	proposée de l' unité de gestion forestière , c'est-à-dire de la parcelle ou de la sous- parcelle en est une claire illustration : il s' agit en effet de la « division élémentaire de la forêt , la plus petite unité territoriale de planification (objectif et suivi de la gestion) , la plus homogène possible , notamment par ses conditions stationnelles [15] » .
6	s5637	Cet article propose plusieurs	définitions	de la « nature ordinaire » et souligne les enjeux de sa conservation et de son étude par la communauté des géographes .
7	s5642	This paper proposes several	definitions	of the `` ordinary '' natural environment , focusing on its conservation and its study by geographers .
8	s26044	En faisant appel non seulement à la « maxime » comme forme abstraite , mais à des moralistes modernes , l' auteur de l' Éloge s' appuie sur le développement de l' institution littéraire et participe à une intégration du roman dans l' univers des lettres , en proposant une	définition	du romancier comme moraliste .

Figure 4-5 : Extrait des concordances du pivot *définition* avec le collocatif *proposer* du Lexicoscope⁶²

Du point de vue syntagmatique, deux relations syntaxiques peuvent être identifiées dans cet extrait : l'objet profond *proposé* → *définition* dans les constructions passives, l'objet *proposer* → *définition* dans les constructions canoniques. Sur l'aspect paradigmatique, des déterminants (*une, la, plusieurs*), des noms et adjectifs modificateurs comme *risque, caractéristique, paysage, romanciers, différent* sont affichés. Au-delà des fragments de phrases listés, le Lexicoscope ouvre l'accès aux métadonnées des contextes élargis de chaque fragment de façon à ce qu'une vérification en contexte étendu soit faisable. Pour l'étude linguistique des marqueurs ou routines dans les écrits scientifiques, l'identification des fonctions remplies par ces phénomènes linguistiques constitue une question cruciale. Par exemple, l'adverbe *enfin* peut réaliser au moins deux fonctions : la fonction de conclusion et la fonction de temporalité

⁶² L'exemple 7 est une erreur qui enregistre des constructions en anglais. Nous les avons exclues lors du traitement des données brutes.

indiquant l'énumération ou la succession dans le temps. Entrer dans les contextes étendus permet d'identifier la fonction exacte d'*enfin*. Cette fonctionnalité est très utile pour le tri manuel des routines de guidage à partir des expressions candidates, car les expressions d'une même forme peuvent remplir plusieurs fonctions, pas forcément liées au guidage.

4.3.1.3. Recherche d'expressions polylexicales

La troisième fonctionnalité – la recherche d'expressions polylexicales permet de chercher des expressions complexes à travers les paramètres avancés définis par l'utilisateur. Le Lexicoscope intègre la formulation de requêtes avancées, dédiées davantage aux chercheurs voulant faire des recherches plus ciblées et adossées à leurs hypothèses. Ainsi, l'utilisateur peut créer des classes lexicales en nommant ces classes, et redéfinir des paramètres de filtrage des lexicogrammes tels que les seuils de fréquence, les seuils de mesure d'association (log-likelihood), la dispersion (nombre de disciplines dans notre étude), la longueur minimale des expressions, etc. Le résultat extrait est affiché en format HTML avec des accès aux listes triées des expressions, ainsi que leurs concordances et contextes étendus. Nous avons utilisé ce type de recherches paramétrées pour l'extraction des candidats de routines.

4.3.2. Méthode des Arbres Lexico-syntaxiques Récurrents

Si l'outil d'exploration du corpus discuté précédemment fournit des fonctionnalités avancées pour recueillir automatiquement des données brutes, il tire profit de la méthode des Arbres Lexico-syntaxiques Récurrents (ALR désormais) qui y est intégrée.

Contrairement à la méthode des n-grammes, ou segments répétés, qui cherchent des expressions consécutives de n caractères ou tokens, la méthode des ALR vise à chercher des séquences de mots définies par des relations syntaxiques profondes entre les éléments, quels que soit leur distance ou leur ordre dans la phrase. Ainsi, les ALR dépassent la limite de la linéarité en surface des séquences récurrentes, et identifient des structures hiérarchisées de dépendances syntaxiques (Kraif, 2016 ; Kraif & Tutin, 2017).

Nous pouvons décrire plus simplement l'algorithme de la méthode des ALR comme une opération qui se déroule à travers un processus itératif : on part d'un pivot initial pour extraire ses collocatifs, puis ceux-ci sont fusionnés avec le pivot et l'on relance la recherche des collocatifs sur les pivots fusionnés. Ce processus est réitéré jusqu'au moment où les collocatifs ne satisfont plus au seuil de significativité requis (Kraif & Tutin, 2017). Pour illustrer cette méthode des ALR, prenons encore *définition* comme exemple. Le processus illustré ci-dessous comporte trois itérations, et aboutit à l'ALR représenté par la figure 4-6 :

définition_NOUN → reprendre_VERB

définition_NOUN + reprendre_VERB → le_DET

définition_NOUN + reprendre_VERB + le_DET → on_PRON

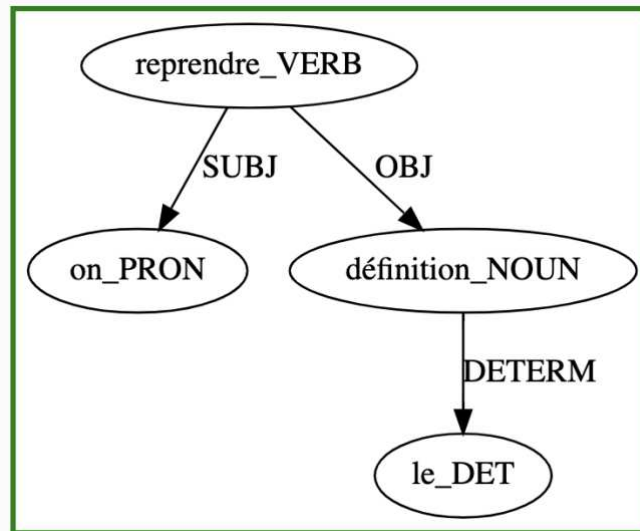


Figure 4-6 : ALR correspondant à on_PRON + reprendre_VERB + le_DET + définition_NOUN

S'il existe encore d'autres éléments comme les verbes modaux (*pouvoir, vouloir*), les adverbes qui peuvent s'intégrer dans cet arbre, le processus d'extraction itérative s'exécute automatiquement afin d'extraire tous les éléments répondant aux critères de significativité. Cette démarche d'extraction automatisée fournit ensuite des exemples triés du corpus TermITH dans le concordancier pour l'ALR de la figure 4-4 :

[1] **On reprend les définitions** de Saad et Schnetzler (1994). (Psychologie)

[2] **On pourrait reprendre la définition** de Pierre Lascoumes et de Patrick Le Galès en la transposant au champ de la participation. (Sciences politiques)

[3] **On peut ainsi reprendre la définition** qu'en donne Alain Lancelot. (Sciences politiques)

L'efficacité de l'application des ALR pour l'identification des routines a été testée par rapport à la méthode des n-grammes dans une étude comparant ces deux méthodes réalisée par Kraif et Tutin (2017). Cette étude a observé la performance des ALR pour le repérage des expressions phraséologiques formées autour des verbes de communication dans les écrits scientifiques. Les résultats montrent que les ALR sont plus efficaces que les n-grammes pour la détection des routines. C'est pour ces avantages de la méthode des ALR que nous l'avons choisie pour extraire les routines.

Comme le Lexicoscope et la méthode des ALR intégrée dans ce dernier étaient déjà disponibles, notre travail consiste en l'extraction automatique des candidats de routines, le filtrage et l'identification des données, et finalement l'élaboration d'un modèle descriptif pour analyser systématiquement les routines. Nous rassemblons les étapes de cette démarche dans la figure 4-7 :

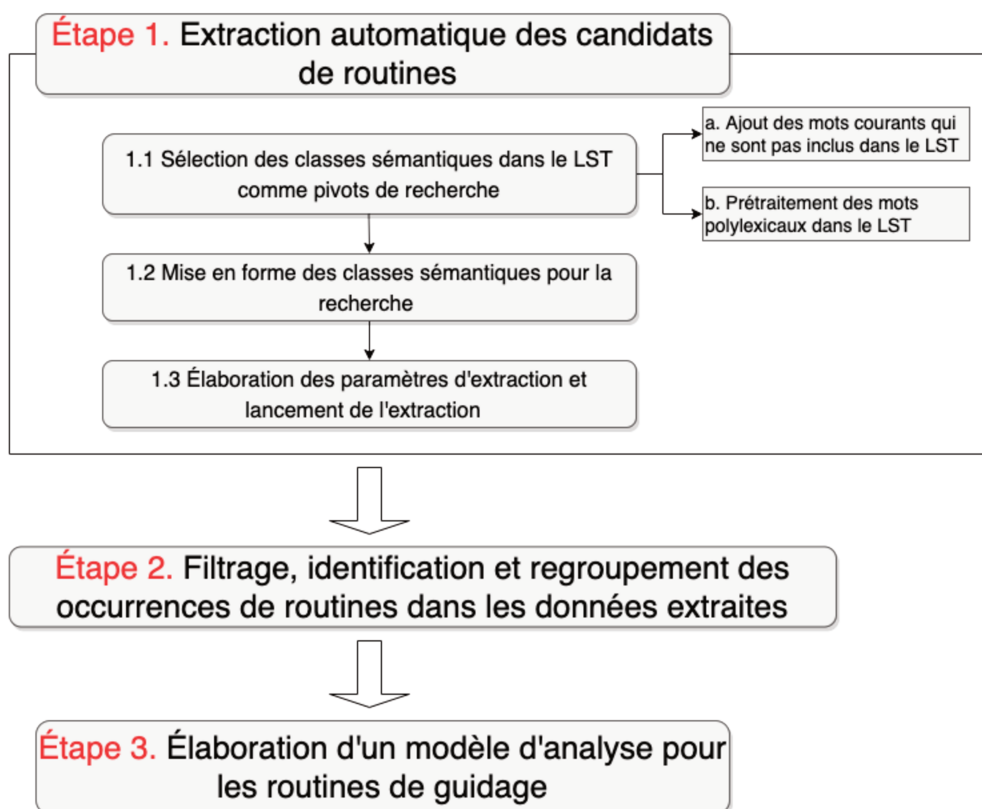


Figure 4-7 : Les démarches pour l'extraction et le traitement de routines de guidage

4.4. Étape 1. Extraction automatique des candidats de routines

Nous avons utilisé la recherche paramétrée proposée par le Lexicoscope pour effectuer l'extraction automatique des candidats de routines. Cette étape comporte encore trois sous-étapes qui ont permis d'adapter les requêtes à notre étude.

Notre recherche s'appuie sur une approche semi-inductive. Elle est guidée d'abord par les éléments définitoires des routines (*cf.* Chapitre 1) et les classes proposées dans les modèles métadiscours (*cf.* Chapitre 3), afin de sélectionner des mots-clés pouvant constituer les « noyaux », c'est-à-dire les parties fixes d'une routine de guidage. Dans l'exemple cité depuis le début de la présente section (l'exemple : *on reprend la définition*), le nom *définition* et le verbe *reprendre* sont des noyaux de cette expression. Pour avoir des données plus ciblées, nous avons choisi tout d'abord des mots-clés et les avons utilisés comme les pivots de la requête pour lancer une extraction automatique. Ces mots-clés proviennent du Lexique Scientifique

Transdisciplinaire⁶³ (LST) que nous allons présenter à la section suivante. Plus précisément, au lieu d'utiliser des mots simples, nous avons décidé d'exploiter les classes sémantiques du LST comme pivot de recherche, dans le but de découvrir les relations entre des dimensions sémantiques (ou entre des classes de mots) (Hatier *et al.*, 2016). Cette décision a été prise en tenant compte des variations lexicales des routines. Comme nous l'avons montré dans le Chapitre 1, les routines pourraient regrouper sur l'axe paradigmatique des réalisations lexicales du même champ sémantique (par exemple, les adjectifs dans la routine *il est [important|essentiel|central|fondamental] de [noter|constater|relever|voir|mentionner|montrer]*) appartiennent à la classe des adjectifs d'IMPORTANT_POSITIVE). C'est pour cette raison que nous avons utilisé finalement les classes sémantiques du LST. La sélection des classes sémantiques comme pivot de recherche constitue la première sous-étape, suivie par deux sous-étapes visant à adapter les classes sémantiques pour correspondre aux besoins de requête du Lexicoscope et à élaborer des paramètres d'extraction avant d'effectuer enfin l'extraction automatique.

4.4.1. Sélection des classes sémantiques dans le LST comme pivots de recherche

Le lexique scientifique transdisciplinaire est transversal aux différentes disciplines, indépendant des spécificités disciplinaires, et renvoie au discours sur les objets et les procédures scientifiques. Il constitue donc le « fond commun lexical propre à la démonstration » et au raisonnement scientifique (Tutin & Grossmann, 2013, p. 12).

Le LST enregistre environ 1600 mots simples et polylexicaux correspondant à quatre catégories grammaticales : verbes, noms, adjectifs, adverbes. Les mots sélectionnés validés comme appartenant au LST doivent répondre aux deux critères suivants (Hatier, 2016, p. 72–74) :

- (1) Critère de spécificité : ils ont des fréquences significativement plus élevées dans le corpus des articles de recherche que dans le corpus de contraste.
- (2) Critère de dispersion : ils se présentent dans au moins 40 tranches⁶⁴ sur 100 avec un taux de transdisciplinarité d'au moins 4/10.

⁶³ Le travail de l'extraction du LST a été effectué par un groupe de chercheurs du laboratoire Lidilem, dans le cadre du projet TermITH. Nous citons notamment l'ouvrage édité par Jacques et Tutin (2018) qui centralise tous les travaux autour du LST. Les ressources lexicales ont été mises en ligne : <http://lst.demarre-shs.fr/classification>.

⁶⁴ Hatier (2016, p. 74) a segmenté le corpus TermITH en 100 tranches égales. Comme ce corpus comporte 5 millions de mots, une tranche comporte environ 500 000 mots.

Ensuite, les unités lexicales ont été classées selon des critères sémantiques à partir des acceptions des mots. Comme un mot pourrait avoir plusieurs acceptions, il est possible de trouver des mots classés dans plusieurs sous-classes sémantiques. Par exemple, le nom *définition* possède deux significations lors de son usage dans les écrits scientifiques :

	Acceptions	Classes et sous-classes
Définition_1	Explication de mot	Classe : Communication_expression : Acte de communication et de transfert des idées (selon le type de message). Sous-classe : Énoncé Élément d'un énoncé.
Définition_2	Détermination des propriétés	Classe : Processus_cognitif Processus cognitif permettant de structurer/analyser les observables et/ou objets construits. <i>une personne fait/effectue/propose un N</i> Sous-classes : Détermination Processus de délimitation des propriétés de qqc.

Tableau 4-3 : Définitions dans deux classes sémantiques

Selon ce tableau, nous pouvons voir que dans l'exemple cité plus haut *nous reprenons la définition*, le mot *définition* est plutôt à prendre dans le sens 1 d'« explication de mot » plutôt que dans le sens 2. La classification sémantique du LST s'appuie sur plusieurs ressources en matière de lexique sémantique, ainsi que sur l'expertise des linguistes. À titre d'exemple, le classement du nom *définition* est basé sur le *Dictionnaire Électronique des Mots* de Dubois & Dubois-Charlier (2010, cité par Hatier, 2016). La typologie des verbes s'est faite cependant à l'aide du dictionnaire *Les verbes français* (LVF) de Dubois et Dubois-Charlier (1997, cité par Yan, 2017). Au total, 50 classes sémantiques et 267 sous-classes sémantiques ont été recensées.

L'utilisation des sous-classes sémantiques du LST nous permet d'obtenir des résultats plus ciblés et systématiques en interprétant les relations syntaxiques de dépendance entre des paradigmes lexicaux. Nous reviendrons sur ce point dans la section 4.4.1.3. Avant de sélectionner les classes sémantiques comme pivots de requête, il a fallu encore faire deux prétraitements du LST :

- a. Ajout des mots courants qui ne sont pas inclus dans le LST ;
- b. Prétraitement des mots polylexicaux dans le LST.

4.4.1.1. Ajout des mots courants qui ne sont pas inclus dans le LST

Comme le LST a été recensé à l'aide des critères de significativité dans les écrits scientifiques par rapport à un grand corpus de contraste qui est constitué de genres variés (littérature, presse, sous-titres, etc.) (Hatier, 2016), il exclut donc certains mots de la langue courante comme *voir*, *dire*, *revenir*, etc. qui très probablement permettent aussi de constituer des routines de guidage. En vue de combler cette lacune, nous avons extrait les verbes, les noms,

les adjectifs et les adverbes les plus fréquents repérés automatiquement dans le corpus Scientext⁶⁵, et complété les mots du LST en choisissant ceux qui pourraient participer à la construction d'une routine. Pour les verbes et les adjectifs, nous avons choisi ceux dont la fréquence absolue dépasse 100 occurrences, et pour les adverbes, plus de 50 occurrences. Comme les noms sont beaucoup plus fréquents, nous avons enfin choisi les 1000 noms les plus fréquents (pour lesquels la fréquence minimale est de 123). Le tableau ci-dessous résume ces statistiques :

	Nombres à observer	Fréquences minimales ⁶⁶
Verbes	502	100
Noms	1000	123
Adjectifs	458	100
Adverbes	298	50

Tableau 4-4 : Les mots les plus fréquents dans le Scientext

La procédure de sélection est la suivante :

1. Identifier les mots non inclus dans le LST.
2. Sélectionner les mots liés à la description scientifique ou à l'énonciation des activités scientifiques.
3. Observer les concordances des mots sélectionnés à l'étape 2 dans le corpus TermITH à l'aide du Lexicoscope, afin de retenir ceux qui pourraient constituer une routine de guidage.
4. Classer les mots retenus dans la classe sémantique la plus adéquate du LST, ou créer de nouvelles classes sémantiques.

Finalement, nous avons retenu :

- Pour les verbes, 37 entrées de plus, réparties dans 19 sous-classes sémantiques dont la classe « Passage », une nouvelle classe que nous avons créée.

⁶⁵ Les verbes, noms, adjectifs et adverbes sont recensés à partir de la plateforme ScienQuest qui propose des listes de mots en fonction de catégories grammaticales avec des nombres de fréquences. Nous l'avons choisi grâce à cette commodité.

⁶⁶ Ces seuils sont évidemment arbitraires, mais ils répondent à un besoin concret : enregistrer des mots généraux fréquents susceptibles d'entrer dans la construction des routines de guidage. Les unités finalement identifiées ne prétendent donc pas à l'exhaustivité, mais elles satisfont le critère de fréquence.

verbes ajoutés	classes sémantiques
<i>comprendre</i>	analyse info compréhension
<i>voir, savoir, connaître, remarquer</i>	analyse info constat
<i>ajouter</i>	analyse info réexamen
<i>préférer</i>	ancrage construction choix
<i>chercher, destiner</i>	ancrage construction finalité
<i>appeler</i>	ancrage-construction dénomination
<i>dire, parler, énoncer</i>	communication diffusion
<i>rester</i>	état existence
<i>rapporter</i>	état rapport
<i>entendre</i>	état représentation
<i>falloir</i>	modalité nécessité
<i>devoir</i>	modalité obligation
<i>passer, suivre, revenir, rappeler, poursuivre, retourner</i>	passage
<i>trouver, penser, percevoir, croire</i>	processus cognitif psycho opinion neutre
<i>manquer, oublier</i>	processus cognitif psycho opinion défavorable
<i>prendre, servir, reprendre</i>	processus humain usage
<i>orienter</i>	relation influence
<i>commencer, lancer</i>	temporalité début
<i>finir</i>	temporalité fin humain

Tableau 4-5 : Verbes ajoutés dans le répertoire du LST pour l'extraction des routines

- 8 adjectifs, dont la plupart sont des adjectifs numériques, qui peuvent apparaître dans certaines routines qui structurent le texte.

adjectifs ajoutés	classes sémantiques
<i>nombreux</i>	quantité grande
<i>trois, quatre, cinq, deuxième, troisième, quatrième dernier</i>	temporalité espace énumération

Tableau 4-6 : Adjectifs ajoutés dans le répertoire du LST pour l'extraction des routines

- Aucun nom et aucun adverbe n'ont été ajoutés après ces procédures de vérification.

4.4.1.2. Prétraitement des mots polylexicaux dans le LST

Dans la liste du LST, il existe encore des mots polylexicaux comme *mettre en relief, de premier ordre, au premier chef*, qui ne peuvent pas être utilisés comme amorce dans la procédure des ALR. Cela est dû au fait que l'analyseur XIP ne les a pas repérés comme des unités, mais comme un ensemble de mots simples. Cependant, le LST comporte de nombreux mots polylexicaux. Si l'on veut les utiliser, il faut régler ce problème. La solution que nous avons proposée consiste à projeter l'expression polylexicale dans le lemme de la tête de l'expression, c'est-à-dire, les mots polylexicaux vont recevoir des étiquettes catégorielles afin qu'ils travaillent comme une unité simple. Prenons l'exemple de l'expression verbale *prendre en considération*. On peut le trouver avec le Lexicoscope dans le corpus TermITH en utilisant l'expression de requête suivante :

<l=prendre, c=VERB, # 1><l=en, c=PREP, # 2>& &<l=considération, c=NOUN,
3> ::(PREPOBJ, 3, 2) (VMOD, 1, 3)

Après projection des polylexicaux⁶⁷, le lemme *prendre* va être remplacé par l'expression complète (seulement bien sûr pour les occurrences de ladite expression), et l'on pourra par la suite rechercher le lemme sans prêter attention aux relations syntaxiques internes à l'expression polylexicale, avec une requête comme :

<l=prendre en considération, c=VERB>

Cette reconfiguration des mots polylexicaux par un seul token simplifie beaucoup la requête correspondant aux routines, lorsqu'il y a des mots polylexicaux qui y participent. Enfin, nous avons procédé à l'annotation de 176⁶⁸ éléments parmi les 193 mots polylexicaux du LST dans le corpus.

L'ajout des mots courants, mais manquants dans le LST, ainsi que le prétraitement des mots polylexicaux du LST nous permet d'avoir une meilleure exactitude des données sélectionnées. Mais cela ne signifie pas que nous devons prendre toutes les classes sémantiques comme pivots de recherche lors de l'extraction.

4.4.1.3. *Sélection des classes sémantiques comme pivots de recherche*

Pour choisir des classes sémantiques à utiliser en tant que pivots de recherche, nous nous sommes référée aux modèles du métadiscours discutés dans le chapitre 3. Nous partons de marqueurs lexicaux classés dans ces modèles pour les fonctions de guidage de la lecture (dans le modèle de Hyland), les fonctions métatextuelles et métalinguistiques (dans le modèle d'Ädel) et arrivons à sélectionner des mots clés avec les classes sémantiques correspondantes. Plus précisément, bien que les marqueurs soient en anglais, leur traduction en français peut souvent remplir des fonctions similaires, comme les marqueurs endophoriques pour la référence intratextuelle qui couvrent souvent des mots renvoyant aux parties textuelles du discours comme *section*, *partie*, *chapitre*, *etc.* Dans ce cas, nous pouvons choisir la sous-classe COMMUNICATION_SUPPORT_SECTION (*annexe*, *bibliographie*, *chapitre*, *conclusion*, *développement*, *extrait*, *introduction*, *note*, *section*, *page*) et la sous-classe COMMUNICATION_SUPPORT_GRAPHIQUE (*figure*, *illustration*, *image*, *tableau*).

Cette méthode impliquant les classes sémantiques n'est pas récente, Hatier (2016) a déjà testé sa faisabilité pour en extraire des routines, et Jacquy *et al.* (2013, p. 126) ont aussi remarqué la nécessité d'utiliser des classes sémantiques pour filtrer des données : une

⁶⁷ Merci à Olivier Kraif, mon co-directeur, pour son aide dans l'annotation des 176 mots polylexicaux dans le corpus TermITH, dans le cadre de cette étude sur les routines de guidage.

⁶⁸ Les 17 mots polylexicaux qui n'ont pas été traités sont plutôt des adverbes. Leur présence n'a pas d'incidence ni sur l'extraction, ni sur la construction des routines.

« classification sémantique des noms du lexique transdisciplinaire apparaît donc indispensable pour effectuer un filtrage [des termes] plus efficace » (*ibid.*).

À l'aide de l'observation des données, nous avons finalement retenu 65 catégories parmi 267, et ajouté une classe « passage » avec 7 verbes *passer, suivre, revenir, retourner, rappeler, poursuivre, revoir* (voir l'annexe 1 pour les détails ces 66 classes).

4.4.2. Mise en forme des classes sémantiques pour la recherche

Nous avons choisi finalement 66 sous-classes sémantiques pour être pivots de requêtes⁶⁹. Mais les classes non sélectionnées sont aussi utilisées dans les paramètres que nous allons présenter plus tard comme les classes lexicales de répertoire.

La figure 4-8⁷⁰ montre comment nous avons pris en compte ces classes pendant l'extraction des données : la déclaration des *pivots* est dédiée aux classes sémantiques que nous avons choisies (66 classes), la déclaration des *lexicalClasses* comporte les 268 classes formatées. De la sorte, les relations entre le pivot \$VPROCESHUMUS et les autres classes pourront d'être extraites par la méthode des ALR.

```

1 pivots=$VPROCESHUMUS#VERB
2 corpus=TERMITH_Yujing_FR
3 pivot_attr=l#c
4 colloc_attr=l#c
5 lexicalClasses=$JE=(je|nous|on)//$PAS=(pas|plus|jamais|guère)//$SON=(son|leur|mon|ton|notre|
votre)//$UN=(un|une|des)//$LE=(le|la|les|l|u|ux|es)//$ADJAPPAREXCLU=(autonome|distinct|
exclusif|extérieur|indépendant|individuel|isolé|séparé)//$ADJAPPARINCLU=(constitutif|
indissociable|inhérent|particulier|propre|respectif|spécifique)//$ADJAXION=(critique|
défavorable|flou|insuffisant)//$ADJAXIOP=(ad hoc|adéquat|approfondi|approprié|apte|cohérent|
correct|de premier ordre|de qualité|détaillé|en profondeur|exemplaire|explicite|favorable|
fiable|fin|idéal|intéressant|logique|pertinent|privilegié|profond|propice|reconnu|remarquable|
rigoureux|satisfaisant|strict|suffisant|utile|valable)//$ADJCHANGN=(fixe|stable)//
$ADJCHANGP=(dynamique|variable)//$ADJCOMPLEXN=(simple)//$ADJCOMPLEXP=(complexe|difficile|
élaboré|problématique)//$ADJCOMPOS=(divers|global|hétérogène|homogène|multiple|partiel|simple|
total|varié)//$ADJESPDIMEN=(court|étroit|grand|large)//$ADJESPLOCA=(central|externe|
horizontal|interne|proche|vertical|voisin)//$ADJESPMOUV=(fixe)//$ADJEVALU=(ambigu|attentif|
bref|conforme|décisif|définitif|effectif|exhaustif|général|informel|marqué|négatif|net|objectif|
opérationnel|porteur|positif|possible|précis|prégnant|radical|relatif|subjectif|véritable)//
$ADJIMPORIN=(limité|marginal|négligeable)//$ADJIMPORIP=(central|critique|crucial|de base|de

```

⁶⁹ Le Lexicoscope nécessite une déclaration des classes selon un format adapté à l'outil. Après la mise en forme, la classe des verbes de PROCESSUS_HUMAIN_USAGE (*accéder, adapter, appliquer, employer, emprunter, exploiter, mettre en œuvre, mobiliser, pratiquer, recourir, tirer profit, utiliser, prendre, reprendre, servir*) se présente comme :

```
$VPROCESHUMUS=(accéder|adapter|appliquer|employer|emprunter|exploiter|mettre en
œuvre|mobiliser|pratiquer|recourir|tirer profit|utiliser|prendre|reprendre|servir)
```

Toutes les classes du LST ainsi que la classe ajoutée (verbes de PASSAGE) doivent respecter cette règle de formatage. Ici, « \$ » est obligatoire et signifie qu'il s'agit d'une classe lexicale, « VPROCESHUMUS » est le nom abrégé que nous avons attribué à cette classe lexicale.

⁷⁰ Dans la figure, le corpus est nommé TERMITH_Yujing. Il s'agit toujours du même corpus TermITH, mais avec le prétraitement des mots polylexicaux discuté plus haut. En conséquence, sauf pour ce point, la constitution, l'annotation syntaxique et sémantique du corpus sont identiques. La dénomination de ce corpus pendant l'extraction consiste seulement à distinguer le corpus que nous utilisons de celui non prétraité.

Figure 4-8 : Formatage des classes sémantiques du LST pour le Lexicoscope

De cette manière, la méthode des ALR combinée avec les groupements lexicaux permet d'observer des relations de cooccurrence entre les dimensions sémantiques. Nous pouvons reprendre l'exemple d'un ALR cité plus haut <on_PRON + reprendre_VERB + le_DET + définition_NOUN>. Les positions du verbe *reprendre* et du nom *définition* sont remplacées respectivement par la classe PROCESSUS_HUMAIN_USAGE (*accéder, adapter, appliquer, employer, emprunter, exploiter, mettre en œuvre, mobiliser, pratiquer, recourir, tirer profit, utiliser, prendre, reprendre, servir*) et la classe COMMUNICATION_EXPRESSION_ENONCÉ (*affirmation, définition, discours, expression, formulation, formule, mot, proposition, terme, vocabulaire, reformulation, phrase, citation, notion*). Le nouvel ALR est illustré par la figure 4-9⁷¹ :

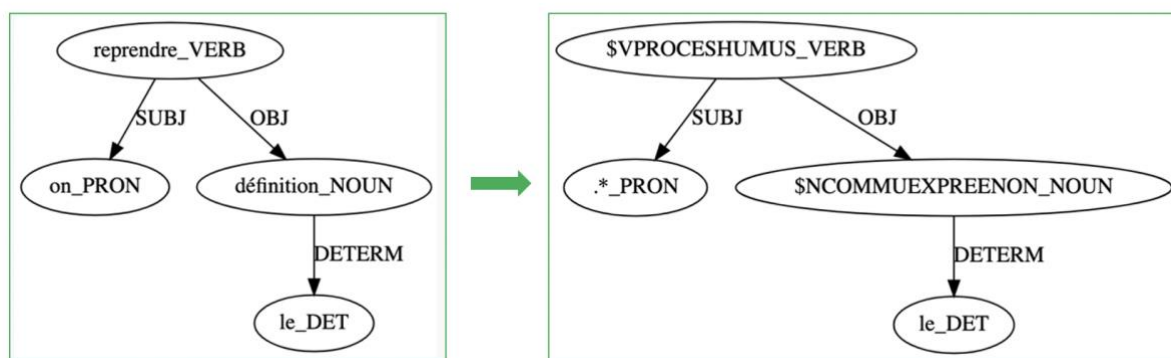


Figure 4-9 : Un ALR avec des classes sémantiques

Notre méthode semi-inductive avec une présélection des classes sémantiques ayant le potentiel de configurer une routine de guidage donne des résultats assez satisfaisants. Ce nouvel ALR, par exemple, permet de dégager des expressions comme :

- *J'emploie le terme de Cyrulnik...*
- *On reprend les définitions de Saad et Schnetzler...*
- *Nous reprenons le terme à Philippe Le Guern...*

Après cette présélection des classes sémantiques comme pivots et la mise en forme de toutes les classes, il est nécessaire de fixer les paramètres pour extraire les expressions récurrentes.

⁷¹ Dans le Lexicoscope, les classes lexicales définies par l'utilisateur doivent commencer par \$ suivi le nom de classe en Majuscule. Dans notre cas, nous avons simplifié les noms des classes dans LST pour faciliter la présentation.

4.4.3. Élaboration des paramètres de l'extraction automatique et lancement de l'extraction

Nous avons pu accéder directement au fichier qui spécifie les paramètres d'extraction du Lexicoscope, et nous avons donc reconfiguré des paramètres pour qu'ils répondent à nos critères de sélection. En général, 4 paramètres nous concernent ici :

1. taille maximale des arbres ;
2. seuil de fréquence pour les itérations 1 à n ;
3. seuil de dispersion ;
4. seuil de mesure d'association.

Nous avons d'abord fixé la taille maximale des arbres extraits à 8, c'est-à-dire que les expressions les plus longues comporteront 8 mots. D'après notre observation sur les études précédentes portant sur les routines sémantico-rhétoriques (*cf.* section 1.2.6 dans le premier chapitre), les routines ont souvent entre 3 et 8 mots. Nous pensons qu'une proposition contenant 8 mots est déjà très large.

Le seuil de fréquence pour les itérations de 1 à n indique précisément la fréquence minimale du collocatif de chaque étape d'itération des sous-arbres, comme expliqué dans la section 4.3.2 de ce chapitre. Avant de fixer ces seuils, il faut connaître préalablement la fréquence minimale pour justifier qu'une expression est récurrente. Comme les routines sont souvent longues, nous avons fixé 5 comme étant la fréquence minimale. Ce seuil de fréquence est le fruit d'une réflexion à partir de l'étude de Tran (2014) pour les séquences discursives et les études déjà effectuées sur les routines (Ji & Tutin, 2019, 2020). Tran (2014) a fixé la fréquence minimale à 10 pour des séquences discursives qui sont souvent moins complexes que les routines. Celles de Ji & Tutin (2019, 2020) s'appuient sur un seuil de 4 occurrences. Mais après observation des données dans leurs études, nous remarquons qu'il existe très peu de routines dont la fréquence est inférieure à 5⁷². Nous pouvons donc élever au moins la fréquence jusqu'à 5.

Les seuils sont fixés respectivement à : 10, 5, 5, 5, 3, 3, 3. Pour le pivot, la première extraction doit faire apparaître des collocatifs (collocatif 1) avec une fréquence au moins égale à 10. La deuxième itération doit extraire des collocatifs (collocatifs 2) du sous-arbre < pivot-collocatif 1 > avec une fréquence au moins égale à 5, et ainsi de suite. Ces seuils permettent de

⁷² D'autant plus que notre classement des routines permet de regrouper des routines dans les travaux de Ji & Tutin (2019,2020) en une routine plus générale.

traiter les expressions de 5 mots plus de 5 fois dans le corpus. Quant aux 5^{ème}, 6^{ème}, 7^{ème} collocatifs, ils doivent apparaître au minimum 3 fois dans l'ALR. Nous avons fixé un seuil moins élevé que 5 pour ouvrir la possibilité d'observer plus de données, spécialement les éléments périphériques (nous y reviendrons dans la section suivante). Ces seuils, fixés empiriquement, sont évidemment étroitement liés à notre corpus, et notamment, à sa taille : ils ne sont aucunement généralisables à un corpus quelconque.

Pour que les expressions extraites soient bien transdisciplinaires, nous avons choisi de fixer le seuil de dispersion à 3 disciplines au minimum sur 10 – rappelons ici que la dispersion indique le nombre de disciplines où l'ALR apparaît. Nous pensons qu'une routine doit être repérée dans au moins trois disciplines pour qu'elle soit jugée « transdisciplinaire ». Le seuil d'association (*log.likelihood*) a été fixé à 10,83. Il s'agit d'une estimation de vraisemblance qui permet de considérer que la fréquence d'un collocatif par rapport à un pivot ne peut être seulement due au hasard⁷³. L'association haute est indiquée par un chiffre haut, ce qui montre que l'expression extraite n'est pas aléatoire, mais résulte bien d'associations deux à deux marquées dans le corpus.

La méthode d'extraction automatique n'est pas sans problème. Ainsi, elle ne permet pas d'extraire des constructions impersonnelles qui sont pourtant très courantes dans les écrits scientifiques. Il s'agit de la construction *Il est à* + Verbe (*il est à noter/constater...*), car la relation syntaxique entre les éléments de cette routine est très compliquée à détecter avec la méthode des ALR, en suivant le modèle syntaxique de XIP : *noter* est à la fois le verbe modifieur du prédicat *être*, et l'objet de la préposition *à*. Pour remédier à cette lacune, nous avons confectionné la requête TQL suivante :

```
<l=il, c=PRON, # 1>& &<l=être, c=VERB, # 2>& &<l=à, c=PREP, # 3>&
&<l=$VANACONSTAT|SVANADEMON|VCOMMUDIF,c=VERB,#4>::(PREPOBJ, 4, 3)
(SUBJ, 2, 1) (U3_A_VMOD, 2, 4)
```

Ici trois classes de verbes ont été choisies :

- \$VANACONSTAT (verbes d'ANALYSE_INFOMATION_CONSTAT) : *constater, identifier, noter, observer, prendre acte, prendre connaissance, relever, repérer, voir, remarquer* ;
- \$VANADEMON (verbes

⁷³ Avec une p-value de 0,001, cf. <http://ucrel.lancs.ac.uk/llwizard.html>.

d'ANALYSE_INFORMATION_DEMONSTRATION) : *démontrer, mettre au jour, mettre en évidence, mettre en exergue, mettre en lumière, mettre en relief, mettre en valeur, établir, montrer* ;

- \$VCOMMUDIF (verbes de COMMUNICATION_DIFFUSION) : *dire, parler, citer, diffuser, exposer, exprimer, garantir, indiquer, mentionner, préciser, reporter, reformuler, signaler*.

Nous avons aussi effectué une requête simple pour trouver les routines autour de *cf*. C'est l'abréviation d'origine latine (*confer*) qui est souvent utilisée dans les articles de recherches pour renvoyer le lecteur à un endroit précis. Néanmoins, elle n'est pas classée dans le LST, ni dans les mots les plus fréquents du corpus Scientext. Nous l'ajoutons finalement dans nos données afin d'avoir des données plus complètes.

Les résultats de l'extraction automatique à l'aide de la recherche paramétrée sont listés dans un fichier HTML, à travers une arborescence qui peut interactivement être contractée ou développée, afin d'afficher des exemples de réalisation en contexte, comme l'illustre la figure 4-7. Pour les 66 classes sémantiques de pivots, nous avons obtenu un tel fichier HTML. Chaque fichier HTML possède une liste triée des expressions comme le montre la figure 4-10. Cette liste fournit toutes les combinaisons entre le pivot (verbes de PROCESSUS_HUMAIN_USAGE, désormais verbes d'USAGE) et les autres classes sémantiques du LST. Par conséquent, elle peut comporter des collocations, des patrons et aussi des routines. Dans ce cas, il faut établir des critères de sélection pour le filtrage, l'identification et le regroupement des occurrences de routines de guidage, qui font l'objet de la section suivante.

Liste triée des expressions (terminales ou non)

- ☐ Expression : \$VPROCESHUMUS_VERB \$JE_PRON \x5d_PUNCT \$ADVESPLOCA_ADV (5) ▾
- ☐ Expression : \$VPROCESHUMUS_VERB \$NCOMMUEXPREENON_NOUN \$JE_PRON \x5d_PUNCT \$ADVESPLOCA_ADV (3) ▾
- ☐ Expression : \$VPROCESHUMUS_VERB \$NCOMMUEXPREENON_NOUN \$JE_PRON \x5d_PUNCT (7) ▾
- ☐ Expression : \$VPROCESHUMUS_VERB pour_PREP \x5d_PUNCT \$NCOMMUEXPREENON_NOUN \$LE_DET (6) ▾
- ☐ Forme canonique : \$ADVDISAD se \$VPROCESHUMUS il (p.ex. : encore s' prend -elle) (6) ▾
- ☐ Forme canonique : \$ADVDISCONCLU , avoir \$VPROCESHUMUS (p.ex. : donc , ont servi) (5) ▾
- ☐ Forme canonique : \$ADVTEMPFREQ \$VPROCESHUMUS \$NCOMMUEXPREENON (p.ex. : souvent utilise phrases) (6) ▾
- ☐ Forme canonique : \$ADVTEMPFREQ \$VPROCESHUMUS par \$NOBJSCIENEXPLISIMP (p.ex. : souvent utilisé par exemple) (6) ▾
- ☐ Forme canonique : \$ADVTEMPFREQ \$VPROCESHUMUS (p.ex. : souvent utilisé) (207) ▾
- ☐ Forme canonique : \$JE \$LE avoir \$VANCRAFINA (p.ex. : nous l' avons employé) (7) ▾
- ☐ Forme canonique : \$JE \$LE avoir \$VPROCESHUMUS (p.ex. : nous l' avons utilisé) (7) ▾
- ☐ Forme canonique : \$JE \$VANAEXM (p.ex. : l'on prend en considération) (9) ▾
- ☐ Forme canonique : \$JE \$VANCRAFINA \$NCOMMUEXPREENON (p.ex. : on emploie terme) (25) ▲

```

graph TD
    A($VPROCESHUMUS_VERB) -- OBJ --> B($NCOMMUEXPREENON_NOUN)
    A -- SUBJ --> C($JE_PRON)
  
```

Search:

on emploie terme ▲

1981 [8] On emploie en Suède le terme de « bourgeois » pour désigner les partis de droite et du centre .

6105 Fruits du travail de plusieurs protagonistes , ces grossesses sont en outre le plus souvent multiples , et d'ailleurs lorsqu' il n' y a qu' un fStus , on emploie le terme « singleton » .

Showing 1 to 2 of 2 entries

Search:

Figure 4-10 : Extraite du fichier HTML pour le pivot \$ PROCESHUMUS

4.5. Étape 2. Filtrage, identification et regroupement des occurrences de routines dans les données extraites

Nous avons présenté dans la section précédente notre méthode de recueil automatique dans le corpus TermITH. Néanmoins, l'extraction automatique ne suffit pas pour identifier les routines de guidage du lecteur. D'une part, les expressions extraites par la méthode des ALR peuvent contenir toutes sortes de séquences récurrentes, comme des collocations, des locutions, des patrons, des fragments d'expression et pas seulement des routines. D'autre part, l'emploi des classes sémantiques comportant des marqueurs de fonction de guidage ne garantit pas forcément la présence de la fonction visée dans une routine. De ce fait, les données brutes nécessitent un tri manuel afin d'identifier des routines pertinentes.

La méthode d'extraction automatique des ALR avec l'apport des classes sémantiques du LST permet de moissonner des expressions candidates. Mais en observant des données extraites, nous avons remarqué comme Tutin & Kraif (2016) que les résultats de l'extraction automatique comportent beaucoup de bruit à écarter. Nous avons effectué un filtrage manuel en observant les occurrences en contexte pour identifier les routines pertinentes à partir des fichiers HTML. Dans ce processus de traitement manuel, des critères quantitatifs et qualitatifs ont été appliqués :

1. Nous n'avons classé que des routines comportant au moins 5 occurrences comportant au moins 3 mots qui apparaissent dans 3 disciplines au minimum⁷⁴.
2. Nous avons effectué la vérification des candidats routines en contexte, en particulier pour identifier leurs formes linguistiques, leurs paramètres énonciatifs et leurs fonctions.
3. Nous avons regroupé les occurrences d'une routine en fonction de critères lexicosyntaxiques et sémantiques, au lieu des rôles sémantiques pour présenter des formulations des structures syntaxiques plus simples.

Dans les fichiers HTML, en observant les formes canoniques des expressions extraites, nous pouvons très vite exclure :

- a. des patrons syntaxiques qui ne constituent pas des routines, par exemple, dans la figure 4-7 :

\$ADVDISCONCLU⁷⁵, avoir \$VPROCESHUMUS (p.ex. : *donc, ont servi*)

des collocations :

\$ADVTEMPFREQ⁷⁶ \$VPROCESHUMUS (p. ex. : *souvent utilisé*)

- b. des expressions employant des pronoms personnels de la troisième personne (*il, elle, ils, elles*). Ces expressions qui ne s'inscrivent pas dans le scénario scientifique ne

⁷⁴ Bien que l'extraction automatique soit fixée avec une fréquence à 5, le nettoyage manuel peut éliminer des occurrences non pertinentes. C'est-à-dire qu'une expression de 5 occurrences (ou plus) après l'extraction automatique, peut comporter en réalité moins de 5 occurrences valides, dans ce cas, nous n'avons pas pris cette expression. C'est pour cette raison que nous avons encore une fois insisté ici sur la fréquence minimale de 5 occurrences.

⁷⁵ Adverbes DISCURSIFS_CONCLUSION : *donc, en conclusion, enfin, in fine, pour conclure*.

⁷⁶ Adverbes de TEMPORALITE_FREQUENCE : *à l'occasion, à plusieurs reprises, chaque jour, constamment, couramment, de nouveau, de temps en temps, des fois, encore une fois, fréquemment, habituellement, la plupart du temps, le plus souvent, parfois, plusieurs fois, rarement, régulièrement, sans cesse, souvent, tous les jours, une fois de plus, une nouvelle fois*.

portent pas sur le discours en cours (sans actualisation énonciative) : elles ne présentent pas d'intégration interlocutive, et renvoient plutôt à l'énonciation interdiscursive qui fait référence aux autres auteurs ou aux faits scientifiques déjà observés par d'autres chercheurs.

II \$VPROCESHUMUS \$NOBJSCIENMETHO⁷⁷ (p.ex. : elle emprunt des voies)

Ce qui nous intéresse, ce sont des propositions qui parlent du discours en cours, ayant au moins une fonction rhétorique, comme :

\$JE \$VPROCESHUMUS \$NOBJSCIENMETHO (p.ex. : *on utilise méthode*) (1)

ou

\$JE \$VANCRAFINA \$NCOMMUEXPEREENON⁷⁸ (p.ex. : *on emploie terme*) (2)

On exclut aussi des routines qui n'ont pas pour fonction de guider la lecture comme (1) [*je|nous|on*] [*utiliser|adaptermobiliser|appliquer*] [*méthode|approche|procédure*], qui remplit plutôt une fonction de 'faire un choix méthodologique', comme le montre l'exemple 1.

[1]Pratiquement, **on utilise une méthode** analogue à celle des trapèzes, en remplaçant l'intégration par une sommation sur 50 intervalles. (Économie)

Pour des routines de guidage, il faut également entrer dans les contextes pour une vérification plus précise, en particulier pour les routines avec les pronoms personnels renvoyant à l'auteur ou au lecteur (*nous, je, on*). Il faut exclure les pronoms *nous* et *je* dans les citations ou dans les extraits d'entretiens, etc. qui ne renvoient pas à l'auteur. Le filtrage sur *nous* et *je* est relativement simple, tandis que le pronom *on* mérite une étude minutieuse en contexte, car il peut désigner des référents humains extérieurs à la situation de communication, mais aussi l'auteur du discours scientifique. L'objectif de la vérification consiste à éliminer des occurrences qui ne concernent pas l'auteur, donc ne concernent pas le guidage, mais plutôt une représentation d'une connaissance scientifique. Dans l'exemple 2, *on utilisait le terme* ne peut pas être considéré comme relevant d'une routine métalinguistique d'emprunt terminologique (une classe de notre typologie, nous y reviendrons dans la partie analyse), car le pronom *on* n'est pas actualisé énonciativement dans le discours en production : il ne renvoie pas à l'auteur ni au lecteur, mais se généralise sans référent précis. Par conséquent, dans l'exemple cité

⁷⁷ Noms d'OBJET SCIENTIFIQUE_MÉTHODOLOGIE : *analyse de discours, analyse de régression, approche, cahier des charges, démarche, dispositif, grille de lecture, logique, méthode, méthodologie, plan, procédé, procédure, programme, stratégie, théorie, voie.*

⁷⁸ Noms de COMMUNICATION_EXPRESSION_ÉNONCÉ : *affirmation, définition, discours, expression, formulation, reformulation, formule, mot, phrase, proposition, terme, vocabulaire.*

l'expression *on utilisait le terme* renvoie plutôt à une situation générale de l'époque en question, et nous l'avons exclue de nos données.

[2]Jusqu'en 1960, **on utilisait** de préférence **le terme** « ethnologie ».
(Anthropologie)

Par ailleurs, le regroupement des occurrences en tant que routine nécessite d'appliquer des critères lexico-syntaxiques et sémantiques. Il faut rappeler ici qu'une routine n'est pas une simple phrase, mais un groupe de phrases que nous appelons « occurrences de routine ». En effet, les routines peuvent comporter plusieurs types de phrases en respectant les mêmes rôles sémantiques, comme *nous reprenons la définition de Duschmoll et la définition adoptée ici est celle de Duschmoll* (Kraif & Tutin, 2017) (*cf.* section 1.2.6). Par commodité, nous n'avons cependant pas effectué le regroupement des occurrences de routines à partir des critères de rôles sémantiques. Notre regroupement est plutôt basé sur les matériels lexicaux et les paramètres syntaxiques. Nous avons affaire à deux cas de figure :

1. Cas de figure 1 : quand les occurrences comportent des matériels lexicaux de même paradigme et de même modalité phrastique, elles sont regroupées dans une seule routine, comme *on peut préciser, on peut noter, nous constatons que*.
2. Cas de figure 2 : Quand les phrases ont le même matériel lexical, mais pas de même modalité, le cas est plus complexe. Nous regroupons les occurrences d'abord en fonction de nature phrastique, à savoir, la phrase déclarative, la phrase injonctive et la phrase interrogative. Mais il faut préciser quelques points spécifiques. Par exemple, pour les occurrences *on peut préciser, on peut noter, il faut préciser, il est à noter*. Les deux premières phrases sont de nature déclarative principale, tandis que les deux autres sont aussi déclaratives, mais ont une fonction injonctive. Dans ce cas précis, nous classons les quatre occurrences dans deux routines différentes d'une même fonction rhétorique. Par contre, les occurrences aux phrases déclaratives principales, subordonnées et clivées ou pseudoclivées partageant les mêmes paradigmes lexicaux peuvent se classer comme la même routine. Ainsi, les occurrences *il faut préciser X et c'est X qu'il faut préciser* constituent une même routine.

Ainsi, une fonction rhétorique peut être réalisée par plusieurs routines qui comportent à leur tour plusieurs occurrences qui correspondent de manière générale à la même construction syntaxique et partagent les mêmes paradigmes lexicaux du LST.

À l'aide de ces critères de tri, nous avons éliminé les expressions formellement valides, mais énonciativement ou fonctionnellement non valides en les classant en tant que routines.

En outre, nous avons ajouté une dernière étape qui consiste à élargir l'observation sur les routines de guidage déjà identifiées, afin d'ajouter des extensions intéressantes. Deux types d'extension ont été pris en compte : l'extension lexicale des constituants essentiels d'une routine et l'extension des éléments facultatifs et périphériques d'une routine. Par exemple, pour une routine d'annonce de l'objectif de l'étude comme :

[l'objectif|le but|l'objet] de [cet article|ce papier|cette contribution]+ [est de|consiste à]+
[analyser|examiner|explorer...]

Les verbes apparaissant dans la dernière partie de la routine sont souvent des verbes d'ANALYSE_INFO et des verbes de COMMUNICATION. Mais quand nous avons élargi l'observation sur cette position, nous avons retenu deux verbes supplémentaires qui ne sont pas enregistrés par le LST, mais qui sont pourtant très intéressants : *élucider*, *dresser*.

Voici une occurrence avec le verbe *élucider*, ci-dessous :

[3] Partant de l'idée que la notion d'apprentissage connote en priorité la posture de l'élève, et que la notion d'alternance connote en priorité l'organisation pédagogique, **l'objectif de cet article consiste à élucider** les liens et séparations entre ce qui procède des modalités d'enseignement et ce qui procède des processus d'apprentissage. (Sciences de l'information)

Pour retenir ces verbes dans nos données, au lieu de créer de nouvelles classes, nous les avons intégrés dans les classes existantes, en adéquation avec leur contenu sémantique. Par exemple, *élucider* a été classé dans les verbes de RÉVÉLATION⁷⁹, *dresser* a été considéré comme un verbe de RÉALISATION⁸⁰.

Le deuxième type d'extension des routines s'intéresse aux éléments facultatifs ou périphériques. Il s'agit souvent de locutions, de déictiques ou de connecteurs logiques. Les exemples 4 et 5 montrent deux cas d'éléments facultatifs ou périphériques : le verbe de MODALITÉ *pouvoir* et le marqueur d'EXEMPLIFICATION *en particulier*. Même si ces éléments ne sont obligatoires pour une routine, leur présence permet de découvrir l'intégration des routines de guidage dans le processus d'argumentation ou d'énonciation du discours

⁷⁹ Il s'agit ici d'une sous-classe de la classe des verbes d'ANALYSE_INFO qui comporte des verbes : *affirmer (s')_2, décrire_2, démontrer_2, dessiner_2, éclairer_2, expliciter_2, exprimer_2, illustrer_2, indiquer_2, marquer, montrer_1, refléter_1, révéler_1, signaler_2, spécifier_2, témoigner, véhiculer*. Les numéros *_1, _2* indiquent l'acceptation 1 et l'acceptation 2 des verbes. Comme nous l'avons expliqué dans la partie 4.3.3 concernant le LST, un mot peut avoir plus sens dans les écrits scientifiques.

⁸⁰ Une sous-classe des verbes d'ANALYSE_INFO : *atteindre_3, construire, élaborer, esquisser, établir_1, mettre au point, mettre en forme, mettre en place, organiser_1, produire_1, publier, réaliser_1, reconstituer, reconstruire, reproduire_1, restituer*.

scientifique, dont l'analyse s'avère incontournable pour une étude des routines inscrites à la fois dans l'approche énonciative et fonctionnelle-rhétorique.

[4] Pour une revue de ces questions, **on peut se référer à** Gineste (1997). (Psychologie)

[5] [81] **On se reportera, en particulier, à** David Moore Bergeron, *The wax figures in the Duchess of Malfi*, *Studies in English Literature* (Houston), 1978, p. 331 - 339, et *Monuments and ruins as symbols in The Duchess of Malfi*, *Drama and Symbolism*, *Themes in Drama*, 4, Cambridge University Press, 1982, p. 71 - 87. (Histoire)

Une fois les routines sélectionnées, nous avons procédé à un bilan dans une feuille de calcul afin de noter tous les points à retenir. Neuf colonnes ont été mises en place : Exemples, Discipline, Fonction, Type de phrase, Structuration syntaxique, Variations lexicales, Éléments périphériques, Paramètre énonciatif, Lieu textuel, Notes. Voici un échantillon de nos données ainsi structurées après nettoyage :

1 Description sémantique : l'article se préoccupe d'un thème										
2 Exemples	disciplines	fonctions	types de phrase	structures syntaxiques	réalisations lexicales	éléments périphériques	paramètres énonciatifs	dimensions dialogiques	lieu textuel	notes
15589 Le présent article s' intéresse au coût relatif de la décentralisation des politiques budgétaires dans une Union monétaire par rapport à un mode de gestion " coordonné " (ou centralisé), et sur l' évolution de ce coût en cas d' " élargissement " de l' Union (accroissement du nombre de pays membres) .	économie	annonce le thème, préciser/serrer le thème à aborder	déclarative au présent	Article défini+ présent Nom de document + Verbe de thème (S + V + SN)	le présent article s'intéresse au		cet	interlocutif	introduction	
10449 Alors que la capitale vietnamienne change d' échelle en s' inscrivant dans un processus de métropolisation , l' article s' interroge sur les modalités de gestion de cette transition urbaine , dans les cadres administratifs , législatifs et socio-économiques actuels.	géographie	annonce le thème, préciser/serrer le thème à aborder	déclarative au présent	Article défini+ Nom de document + Verbe de thème (S + V + SN)	l'article s'interroge sur		l'	interlocutif	résumé	
4997 À l' heure où les espaces Est européens s' intègrent dans la mondialisation par les réseaux d' entreprises , l' article s' interroge sur les facteurs principaux qui favorisent leur attractivité et sur l' effet de retour des investissements sur les structures territoriales d' accueil .	géographie	annonce le thème, préciser/serrer le thème à aborder	déclarative au présent	Article défini+ Nom de document + Verbe de thème (S + V + SN)	l'article s'interroge sur		l'	interlocutif	résumé	
15422 Cet article s' intéressera plus										

Figure 4-11 : Structuration des données après nettoyage

Dans ce tableau, la colonne *Exemples* liste toutes les occurrences trouvées pour la routine avec leur contexte. La colonne *disciplines* indique la discipline à laquelle appartient l'occurrence correspondante. La colonne *fonctions* précise les fonctions rhétoriques primaires, ou secondaires. La colonne *types de phrase* mentionne la nature phrastique de la routine actualisée (aussi la temporalité et la modalité). La partie textuelle où se trouve cet exemple (résumé, introduction, corps, conclusion, note) est indiquée dans la colonne *lieu textuel*. La

colonne *structures syntaxiques* montre la structure syntaxique de l'occurrence représentée par des catégories grammaticales et des relations syntaxiques, suivie par la colonne *réalisations lexicales* qui détaille la réalisation de surface. La colonne *éléments périphériques* recense les adverbes et les locutions à fonction discursive ou argumentative qui peuvent participer à l'énonciation de la routine. La colonne *paramètres énonciatifs* indique les éléments énonciatifs comme des déictiques, etc. Les dimensions dialogiques où s'inscrit la routine sont signalées dans la colonne *dimensions dialogiques*. Enfin, la colonne *notes* permet d'ajouter des remarques spécifiques pour chaque occurrence.

Finalement, nous avons obtenu 37 routines de guidage du lecteur. Le nettoyage des données dans l'intention d'identifier les routines de guidage du lecteur avec tous les paramètres retenus dans le fichier Excel offre une facilité de modélisation des routines qui sera l'un des objectifs principaux de la partie III de la thèse. Les routines dont les fonctions primaires ne visent pas à faciliter la lecture ne sont pas retenues, tandis que les routines de guidage validées ont été catégorisées selon leurs fonctions primaires, ce qui permet d'élaborer notre typologie fonctionnelle (nous allons approfondir ce point dans la partie III).

4.6. Étape 3. Élaboration d'un modèle d'analyse pour des routines de guidage

Notre objectif ne s'arrête pas au recueil et à l'identification des routines. Un modèle d'analyse doit maintenant être mis en place pour décrire systématiquement les caractéristiques linguistiques des routines.

4.6.1. Quelques références pour la modélisation des expressions phraséologiques

L'étude linguistique ne se limite jamais seulement à la découverte d'un phénomène linguistique, mais a surtout pour objectif de trouver des descriptions adaptées pour celui-ci. Notre étude sur les routines de guidage vise à élaborer un modèle d'analyse qui permette de schématiser systématiquement leurs caractéristiques linguistiques. Dans cette partie, nous allons d'abord rappeler quelques études phraséologiques envisagées dans plusieurs domaines, comme le traitement automatique des langues ou la didactique du français. Nous allons voir que ces études proposent des modèles d'analyse intéressants pour des expressions phrastiques. En nous basant sur ces modèles, nous proposerons enfin une modélisation systématique spécifiquement adaptée à notre objet d'étude.

Nous allons détailler dans cette section des modèles d'analyse des expressions récurrentes, dont deux dans le champ des études phraséologiques du discours scientifique, et d'autres dans le domaine du traitement automatique des langues.

Dans le domaine de la phraséologie du discours scientifique, nous proposons d'examiner les modèles d'analyses dans l'étude sur les routines sémantico-rhétoriques des articles de recherche de Chevalier (2019), ainsi que dans l'étude classique sur la phraséologie bilingue français-anglais de Pecman (2004). Les autres modèles ont été développés en linguistique de corpus avec un recours aux outils textométriques afin d'effectuer un « moissonnage automatisé » (Bendinelli, 2017, p. 6) des expressions récurrentes à partir des corpus. Nous relevons dans ce domaine des propositions de modélisation de ces phrases routinisées associées à des fonctions rhétoriques ou discursives. Ces modélisations visent initialement à repérer automatiquement dans les corpus les propositions métadiscursives ayant une stratégie rhétorique (Sandor, 2007) ou les propositions véhiculant des informations stratégiques (Ibekwe-Sanjuan, Chen, & Robert, 2008). Elles proposent, pour nous, plusieurs pistes intéressantes pour modéliser des séquences complexes comme les routines de guidage du lecteur.

En fonction du niveau d'abstraction de ces propositions pour la modélisation, nous les divisons en trois types : le premier type de modélisations met en place une description relativement simple en listant les variations lexicales et syntaxiques des phrases. Le deuxième type propose une schématisation plus abstraite avec des étiquettes notionnelles ou cognitives pour un groupe d'éléments phraséologiques. Le troisième type essaie non seulement d'attribuer les étiquettes notionnelles aux phrases, mais aussi de prendre en compte les caractéristiques syntaxiques des expressions, en attribuant des étiquettes sémantiques.

4.6.1.1. Modèles simples

En ce qui concerne le premier type qui est relativement simple, nous citons deux études : l'étude de Chevalier (2019) sur les routines sémantico-rhétoriques dans les écrits scientifiques en français et celle d'Ibekwe-Sanjuan et al. (2008) sur les expressions lexico-syntaxiques exprimant la contribution de l'auteur dans les articles scientifiques en anglais.

4.6.1.1.1. Modèle de Chevalier (2019)

Chevalier (2019) propose une analyse des routines sémantico-rhétoriques renvoyant à la description des démarches et procédures scientifiques dans les écrits scientifiques. Comme nous, elle a repris la définition de Tutin (2016) pour ces routines sémantico-rhétoriques. Elle a recensé des routines d'intention de communication des auteurs dans un corpus de 40 articles de

recherche, dans des revues québécoises et canadiennes pour la période 2015 – 2017. Les intentions de communications inventoriées comportent des routines ayant des fonctions comme ‘donner un aperçu de la problématique, de la méthodologie, de la discussion’, ‘rappeler le contenu de l’article’, etc. (cf. Chapitre 3). Toutes les routines identifiées ont également été associées aux parties textuelles, à savoir « introduction, problématique, cadre conceptuel, méthodologie, résultats, discussion des résultats et conclusion » (*ibid.*). En s’appuyant sur trois fonctions particulières ‘présenter les objectifs’, ‘établir les constats, la recommandation’, cette étude essaie principalement de mener une analyse des routines sur le plan linguistique en mettant l’accent sur leurs constituants lexicaux et leurs variations syntaxiques. Prenons l’exemple des routines qui présentent les objectifs. Trois routines reprises ici dans le tableau 4-7 ont été recensées :

Le présent travail Cet article Notre recherche L’étude Notre contribution	a pour but de a pour objectif de a comme objectif de chercher à tente de vise à	<i>déterminer</i> <i>documenter</i> <i>évaluer</i> <i>examiner</i> <i>explorer</i> <i>identifier</i> <i>savoir si</i> ...	
Nous	cherchons à désirons tentons de voulons nous proposons de		
Le but de L’objectif de Les objectifs de	le présent travail cet article notre recherche l’étude	est de	<i>décrire...</i> <i>cerner...</i> <i>mesurer...</i> <i>comprendre...</i> <i>mettre en lumière...</i> <i>expliquer comment...</i>

Tableau 4-7 : Routines de présentation des objectifs (Chevalier, 2019)

Outre le recensement des variations des routines dans les tableaux, le modèle développe également une description portant sur quatre paramètres que nous pouvons résumer ici ainsi :

1. Distinction des parties syntaxiques récurrentes et non récurrentes ;
2. Identification des paradigmes sémantiques pour les parties récurrentes ;
3. Alternances syntaxiques des routines ;
4. Parties textuelles auxquelles appartiennent les routines identifiées.

D’abord, l’observation des parties syntaxiques des trois routines montre qu’elles partagent les mêmes parties syntaxiques récurrentes : le sujet et le prédicat. Pour les deux premières routines : les sujets peuvent être variés, comportant donc deux paradigmes sémantiques : le lexique des objets scientifiques comme *recherche/article/étude*, celui de

l’auteur de la recherche (*nous*), alors que leurs prédicats font tous partie d’un même paradigme que l’on peut gloser par ‘viser à’. Ensuite, Chevalier (2019) indique que les compléments en italique dans le tableau ne sont pas récurrents, mais thématiques, et peuvent être choisis par l’auteur en fonction du contenu thématique du texte scientifique. La troisième routine est une alternance syntaxique des deux premières routines avec les mêmes éléments sémantiques en jeu. Le prédicat ‘viser à’ est réalisé par la construction attributive (*l’objectif de cette recherche est de...*) (*ibid.*). Finalement, ces routines présentant les objectifs d’étude se répartissent dans trois sections textuelles : introduction, problématique et description du cadre conceptuel.

Cette description récapitulant quatre paramètres montre les points essentiels à retenir pour la description linguistique des routines que nous reprendrons en compte dans notre analyse des routines de guidage du lecteur, en particulier les variations linguistiques.

4.6.1.1.2. *Modèle d’Ibekwe-Sanjuan et al., (2008)*

La deuxième étude que nous citons ici est semblable à celle de Chevalier (2019), malgré l’objectif TAL qui vise à identifier automatiquement des phrases présentant des informations stratégiques dans la partie introductive des écrits scientifiques. Sept types d’informations stratégiques sont visés dans cette étude : ‘objectif’, ‘résultat’, ‘nouveau’, ‘hypothèse’, ‘recherche associée’, ‘conclusion’ et ‘perspective’. Pour récupérer ces patrons, l’étude propose des grammaires locales en énumérant les variations d’un patron après une étude manuelle sur un corpus plus petit (Ibekwe-Sanjuan *et al.*, 2008, p. 1521) :

Types des informations véhiculées	Patrons lexico-syntaxiques repérés
Objectif	In this <i>{article paper study research work}</i> ... We <i>{examine investigate describe present outline introduce consider}</i> ... DET <i>{motivation: aim goal objective}</i> ...
Nouveauté	Here, we propose a novel (...) approach... This analysis reveals... Emerging evidence suggests that... Interestingly, our results indicate that...
Hypothèse	DET_NP <i>{may might}</i> <i>{ADV V_NP}</i> ... Our findings support the view that... DET NP <i>can</i> <i>{V NP}</i> ...
Perspective	<i>{Further Future more}</i> <i>{work investigation observation}</i> <i><verb></i> ...

Tableau 4-8 : Exemples de patrons lexico-syntaxiques dans l’introduction des écrits scientifiques en anglais

Comme celui de Chevalier (2019), le recensement des variations des expressions dans ce modèle touche le lexique ainsi que ses variations, et la structure syntaxique. De plus, il introduit les annotations des catégories syntaxiques d’un niveau plus abstrait comme DET, VERB, ADV. Dans ce modèle, l’auteur observe également l’aspect morphologique en ce qui concerne le genre, le nombre, la flexion, etc. Nous trouvons que les propriétés morphologiques sont nécessaires dans l’analyse des expressions assez longues et complexes, parce que les

changements temporels ou modaux peuvent largement influencer le sens d'une expression de même structure syntaxique. Dans notre corpus, la routine <(comme) + Pronom AUTEUR + avoir/venir de/aller + Verbe d'ANALYSE_INFO (constater/voir/dire/mentionner...)> peut avoir deux fonctions : la fonction de navigation intertextuelle vers la partie précédente quand elle est fléchie au passé composé avec l'auxiliaire *avoir*, au présent avec *venir de* (*nous avons vu, on vient de dire, comme on a constaté*), et la fonction d'annoncer le thème bientôt abordé quand la routine est une phrase au présent avec *aller* ou au futur avec des verbes (*nous allons voir..., nous montrerons que...*). Nous retenons dans ce modèle d'analyse deux points de plus par rapport au modèle de Chevalier : les propriétés morphologiques et les catégories syntaxiques plus abstraites, que nous allons introduire dans notre propre modèle d'analyse des routines de guidage du lecteur.

Mais ces études n'ont pas pour finalité d'établir un modèle linguistique permettant de décrire des caractéristiques des expressions routinisées. Le travail de Chevalier (2019) s'inscrit dans le domaine de la didactique qui vise à satisfaire à la didactisation des ressources linguistiques, et celui d'Ibekwe-Sanjuan *et al.* (2008) dans le champ du traitement automatique des langues. Ces modèles, malgré leur faible complexité, s'adaptent bien à leurs objectifs de recherche. Néanmoins, l'objectif de notre étude, nous le rappelons, est de mettre en place une modélisation linguistique plus schématisée qui ne recense pas uniquement des variations linguistiques, mais qui examine aussi les autres aspects des routines, comme par exemple la dimension dialogique. En ce qui concerne des parties récurrentes dans ces études, il est tout à fait possible d'attribuer des étiquettes sémantiques au lexique partageant le même paradigme sémantique. Par exemple, les verbes en tant que compléments du prédicat dans les routines sémantico-rhétoriques citées plus haut sont des verbes d'analyse des informations scientifiques très récurrents dans le registre de l'écrit scientifique, et classés comme tels dans le LST. Leur étiquetage avec des traits sémantiques permet de mettre en évidence les actes scientifiques les plus représentatifs dans le domaine scientifique.

4.6.1.2. *Modèle à schéma collocationnel de Pecman (2004)*

Le deuxième modèle d'analyse est élaboré dans le cadre de la thèse de Pecman (2004), pour les unités phraséologiques dans les écrits scientifiques. À travers un corpus parallèle en français et en anglais dans trois domaines tels que la chimie, la physique et la biologie, l'étude de Pecman a pour objectif de trouver le moyen de faciliter la rédaction des textes scientifiques par l'analyse des phénomènes collocatifs dans le discours scientifique.

Cette étude propose une approche onomasiologique en mettant en place une ontologie de la *Langue Scientifique Générale*⁸¹. Il s'agit plus précisément d'un système d'étiquetage à la fois notionnel et rhétorique qui permet de modéliser 2000 unités phraséologiques sous forme de schémas collocationnels. L'accès onomasiologique permet de structurer des unités phraséologiques à partir d'un concept. Par exemple, le concept |EXPECTATION| pour exprimer une attente regroupe des unités phraséologiques comme *Nous nous attendons à ce que/we expect that, comme prévu/as expected*. Les étiquettes notionnelles servent à grouper des concepts sémantiquement proches autour de quatre grandes sphères qui « agissent comme des noyaux sémantiques fondamentaux » (Pecman, 2007, p. 89) de la *Langue Scientifique Générale* : la scientificité, l'universalité, la modalité et la discursivité, qui regroupent plus précisément 125 « concepts spécifiques ou hyponymiques » au total (Pecman, 2004, p. 296).

La scientificité et l'universalité occupent une place relativement importante dans les écrits scientifiques, manifestant respectivement la nature scientifique des écrits scientifiques et l'emploi des notions universelles de la communauté linguistique propre au discours scientifique. D'une part, les écrits scientifiques emploient largement des notions liées à la procédure scientifique, aux activités scientifiques, aux entités, aux descriptions abstraites, à la caractérisation de la qualité, ou aux concepts dont la valeur devient spécifique dans le domaine scientifique (Pecman, 2007). Nous trouvons par exemple dans la sphère de la scientificité des notions comme |EXPERIMENT| ou |OBSERVATION| qui sont des concepts indispensables dans les écrits scientifiques. D'autre part, certains concepts plus larges que la scientificité sont également importants pour la description scientifique, comme ceux de |TEMPORALITY|, |QUANTITY|, |CONSEQUENCE| et relèvent de la sphère de l'universalité.

En parallèle, les sphères de la modalité et de la discursivité ne sont pas moins importantes. La première vise à montrer l'intégration de l'auteur en tant qu'énonciateur de son propre discours, la seconde permet d'organiser le discours. Plus précisément, la sphère de la modalité est davantage liée au sujet parlant des écrits scientifiques et regroupe des unités phraséologiques permettant à l'auteur d'exprimer un point de vue, une évaluation ou une appréciation de son propre discours, par exemple, la modalité épistémique |PROBABILITY| ou déontique |NECESSITY|. La sphère de la discursivité regroupe toutes les unités

⁸¹ La *Langue Scientifique Générale* (LSG) est « une pratique langagière spécifique à une communauté de discours composée de chercheurs en sciences exactes et sciences humaines dont les objectifs communicatifs émanent de préoccupations partagées par des scientifiques à travers le monde et indépendamment de leurs spécificités disciplinaires » (Pecman, 2007, p. 85).

phraséologiques impératives pour la construction du discours scientifique comme la présentation du sujet d'étude |PRESENTATION|, l'introduction d'une citation |CITATION|, etc. (Pecman, 2007). Nous avons remarqué cependant que la classe |CAUSE| est incluse dans la sphère discursive, qui renvoie plutôt à l'argumentation de la scientificité au lieu de la discursivité, tandis que la classe |CONSEQUENCE| est bien classée dans la scientificité.

Pecman (2004) n'a pas seulement créé ces étiquettes notionnelles. Elle a également proposé des schémas collocationnels pour analyser les unités phraséologiques en montrant leurs propriétés lexico-syntaxiques et leurs extensions. Voici un exemple de citation :

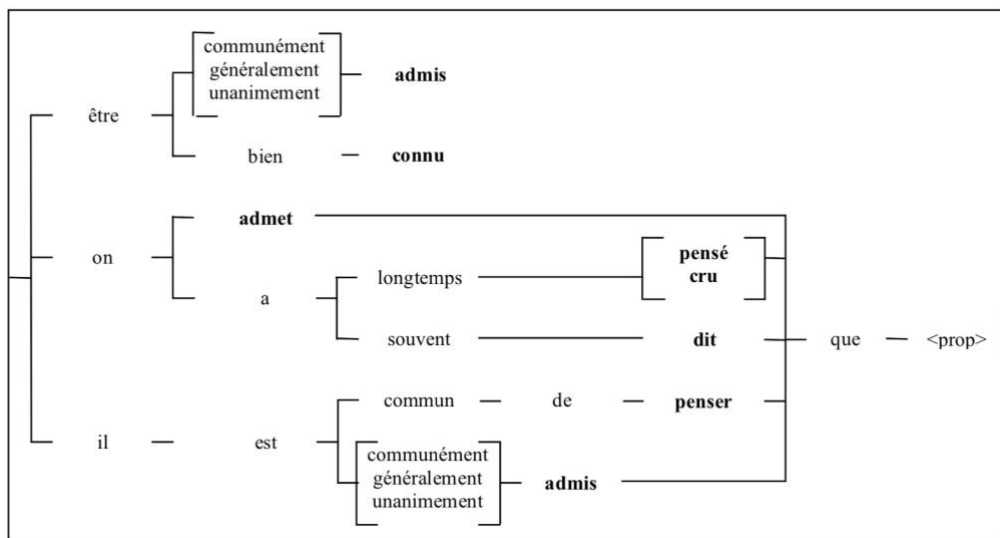


Figure 4-12 : Schéma collocationnel de la citation en français (Pecman, 2004, p. 414)

Nous pouvons voir, dans cet exemple, que le concept de citation en français comporte trois formes sur l'axe syntagmatique : forme passive, forme canonique, et forme impersonnelle. Sur l'axe paradigmatique, différentes variations lexicales liées à l'idée de citation ont été intégrées. Si le verbe *admettre* participe à toutes les unités phraséologiques, les verbes *connaître*, *penser*, *croire*, *dire* présentent une attraction plus forte pour une ou deux structures syntaxiques. De plus, les schémas collocationnels prennent en compte également des éléments facultatifs qui constituent une extension des unités phraséologiques. Ces éléments peuvent être des propositions subordonnées, des arguments nominaux ou verbaux. L'exemple de la citation a une extension propositionnelle introduite par le pronom relatif *que*. Dans d'autres exemples, comme l'une des unités phraséologiques de la nécessité *il faut <faire qch>/que <prop_sub>/que <qn fasse qch>*, les extensions sont respectivement, un syntagme verbal *faire qch*, une proposition subordonnée introduite par *que*.

Il faut souligner que l'avantage de ce modèle d'analyse à l'aide des schémas collocationnels est évident : les variations lexico-syntaxiques, ainsi que les extensions d'unités phraséologiques sont bien structurées. Néanmoins, à l'instar de Tutin & Ji (2022), nous avons relevé deux aspects qui pourraient améliorer encore ce schéma collocationnel. Premièrement, les composants des éléments phraséologiques peuvent être décrits d'une façon plus abstraite en adoptant également des catégories sémantiques. Par exemple, *longtemps/souvent* peuvent être regroupés sous l'étiquette sémantique en l'indiquant les catégories grammaticales adverbales de DURÉE/FRÉQUENCE et *communément/généralement/unanimement* comme adverbales de GÉNÉRALITÉ. Le regroupement sémantique des composants permet de fournir aux apprenants des variantes lexicales des unités phraséologiques lors de la rédaction des écrits scientifiques. Deuxièmement, les schémas collocationnels ne différencient pas les unités scientifiques de différentes natures, c'est-à-dire que les marqueurs discursifs, les collocations, les expressions propositionnelles ont été analysés par le même modèle sans un traitement différencié. Cependant, ces unités scientifiques manifestent des caractéristiques différentes sur les plans discursifs et rhétoriques.

Malgré ces deux remarques, nous considérons toujours que le travail de Pecman (2004), notamment les schémas collocationnels et le traitement par des concepts notionnels, offrent une entrée intéressante pour traiter des unités scientifiques avec des étiquettes notionnelles qui sont à la fois sémantiques et rhétoriques. Cette approche a influencé l'élaboration de notre propre modèle d'analyse qui va être développé à la fin de ce chapitre.

4.6.1.3. *Modèle de « concept-matching » de Sandor (2007)*

Le dernier type donne une entrée plus abstraite pour réfléchir sur la modélisation des routines de guidage du lecteur. Sandor (2007) propose une modélisation plus fine en créant un modèle de « concept-matching ». Ce dernier associe les marqueurs lexicaux à des fonctions discursives (Sandor, 2007, p. 103), en exploitant des concepts fonctionnels pour des parties syntaxiques des propositions métadiscursives. Avec cette méthode, la phrase *Recently, several authors have suggested*, qui correspond à la fonction 'background knowledge' est modélisée ainsi :

background		knowledge		
general	past	researcher	mental activity	object/result of research activity
Most		evolutionists	consider	
	Recent		get a picture	findings
several	[inherited from preceding sentences] Recently	authors	suggest have suggested	data [referring to « Recents findings »]
	previous		expand	hypothesis
	[inherited from preceding sentences]		explains	Hypothesis [referring to « previous hypothesis »]
	[inherited from preceding sentences]		explains	It [referring to « previous hypothesis »/observations]

Tableau 4-9 : Expressions de la fonction 'background knowledge' reprises de Sandor (2007, p. 103–104).

Ce type de modélisation garantit donc la cohérence entre les notions sémantiques et la fonction discursive liée au contexte scientifique. Il reprend, d'une part, certains avantages des modèles précédemment évoqués : des étiquettes notionnelles aux expressions (*background knowledge*), le recensement des variations linguistiques ; d'autre part, il attribue des étiquettes sémantiques aux parties syntaxiques (<general>, <researcher>). Ces étiquettes sémantiques indiquent les champs sémantiques récurrents dans les expressions préfabriquées, donnent une entrée plus approfondie pour la description sémantique des expressions métadiscursives. Nous remarquons cependant un inconvénient dans ce modèle d'analyse : comme l'auteur étiquette toutes les parties syntaxiques des expressions véhiculant le même sens, mais avec des structures syntaxiques différentes, l'affichage des parties syntaxiques est parfois inversé : *previous hypothesis expand* \Rightarrow *previous expand hypothesis* ou *it explains* \Rightarrow *explains it*. L'inversion de l'ordre pour des parties syntaxiques n'a peut-être pas d'incidence pour une étude de traitement automatique des langues, mais elle risque de créer des confusions dans une étude linguistique ayant pour objectif de décrire les caractéristiques des expressions complexes comme les nôtres. C'est pourquoi dans notre modèle d'analyse, nous allons le voir dans la partie suivante, nous avons essayé d'éviter cet inconvénient en divisant des routines syntaxiquement différentes en plusieurs routines.

En résumé, la modélisation des séquences préfabriquées que nous souhaitons élaborer n'est pas une description simple des caractéristiques lexicales, syntaxiques, sémantiques et fonctionnelles, mais plutôt un système plus abstrait établi à partir de l'observation du contexte et du fonctionnement cognitif, en prenant en compte les caractéristiques du genre discursif.

4.6.2. Élaboration d'un modèle d'analyse pour les routines de guidage du lecteur

Notre modèle d'analyse pour décrire des routines de guidage du lecteur s'appuie sur un modèle déjà développé dans Tutin & Ji (2022) qui a profité des modélisations décrites plus

haut : étiquette conceptuelle, étiquette sémantique, recensement des variations lexicales et syntaxiques. Il s’agit d’un modèle spécialement proposé pour une description approfondie des routines sémantico-rhétoriques.

4.6.2.1. Modèle de base

La modélisation de Tutin & Ji prend racine dans l’observation des paramètres définitoires des routines sémantico-rhétoriques, dont les routines de guidage font partie. Ce modèle prend en compte d’autres paramètres entrant en jeu, et introduit 9 champs qui sont illustrés par la fiche analytique⁸² de la routine <voir/cf. ci-dessous/supra/plus bas... ; ou voir/cf. section/chapitre/page... > (Tutin & Ji, 2022, p. 59–60). Les 9 champs à prendre en considération dans le modèle couvrent tous les aspects linguistiques d’une routine : *exemple*, *type de phrase*, *description sémantique*, *réalisations lexicales*, *structure syntaxique*, *éléments périphériques*, *paramètres énonciatifs*, *structure textuelle* et *fonctions rhétoriques*, c’est-à-dire les éléments que nous avons déjà retenus lors de l’identification des routines de guidage du lecteur (cf. section 4.3.4).

Exemple	<i>voir ci-dessous</i> <i>cf. section 4</i>
Type de phrase	Phrase injonctive.
Description sémantique	L’auteur enjoint le lecteur à « se reporter » au « lieu textuel » pour une information précise.
Réalisations lexicales	1. (<i>voir/cf.</i>) (<i>ci-dessous/supra/plus bas...</i>) 2. (<i>voir/cf.</i>) (<i>section/chapitre/paragraphe/page...</i>) (1/B//)
Structure syntaxique	<p>1. Verbe Adverbe</p> <p>2. Verbe Nom Numéro</p>
Éléments périphériques	Pas d’éléments périphériques.
Paramètres énonciatifs	La phrase s’adresse directement au lecteur (par le biais de l’infinitif jussif) et relève du dialogisme interlocutif.
Structure textuelle	La routine apparaît toujours dans le péritexte, entre parenthèses ou en note de bas de page. Elle n’est pas spécifique d’une partie textuelle, mais apparaît peu dans l’introduction ou la conclusion.
Fonctions rhétoriques	La fonction principale est le guidage du lecteur dans la structure du document (structuration/navigation/renvoi intratextuel). Une fonction secondaire peut être d’appuyer l’argumentation de l’auteur en renvoyant le lecteur vers des éléments de preuve factuels.

Figure 4-13 : Fiche analytique d’une routine (Tutin & Ji, 2022, p. 59–60)

⁸² Dans la fiche, la structure syntaxique adopte une représentation graphique proposée par l’outil Arborator. Voir le site <https://arborator.ilpga.fr/q.cgi>.

1. Exemple

Dans cette modélisation, l'exemple montre les occurrences les plus courantes d'une routine.

2. Type de phrase

Ce champ indique la nature syntaxique de la routine, comme phrase déclarative, phrase subordonnée, phrase injonctive etc. (*cf.* section 1.1.6.4).

3. Description sémantique

La description sémantique est la plus intéressante pour codifier les routines. Ce modèle propose pour chaque routine une description sémantique permettant de mieux expliquer le sens et la fonction de la routine. Il ne s'agit pas seulement d'une étiquette comme dans le modèle de Pecman (2004), mais plutôt d'une glose décrivant à la fois la fonction et la structure sémantique d'une routine. Plus précisément, cette description regroupe les parties syntaxiques essentielles d'une routine sous des étiquettes sémantiques, notamment les parties fixes d'une routine. À titre d'exemple, dans les écrits scientifiques, quand l'auteur manifeste le besoin de signaler à son lecteur les informations dans un lieu textuel précis (nous parlons de navigation intratextuelle, *cf.*, Chapitre 5), il peut utiliser la routine (*voir/cf. ci-dessous/supra/plus bas... ; ou voir/cf. section/chapitre/page...*). Une description sémantique modélisée pour cette routine s'exprime ainsi : <L'auteur enjoint le lecteur à « Se reporter » au « Lieu textuel » pour une information précise> qui fonctionne comme une glose de la fonction de cette routine. Dans cette glose, « Se reporter » et « Lieu textuel » sont deux étiquettes sémantiques attribuées par Tutin & Ji (2022, p. 59), qui renvoient respectivement aux deux parties syntaxiques fixes de la routine : [*voir/cf.*] et [*ci-dessous/supra/plus bas/section/chapitre/page*]. Ces deux cadres représentent en même temps des paradigmes lexicaux : des verbes de RÉFÉRENCE, et des adverbes d'ESPACE_NAVIGATION TEXTUELLE (*ci-dessous, ci-dessus, de là, ibid., infra, plus avant, précédemment_2*) ou des noms de SECTION (*annexe, bibliographie, chapitre, conclusion_1, développement_1, extrait, introduction_1, note_2, section_1, page*).

La description sémantique des routines donne un accès plus concret pour comprendre leurs fonctions, pas simplement via les noms des fonctions que nous allons leur attribuer dans les classements fonctionnels, mais aussi à travers une compréhension plutôt explicite des actes scientifiques qui baptisent les fonctions rhétoriques. Cognitivement, les étiquettes conceptuelles dans la description sémantique permettent, d'une part, d'expliquer l'acte de langage spécifique au discours scientifique, d'autre part, de créer une association entre les

fonctions et les classes sémantiques. En ce sens, la glose équipant des étiquettes sémantiques < L'auteur enjoint le lecteur à « Se reporter » au « Lieu textuel » pour une information > met en évidence l'acte scientifique qui facilite le repérage d'une information dans les parties textuelles pour le lecteur. En résumé, ce type de description conjugue à la fois la compréhension cognitive des actes scientifiques et l'association de celle-ci avec des champs sémantiques récurrents et représentatifs dans les routines.

4. Réalisations lexicales

Les réalisations lexicales visent à montrer les paradigmes lexicaux réellement réalisés pour la routine. Ce champ recense des variations lexicales les plus courantes du corpus. Force est de noter que les variations lexicales présentées dans les modélisations des routines ne sont pas exhaustives.

5. Structure syntaxique

En ce qui concerne la structure syntaxique, ce modèle propose une analyse en syntaxe de dépendance⁸³. Nous proposerons une description plus simple dans la section suivante.

6. Éléments périphériques

Les adverbes ou les locutions sont aussi recensés dans cette modélisation comme éléments périphériques s'il en existe. Ces éléments, bien que facultatifs, semblent très récurrents dans certaines routines, et permettent d'interpréter l'intégration des fonctions rhétoriques des routines dans le scénario d'argumentation du discours scientifique.

7. Paramètres énonciatifs

Les paramètres énonciatifs constituent un facteur très important pour étudier les routines, notamment sur la dimension dialogique. Il convient de rappeler que la définition que nous avons donnée des routines de guidages, et plus largement des routines sémantico-rhétoriques, met en évidence leur potentiel d'actualisation énonciative dans le discours. De ce fait, les paramètres énonciatifs constituent un angle d'observation utile pour comprendre l'actualisation énonciative des routines.

8. Structure textuelle

⁸³ Cette étude exploite un logiciel en ligne – Arborator (<https://arborator.ilpga.fr/>) pour une représentation graphique qui permet d'indiquer à la fois la séquence des éléments et les relations de dépendance entre ceux-ci.

Ce champ s'intéresse à la position textuelle de la routine. La position textuelle vise à préciser les emplacements des routines dans l'article de recherche. Comme l'indique la figure 5-4, cette routine de renvoi textuel se trouve souvent dans le périphrase, entre parenthèses ou en note de bas de page. La précision de la position des routines dans le texte présente un intérêt pour comprendre la pratique de l'écriture scientifique. Les routines ne se trouvent pas dans n'importe quelle section de l'article, comme les routines qui annoncent l'objectif d'étude et qui se trouvent souvent dans l'introduction. Le lieu textuel des routines constitue donc un élément fondamental lors de l'analyse (*cf.* Chapitre 8).

9. Fonctions rhétoriques

Le champ *fonction* fournit non seulement la fonction principale des routines, mais aussi leurs fonctions secondaires, car il est possible qu'une routine puisse remplir plusieurs fonctions. Comme notre typologie (*cf.* Partie III) est élaborée en se basant sur les fonctions primaires des routines, il nous semble nécessaire de décider également les fonctions secondaires des routines, afin d'en retenir les nuances.

4.6.2.2. Notre modèle d'analyse après adaptation

Dans notre étude, nous reprenons la modélisation présentée *supra* comme modèle d'analyse de base, mais nous l'avons adapté à notre étude pour plus de simplicité et de clarté. Nous gardons dans le texte principal (chapitres 5, 6 et 7) ce tableau de modélisation simplifié pour chaque routine en reprenant les 9 champs, et en effectuant quelques améliorations.

- a. changement de noms et de portée pour le champ *structure textuelle* ;
- b. changement de forme de représentation des champs *description sémantique*, *structure syntaxique*, *réalisations lexicales* ;
- c. remise en ordre de ces 9 champs.

Ces changements ne touchent pas les sphères d'analyse des caractéristiques des routines, ils visent seulement à harmoniser et optimiser la représentation de l'analyse.

a. Changement de noms et de portée pour le champ *structure textuelle*

Par rapport au nom *structure textuelle*, nous préférons l'appeler *position textuelle* afin de préciser explicitement qu'il s'agit du lieu textuel où se trouve les routines. C'est aussi pour éviter l'ambiguïté éventuellement causé par le terme *structure textuelle* que nous avons employé pour désigner un groupe de routines aux fonctions d'organisation textuelle et discursive des articles de recherche.

Comme on l'a observé dans la section 2.2.2, les structures des articles de recherches comportent des parties générales comme le titre, le(s) nom(s) de(s) l'auteur(s), le résumé avec les mots-clés, le corps principal séparé en sections selon les sous-thèmes abordés, les notes de bas de page et la bibliographie. En fait, toutes ces métadonnées ne sont pas indiquées spécifiquement dans le corpus TermITH (contrairement au Scienquest). D'ailleurs, les sous-sections du corps principal des articles de recherches en SHS ne suivent pas toutes la structure méthode-résultat-discussion de Swales (1990). Les résultats de l'extraction ne permettent donc pas de fournir la position exacte d'une routine. Mais nous pouvons identifier grossièrement les contextes gauches et droites plus larges d'une occurrence pour identifier les parties principales : résumé, introduction, corps principal (entre l'introduction et la conclusion), conclusion et note. Les positions textuelles ainsi retenues montrent plutôt une tendance qui ne doit pas être comprise comme une position absolue. De surcroît, nous avons observé les dispositifs typographiques dans ce champ s'il y en a, comme les parenthèses, les guillemets qui sont souvent normalisés dans les articles de recherche. Nous proposerons une observation globale de ce point dans le dernier chapitre.

b. Changement de forme de représentation des champs *description sémantique, structure syntaxique, réalisations lexicales*

Concernant le champ *description sémantique*, nous avons décidé de mettre la glose dans les chevrons <>. Les parties fixes dans la description sémantique sont centrales et reçoivent donc respectivement une étiquette conceptuelle liée soit aux participants du discours scientifique (cf. section 1.2.6.2) comme l'auteur, le lecteur, ou le discours en cours, soit à l'acte scientifique ou discursif comme analyser des données, définir une notion. Ces étiquettes ont été mises en valeur par la mise en gras et encadrées par les guillemets-apostrophes ‘’. Les éléments facultatifs et périphériques n'ont pas connu un traitement de mise en valeur, mais sont entourés par les parenthèses () s'il y en a. Par exemple, une routine de planification peut être décrite comme < **L'auteur** **traite** un sous-thème dans une **partie textuelle** **numérotée** > (*dans la troisième partie, nous exposons...*). **Auteur** signifie l'auteur du texte en cours, **traite** renvoie aux actes scientifiques ou communication pour parler d'une problématique ou d'un thème spécifique. Les deux étiquettes **partie textuelle** **numérotée** désignent respectivement la section ou la partie du texte et l'ordre dans le texte attribué à la partie textuelle. Toutes les étiquettes conceptuelles utilisées dans notre étude ont été rassemblées dans l'annexe 2.

Concernant le champ *structure syntaxique*, nous n'employons pas la représentation en syntaxe de dépendance qui complexifie la compréhension du lecteur. Nous avons adopté une

schématisation que nous appelons « formules syntaxiques ». L'idée est d'inclure dans les formules les classes sémantiques du LST pour construire la présentation de la structure syntaxique des routines à la façon de Chevalier (2019). Les classes sémantiques correspondent souvent aux étiquettes conceptuelles définies dans les descriptions sémantiques. Reprenons la routine d'annonce de l'objectif d'étude citée plus haut comme exemple. Sa configuration syntaxique est formulée ainsi :

$$[\text{pronom AUTEUR}] + \left\{ \begin{array}{l} [\text{verbe d'ANALYSE_INFO}] \\ [\text{verbe de COMMUNICATION}] \end{array} \right\} + \left\{ \begin{array}{l} \text{SN} \\ \text{que} \end{array} \right\} + \left\{ \begin{array}{l} \text{dans la} \\ \text{en} \end{array} \right\} + [\text{nom de SECTION}] + \left\{ \begin{array}{l} [\text{chiffre}] \\ [\text{adjectif d'ÉNUMÉRATION}] \end{array} \right\}$$

Dans la formule, les éléments fixes de la routine sont entourés par [] et mis en gras. Quand une position syntaxique peut être occupée par plusieurs paradigmes sémantiques, cette position regroupe les classes sémantiques par les grands crochets {}. S'il existe des éléments facultatifs, ils seront introduits dans les parenthèses (). La mise en valeur de ces éléments fixes crée une correspondance entre les étiquettes conceptuelles et les classes sémantiques du LST retenues. On peut voir ainsi entre la description sémantique et la formule syntaxique de cette routine de planification les relations suivantes :

Correspondance entre les étiquettes conceptuelles et les classes sémantiques du LST	
'auteur'	[pronom AUTEUR]
'traite'	[verbe d'ANALYSE_INFO] ou [verbe de COMMUNICATION]
'partie textuelle'	[nom de SECTION]
'numérotée'	[chiffre] ou [adjectif d'ÉNUMÉRATION]

Tableau 4-10 : Correspondance entre la description sémantique et la construction syntaxique de la routine de planification

Nous rencontrons une difficulté très fréquente en ce qui concerne la représentation des variations lexicales par les classes sémantiques du LST, dans la modélisation des routines. D'une part, les mots classés dans le LST ont souvent plusieurs acceptions, et peuvent ainsi appartenir à plusieurs classes sémantiques, comme nous l'avons vu dans la section 4.3.3. Dans ce cas, il faut faire le choix de la classe la plus pertinente. D'autre part, la variation lexicale des routines dans le corpus ne correspond pas toujours parfaitement à l'intégralité d'une classe sémantique. Elle peut couvrir partiellement plusieurs classes sémantiques du LST, sans oublier que les grandes classes possèdent des sous-classes. Par conséquent, utiliser les sous-classes ou la classe générale devient un choix qui doit se baser sur l'observation de l'ensemble des réalisations lexicales trouvées pour une routine dans le corpus.

Pour des mots polysémiques (classés dans différentes classes), nous proposons de trancher sur la classe dominante de ce type de mot lors de l'élaboration de la formule syntaxique d'une routine.

Pour des routines dont les variations lexicales appartiennent à plusieurs sous-classes ou classes sémantiques, nous privilégions les plus fréquentes, et élargissons parfois la représentation des sous-classes par les classes plus générales. Nous employons encore l'exemple de la routine de planification afin de justifier nos choix. Nous retenons ensemble 27 verbes qui entrent dans cette routine programmant le scénario discursif d'un texte scientifique, dont la majorité (18 verbes) appartient à la classe générale d'ANALYSE_INFO (14 sous-classes au total) et se répartit dans ses 11 sous-classes. 5 verbes sont classés dans 3 sous-classes de la classe de COMMUNICATION (6 sous-classes au total). 4 verbes dans d'autres classes. En voici les détails, dans la figure 4-14 :

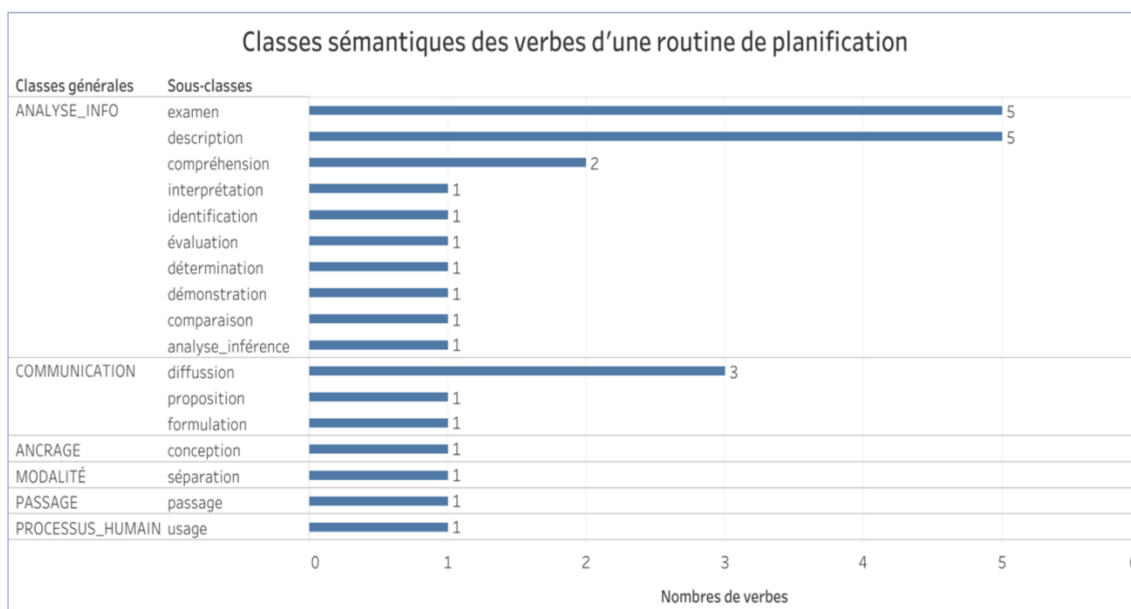


Figure 4-14 : Classes sémantiques des verbes dans une routine de planification

En tenant compte de la portée et la fréquence des verbes montrées par la figure, nous avons finalement choisi d'indiquer les deux classes générales les plus représentatives dans l'équation syntaxique : verbes d'ANALYSE_INFO et verbes de COMMUNICATION.

Par contre, pour des variations assez homogènes, nous préférons utiliser les sous-classes sémantiques, en visant à une description sémantique plus précise. C'est le cas des classes comme [pronom AUTEUR], [nom de SECTION], [adjectif d'ÉNUMÉRATION], [chiffre].

Les (sous-)classes sémantiques ont une valeur indicative pour comprendre la préférence sémantique dans une routine d'une fonction de guidage précise. Cela nous permet de mener des

observations plus synthétiques sur les caractéristiques linguistiques des routines de guidage dans la troisième partie de cette thèse. Leur emploi présente également un intérêt dans la perspective didactique de ces routines. Par exemple, dans notre corpus d'articles de recherche, les occurrences retenues du [nom de SECTION] sont *partie*, *section*, *conclusion* et *annexe*. Mais cette sous-classe comporte aussi d'autres noms comme *chapitre*, *conclusion*, *bibliographie* qui peuvent s'intégrer dans la routine de planification dans d'autres sous-genres d'écrits scientifiques comme la thèse ou le mémoire de master. Une telle extension en didactique sera très utile aux apprentis scripteurs des écrits scientifiques.

Pour la description syntaxique, nous préférons utiliser un tableau structuré, plutôt qu'une structure linéaire. À la structure linéaire suivante :

[*nous/on/je*] [*présenter/analyser/examiner/étudier/montre/se focaliser/conclure*]
 ([*dans/en*] [*premier/deuxième...*] [*section/partie*] [1,2,3/I,II,III...])

correspondra la représentation ci-dessous :

Structure syntaxique	[pronom AUTEUR]+ { [verbe d'ANALYSE_INFO] } + {SN} + ({ dans la } + [nom de SECTION] + { [chiffre] })				
Réalisations lexicales	<i>Nous/on/je</i>	<i>analyser/aborder/analyser/examiner/étudier/considérer/proposer/exposer/discuter/développer...</i>	SN/ <i>que</i>	<i>(dans la Première/deuxième/troisième ou (1,2,3,I,II, III...))</i>	<i>section/partie/conclusion/annexe)</i>

Tableau 4-11 : Démonstration des réalisations lexicales d'une routine de guidage(planification de la structure discursive)

L'idée est de créer une correspondance entre les réalisations lexicales et les classes sémantiques dans l'équation syntaxique. Cette présentation permet une visualisation plus claire de cette correspondance.

c. Remise en ordre de ces 9 champs

Enfin, nous avons réorganisé le tableau en modifiant l'ordre des 9 champs. Ainsi, en reprenant le modèle de Tutin & Ji (2022) avec des modifications, nous avons finalement un modèle d'analyse des routines illustré par le tableau ci-dessous, qui prend toujours la routine de planification comme exemple :

Description sémantique	'L''auteur' 'traite' un sous-thème dans une 'partie textuelle' 'numérotée'
Exemple	<i>Dans la section 2, nous présentons... on formule en conclusion...</i>
Type de phrase	Phrase déclarative au présent, elle peut être voix passive.
Structure syntaxique	[pronom AUTEUR]+ { [verbe d'ANALYSE_INFO] } + {SN} + ({ dans la } + [nom de SECTION] + { [chiffre] })

Réalisations lexicales	<i>Nous/ on/ je</i>	<i>analyser/aborder/analyser/examiner/ étudier/considérer/proposer/exposer/ discuter/développer...</i>	SN/ <i>que</i>	<i>(dans la première/deuxième/troisième ou (1,2,3,I,II, III...))</i>	<i>section/ partie/ conclusion)</i>
Éléments périphériques	Complément de nom : <i>de l'/cet article/contribution</i> Marqueurs discursifs : <i>tout d'abord, enfin, d'une part, d'autre part</i> Une phrase annonce le plan qui marque le début de la planification de la structure textuelle, comme : <i>l'article se présente de la manière suivante.</i>				
Paramètres énonciatifs	cette routine s'adresse implicitement au lecteur pour donner des informations sur l'organisation du texte. Elle relève donc du dialogisme interlocutif .				
Position textuelle	Dans la partie introductive et le début d'une section locale.				
Fonctions rhétoriques	Annonce du plan textuel global et local des écrits scientifiques, avec parfois une fonction de transition entre des sections.				

Tableau 4-12 : Un exemple de la modélisation adaptée dans notre recherche

L'objectif de ce modèle d'analyse adapté consiste donc à fournir un descripteur systématique pour synthétiser de manière condensée et lisible tous les aspects linguistiques des routines de guidage du lecteur.

Conclusion de la partie II

Cette partie est consacrée au recueil, au traitement des données, ainsi qu'à la présentation d'un modèle d'analyse des routines de guidage du lecteur. Notre étude s'inscrit dans la linguistique de corpus et a mobilisé une approche hybride (*corpus-based* et *corpus-driven*) pour le recueil des routines.

Nous avons présenté le corpus à travers sa constitution et son annotation syntaxique. Le corpus TermITH comporte 5 millions de mots pour 10 disciplines de SHS et fournit un terrain d'observation précis et pertinent pour notre étude. De plus, le prétraitement de certains mots polylexicaux que nous jugeons nécessaires a complété l'annotation de ce corpus sans en modifier les propriétés fondamentales. Les ressources et outils comme le Lexicoscope, Scientext, la liste du LST élaborée à l'aide du même corpus et la méthode des ALR couplée avec les classes sémantiques issue du LST nous ont permis de préciser la démarche méthodologique afin d'extraire avec une meilleure efficacité et une meilleure exactitude des expressions polylexicales récurrentes dans les écrits scientifiques. Dans les ALR, 66 classes sémantiques ont été employées comme pivot de requêtes.

À l'issue du moissonnage automatique, la reconnaissance des données brutes comme routines de guidage demande un contrôle à la fois quantitatif et qualitatif qui s'intéresse, d'une part, à la récurrence et à la transdisciplinarité des routines dans le corpus, d'autre part, à la fonction rhétorique mise en œuvre dans ce guidage. L'identification de ces éléments phraséologiques reste une étape décisive de notre recherche linguistique, qui vise à élaborer

une typologie fonctionnelle pour ces phénomènes et à les analyser systématiquement en les inscrivant dans une réflexion plus profonde. Le recensement des exemples en contexte des routines ainsi que celui de leurs variations syntaxiques et lexicales dirigent le travail d'analyse. Il faut savoir que l'utilisation des classes sémantiques du LST ne permet pas de trancher d'un seul coup pour déterminer si la totalité des mots d'une classe peut former une routine avec la même interprétation sémantique ou une fonction semblable. Nous avons remarqué au cours de l'observation de nos données que les routines ont tendance à faire appel à une partie des mots dans les classes sémantiques, mais pas tous. Par conséquent, une observation minutieuse des données dans le contexte a été effectuée afin de recenser des mots qui entrent dans les routines. L'inventaire des routines ainsi que leurs variations lexico-syntaxiques est essentiel pour leurs analyses à travers le modèle d'analyse présenté précédemment qui puise son inspiration dans quelques modélisations intéressantes (telles que Pecman 2004, etc.). Nous avons finalement identifié un total de 37 routines de guidage différentes.

Enfin, nous avons proposé de suivre la modélisation linguistique de Tutin & Ji (2022, p. 59–60), en l'adaptant, pour procéder à une description approfondie en créant des étiquettes sémantiques pour les routines. Nous allons voir que tous les paramètres que nous avons pris en compte lors de l'extraction et du nettoyage des données constituent le fondement indispensable pour accéder à l'élaboration d'une typologie basée sur les fonctions primaires des routines de guidage et aux analyses approfondies qui font l'objet de la partie suivante.

Partie III. Typologie, modélisation et analyses des routines de guidage du lecteur

Nous avons déjà fait le recueil des routines de guidage du lecteur en indiquant leurs fonctions rhétoriques remplies dans les écrits scientifiques. Nous proposons maintenant d'élaborer une typologie fonctionnelle, une modélisation et les analyses linguistiques pour ces expressions phraséologiques. Cette typologie servira à comprendre les stratégies rhétoriques généralement mobilisées permettant de faciliter la lecture dans les écrits scientifiques. Les analyses linguistiques de ces routines mèneront des observations plus approfondies sur leurs caractéristiques syntaxiques, sémantiques, lexicales, énonciatives et fonctionnelles. Cela aidera, d'une part, les linguistes à mieux comprendre ces expressions complexes dans les écrits scientifiques, d'autre part, fournira des pistes intéressantes aux didacticiens pour les enseigner aux apprenants qui manifestent un besoin de maîtriser la rédaction scientifique en français.

Cette partie se divise en quatre chapitres avec une introduction, dans laquelle nous donnerons une explication succincte sur les grandes catégories des fonctions dans notre typologie. Les trois chapitres suivants se focalisent respectivement sur trois grandes fonctions des routines afin de mener une observation plus approfondie. Le dernier chapitre propose une étude synthétique sur des particularités textuelles, disciplinaires et énonciatives des routines.

De manière générale, notre catégorisation a été mise en place à partir de principes onomasiologiques qui partent d'un concept fonctionnel pour aller vers les expressions phraséologiques – les routines de guidage dans notre cas. Notre typologie recense trois grands groupes de fonctions : les fonctions liées à la structuration, les fonctions métalinguistiques et les fonctions qui visent la topicalisation des informations dans le discours scientifique. Avec cette logique, les chapitres 5, 6 et 7 portent respectivement sur un groupe de fonctions des routines de guidage du lecteur afin de mener une observation approfondie sur l'ensemble des routines retenues ainsi que leur fonctionnement dans les articles de recherches.

Notre typologie s'appuie sur des observations menées pour des marqueurs ou expressions métadiscursives, notamment dans l'étude de Hyland (2006) et dans celle d'Ädel (2010) sur les discours scientifiques écrits et oraux, une étude basée sur son propre travail de théorisation du métadiscours plus tôt (Ädel, 2006) (*cf.* Chapitre 3). Plus précisément, nous nous intéressons particulièrement aux fonctions interactives dans le modèle de Hyland et les fonctions métatextuelles dans celle d'Ädel. Ces deux taxinomies proposent ensemble

20 fonctions métatextuelles et métalinguistiques qu'il est possible de trouver dans les écrits scientifiques. Comme notre travail s'inscrit dans une démarche *corpus-driven*, les fonctions retenues dans notre corpus ne sont pas tout à fait pareilles que les fonctions de références. Au niveau des nominations et des définitions, nous n'avons pas repris les noms ou les définitions des fonctions de références, mais les avons ajustés afin de répondre au mieux à notre conceptualisation des classes.

En fonction des routines retenues dans notre corpus des articles de recherche, trois aspects nous apparaissent essentiels pour guider la lecture : premièrement, il faut prêter attention à la structure textuelle du discours ; deuxièmement, il faut clarifier le code ou la langue utilisée, c'est-à-dire, le plan métalinguistique ; enfin, nous incorporons la fonction de topicalisation dans notre typologie, provenue des études de Tutin (2018 a, p. 88) et de Ji et Tutin (2020). Ces routines aident à relever l'élément essentiel du contenu propositionnel pour le lecteur, ce qui montre que l'auteur anticipe la difficulté du lecteur confronté à un grand nombre d'informations et adopte une stratégie rhétorique lui permettant de choisir ce qui est le plus important pour son lecteur.

En tenant compte de ces trois aspects, nous envisageons ainsi trois dimensions pour regrouper des routines de guidage du lecteur : 1) la première dimension prend soin du niveau structurel du texte. Nous regroupons toutes les routines qui permettent de mieux organiser structurellement les écrits scientifiques, que nous nommons routines de **structuration textuelle** ; 2) la deuxième inventorie les routines de **fonctions métalinguistiques**. Cette dimension reflète la propriété réflexive des routines pour commenter la langue ou le contenu précédemment énoncé du discours scientifique en production ; 3) la dernière représente un choix stratégique de l'auteur dans le but de mettre en évidence les informations importantes, les constats essentiels, ou le thème principal de l'écrit scientifique. Nous les désignons par routines de **topicalisation**. Ces trois dimensions se subdivisent chacune en plusieurs fonctions rhétoriques, détaillées par la figure 12.

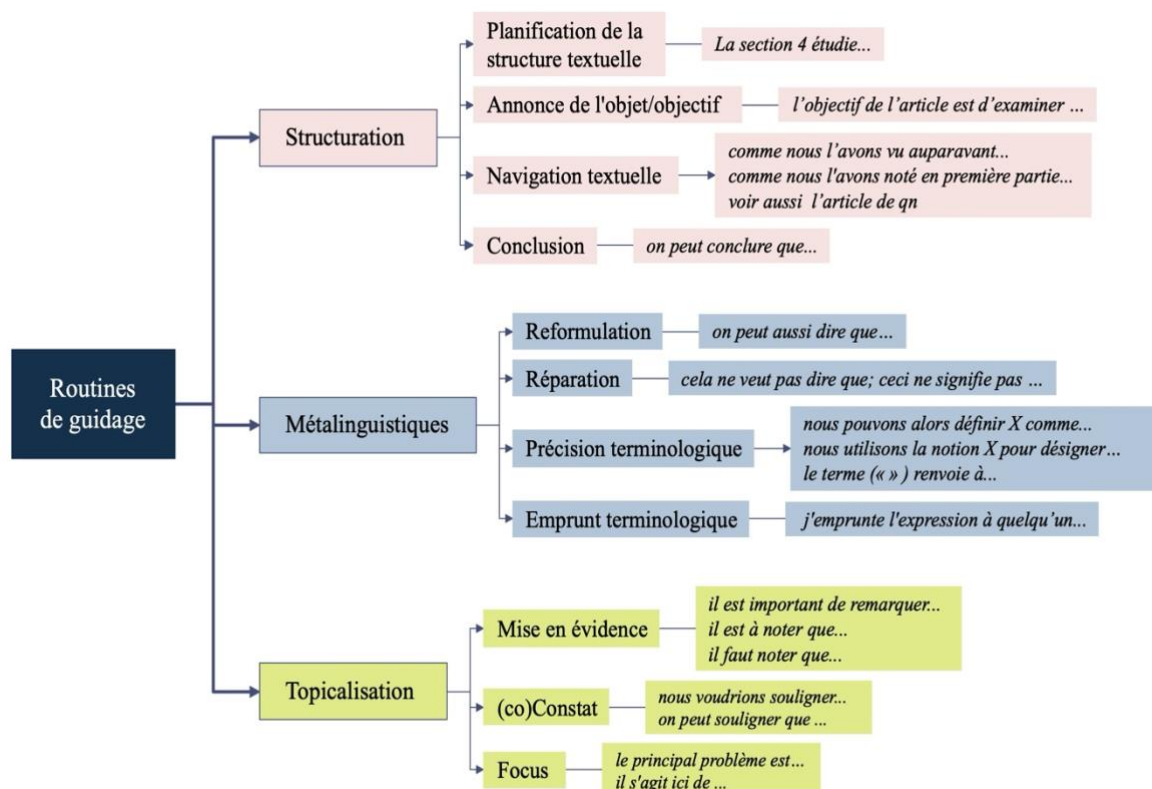


Figure Partie III-1 : Fonctions des routines de guidages dans les écrits scientifiques

Dans les chapitres suivants, nous expliquerons ces trois grandes catégories et détaillerons les définitions pour toutes les sous-fonctions observées avec des modélisations et des analyses des routines.

Chapitre 5. Routines de structuration textuelle : définition, modélisation et analyses

Ce chapitre est consacré à la présentation des fonctions de structuration ainsi que la modélisation et l'analyse des routines que nous avons identifiées dans notre corpus et regroupées dans cette grande dimension. Il convient de préciser d'abord l'organisation de la procédure d'analyse qui s'appliquera par la suite aussi aux chapitres 6 et 7 sur les routines métalinguistiques et les routines de topicalisation. Dans un premier temps, nous proposons une étude théorique sur la notion de structuration textuelle. Dans un deuxième temps, nous expliquerons notre définition de la structuration en la rapprochant des fonctions structurantes adoptées dans les modèles métadiscursifs discutés dans le chapitre 3. L'objectif de ce rapprochement est de montrer la différence et la ressemblance entre nos fonctions et celles des autres auteurs, ce qui permet de délimiter les définitions des fonctions retenues dans notre typologie. Dans cette section, une analyse qualitative (modélisation, analyse) des routines de chaque sous-fonction sera mise en œuvre. Pour finir, nous proposerons une observation sur les positions textuelles, les particularités disciplinaires, l'actualisation énonciative des routines, et nous proposerons une synthèse qui clôturera le présent chapitre.

La classe des routines de structuration textuelle regroupe des fonctions qui aident à organiser la structure du discours. La fonction générale est d'améliorer la lisibilité et le suivi de la progression du discours, par exemple, faire comprendre au lecteur l'objectif ou la problématique de l'article, la composition thématique du discours scientifique, le lieu textuel précis d'une information, la référence d'une source ou d'une information abordées dans le texte, les conclusions.

Même si notre définition des fonctions de structuration textuelle s'appuie sur les études anglophones sur le métadiscours, il faut aussi évoquer certaines études dans la tradition de la linguistique textuelle, qui permettent de comprendre d'un point de vue global les opérations structurant le texte.

5.1. Structuration dans la linguistique textuelle

Dans l'approche de la linguistique textuelle, la structuration du texte a déjà fait l'objet d'études approfondies, notamment dans les ouvrages d'Adam (2005, 2020). Généralement, deux grands types de moyens sont susceptibles de réaliser la structuration d'un discours : l'organisation typographique et l'organisation discursive des contenus propositionnels du discours (Adam, 2020).

Le premier type vise à planifier visuellement le texte à l'aide d'éléments non verbaux qui permettent d'agencer la présentation matérielle d'un texte. Nous avons abordé ce point de façon succincte dans la section 2.5 du deuxième chapitre en évoquant quelques éléments comme la gestion des espaces, des ponctuations, le titrage, le découpage, le surlignage, l'italique, etc. En linguistique textuelle, l'organisation typographique recourt principalement aux deux dispositifs théorisés par Colletta (2013, p. 20) : le paratexte et le péri-texte⁸⁴. Les dispositifs paratextuels dans le discours écrit mettent en place « la ponctuation et des marques sémiographiques [comme] changements de police, alinéa et retrait de paragraphe, bordures et autres caractéristiques de mise en page du texte [...], les notes de bas de page » (*ibid.*), des titres, intertitres, découpage en chapitres/parties (Adam, 2005, p. 52). Ces éléments, qui sont classés sous le terme « segmentation graphique » par Adam (*ibid.*), découpent et regroupent des « unités de niveaux différents de complexité » textuels en fonction des besoins sémantiques, syntaxiques et énonciatifs (*ibid.*), ce qui fournit un accès plus facile à la structure du texte. À côté du paratexte, le péri-texte enrichit l'organisation visuelle du discours par les éléments comme des « illustrations visuelles sur un support écrit, sonores ou audiovisuelles sur un support numérique qui accompagnent le texte » (Colletta, 2013, p. 20). Nous utilisons ici le terme « agencement typographique » pour désigner l'ensemble de ces deux dispositifs.

Cet agencement typographique n'est pas destiné à améliorer l'aspect esthétique du discours⁸⁵, mais joue un rôle fonctionnel qui vise à découper et hiérarchiser non seulement la composition visuelle, mais aussi le contenu sémantique du discours (Jacques, 2005). Dans le discours scientifique, ces dispositifs typographiques jouent un rôle primordial. Jacques (2014, 2017) a mené des observations sur les titres et intertitres des sections dans les discours scientifiques écrits et oraux. Elle a mis en évidence la fonction structurante des titres/intertitres qui s'exerce à la fois sur le plan matériel en facilitant le découpage et la hiérarchisation des parties textuelles, et sur le plan notionnel, en indiquant le contenu sémantique des sections dénommées par les titres/intertitres correspondants. Cette organisation matérielle du discours/texte contribue largement à sa lisibilité, rendant le suivi du texte plus facile pour le lecteur, malgré le fait qu'il ne constitue pas l'objet d'étude de notre recherche.

⁸⁴ La distinction entre le paratexte et le péri-texte est différente chez Lane (2008) : cet auteur envisage le péri-texte en tant que partie constituante du paratexte. Ce dernier est défini comme « un ensemble hétérogène de pratiques et de discours [...] variable suivant qu'il s'agit d'éléments du péri-texte et de l'épitéxte » (*ibid.*, p. 1380). La présence de ces éléments a une finalité pragmatique en permettant à la fois d'« informer, convaincre, asserter et argumenter » (*ibid.*).

⁸⁵ Mais on voit bien un effet esthétique introduit par l'agencement typographique.

Le deuxième moyen de structuration textuelle est lié au niveau discursif. Contrairement à l'organisation matérielle qui vise à structurer le texte en segmentant des parties textuelles (péritexte et paratexte), l'organisation discursive a pour objectif de lier des constituants textuels par les moyens linguistiques sur les aspects de microstructure et de macrostructure. L'aspect macrostructurel, un constituant du liage macro-textuel, préoccupe le plan de texte, les parties du plan de texte et les séquences (Adam, 2020, p. 241). L'aspect microstructurel se focalise sur l'articulation entre des propositions/énoncés (liage des propositions/énoncés) (Adam, 2020, p. 123).

Ce qui nous intéresse ici au niveau macro-textuel, c'est le plan du texte. Il peut être conventionnel pour un genre discursif, ou occasionnel sans modèle préétabli. Il s'agit de « la composition macro-textuelle du sens », souvent dictée par l'instance rhétorique d'un texte (*ibid.*, p. 241). Par exemple, les articles de recherches dans les domaines des sciences « dures » suivent souvent une structure conventionnelle – IMRAD : Introduction, Méthodologie, Résultats et Discussion (Swales, 1990). Chaque partie répond à la cohérence sémantico-pragmatique et s'intègre ensuite dans la globalité du texte académique.

Sur le plan microtextuel, plusieurs éléments verbaux entrent en jeu, allant des éléments grammaticaux comme les anaphores, aux éléments lexicaux comme les marqueurs d'organisation textuelle ou les connecteurs argumentatifs⁸⁶. Ces éléments linguistiques contribuent à la cohérence et la cohésion du texte. En ce qui concerne notre étude, nous nous intéressons particulièrement aux marqueurs d'organisation textuelle qui permettent de baliser le plan du texte et aider le lecteur à suivre la progression du texte. Ce sont souvent des marqueurs spatiaux (*dessus/dessous, plus loin*), temporels (*d'abord, ensuite, après*), énumératifs (*de plus, en second lieu*), les marqueurs de changement de topique (*quant à*), ou d'illustration et d'exemplification (*par exemple, notamment, entre autres*) (Adam, 2020, p. 169–173). Ces éléments sont qualifiés de marqueurs métadiscursifs dans les modèles métadiscursifs pour organiser la structure du discours. Ces marqueurs d'organisation textuelle, omniprésents dans les écrits scientifiques, ont fait l'objet d'une étude approfondie dans la thèse de Tran (2014) sous le terme de « séquences discursives polylexicales » (*cf.* section 3.1). Nous allons voir que ces marqueurs coexistent de façon dynamique avec les routines de structuration pour construire les écrits scientifiques.

⁸⁶ Voir le schéma 13 pour l'opération de liage assurant la continuité textuelle dans Adam, 2020, p. 123.

En résumé, l'approche linguistique textuelle propose deux moyens de structuration textuelle, l'un est typo-dispositionnel et l'autre est discursif. C'est à ce dernier niveau que nous envisageons les routines de structuration textuelle dans le discours scientifique. Il s'agit des moyens linguistiques permettant de présenter la composition textuelle du discours scientifique qui sont étudiés de façon systématique dans le métadiscours que nous allons aborder dans la section suivante.

5.2. Fonctions structurantes dans le métadiscours

Comme nous l'avons dit plus haut, notre étude ne s'intéresse pas à la disposition matérielle qui structure le discours de façon visuelle, mais privilégie la perspective discursive pour la structuration textuelle. La structuration est envisagée comme l'ensemble des fonctions métatextuelles qui guident la lecture à travers l'organisation discursive des contenus propositionnels du texte. Nous restreignons la portée de notre étude aux routines de structuration textuelle qui ne concernent pas le sens des contenus propositionnels, mais aident à mieux les organiser, en excluant des éléments grammaticaux (anaphores, déictiques), ou des connecteurs argumentatifs qui visent à structurer la progression logique du discours.

Notre définition de structuration textuelle s'appuie spécifiquement sur les travaux du métadiscours sur la langue anglaise académique pour un inventaire préliminaire des fonctions de structuration. Comme présentées dans le chapitre 3, ces études proposent concrètement des fonctions liées à la structuration du texte. Plus précisément, nous nous appuyons sur des fonctions de Hyland (2005) et Ädel (2006, 2010). Rappelons ici que Hyland (2005, p. 49) (*cf.* tableau 3-6) identifie quatre fonctions qui contribuent à la structuration du texte : la transition (plutôt argumentative), l'encadrement discursif, la fonction endophorique et la fonction de preuve qui visent à donner des références intratextuelles et intertextuelles. Ädel (2006, 2010) quant à elle recense onze fonctions métatextuelles dans l'écrit scientifique en anglais : introduction du thème, focalisation/délimitation sur le thème, conclusion du thème, addition, exemplification, énumération, rappel/révision, argumentation, contextualisation, prévision, fonction endophorique (*cf.* chapitre 3). Nos sous-classes sont créées sur la base de ces ressources en les adaptant à nos données.

Dans nos catégories de fonctions de structuration textuelle, nous n'avons pas retenu la fonction de transition, car elle renvoie plutôt à l'argumentation. Les marqueurs de transition classés par Hyland (2005, p. 220) sont des marqueurs d'addition (*and, in addition*), de comparaison (*in contrast, rather*), de conséquence (*the result is, thereby*). Nous avons conservé en revanche la fonction endophorique, les fonctions de preuve (référence intertextuelle),

d'encadrement (plan et organisation du texte), rappel/révision, prévision, ainsi que les fonctions d'introduction et de conclusion. Au lieu de reprendre les fonctions telles quelles, nous les avons réorganisées et nous proposons quatre sous-catégories avec de nouvelles dénominations. Dans l'ensemble, quatre fonctions permettant de structurer discursivement le discours scientifique ont été classées dans notre typologie : a) la fonction de planification de la structure discursive ; b) la fonction d'annonce de l'objet/objectif d'étude ; c) la navigation textuelle ; d) la conclusion. Nous les présentons de manière plus précise avec les routines concrètes dans ce qui suit.

5.3. Routines de structuration textuelle

Comme nous l'avons indiqué plus haut, la structuration du texte vise à mieux organiser le discours scientifique dans l'intention d'en faciliter la lecture. Elle ne traite pas le fond des contenus propositionnels, mais aide à mieux les orchestrer. Le discours scientifique, en particulier les articles de recherches, reste un sous-genre conventionnel dont l'organisation doit répondre à plusieurs critères afin de mieux guider son lecteur : savoir bien commencer un discours en annonçant le plan du texte et l'objectif d'étude, situer facilement une information, opérer la transition entre les parties textuelles, clôturer une argumentation, une partie textuelle ou le discours entier. Cette section comporte quatre sous-sections qui présentent respectivement les quatre fonctions évoquées plus haut ainsi que les routines identifiées en indiquant leur modélisation comprenant une analyse synthétique de leurs caractéristiques sémantiques, syntaxiques, lexicales, énonciatives ainsi que leurs fonctionnements rhétoriques. Pour la fonction de structuration textuelle, 18 routines de structuration ont été repérées dans notre corpus.

Nous commençons par le plan du discours que nous appelons « la fonction de planification de la structure discursive ».

5.3.1. Routines de fonction de planification de la structure discursive

La fonction de planification de la structure discursive consiste à hiérarchiser les informations qui seront développées dans le corps d'un article de recherche en SHS. Contrairement aux articles de recherches en sciences dures qui respectent généralement le schéma prototypique Introduction, Méthode, Résultats Et Discussion, les articles en SHS emploient souvent des structures selon les besoins réels du domaine ou de l'auteur. Standardisée en 1972 par l'American National Standards Institute (ANSI), la structure IMRED est devenue le format le plus utilisé pour les articles de recherche en sciences dures ou expérimentales (Nair & Nair, 2014, p. 14). Ce format relève simultanément des dispositifs typographiques comme le périphrase ou le paratexte discuté au début de ce chapitre, et des dispositifs cognitifs qui

permettent de fournir une « trame argumentative » (Pontille, 2007, p. 231). Mais on ne trouve guère de conseils rigides pour les articles de recherches en SHS qui possèdent bien sûr des structures, mais plus flexibles. Comme on n'impose pas de structure universelle dans les disciplines en SHS, la planification de la structure du discours devient essentielle pour clarifier les liens et les segmentations entre les parties textuelles. Elle est un atout pour la réussite d'un discours scientifique bien construit et facile à suivre. Cela s'observe particulièrement dans les articles de recherche en SHS où nous trouvons des moyens linguistiques pour communiquer explicitement l'organisation du contenu.

La fonction de planification ressemble à un type de marqueurs d'encadrement qui réfère au séquençement du discours du modèle de Hyland (2005). Ces marqueurs métadiscursifs indiquent l'ordre des matériels propositionnels (*cf.* Chapitre 3) à travers quatre fonctions métatextuelles précises : la mise en ordre (*first of all, in this section, seconde, then*), l'étiquetage des étapes discursives (*at this point, to conclude*), la prédiction de l'objectif d'étude (*desire to, in this chapter*) et le changement thématique (*back to, in regard to*). La planification structurelle dans notre définition est proche de la fonction de mise en ordre, mais concerne davantage le plan global du discours, permettant ainsi au lecteur de comprendre le scénario logique et argumentatif qui va se dérouler dans le discours. Quant à l'étiquetage séquentiel du discours et l'annonce de l'objectif, nous les classons dans d'autres fonctions de structuration.

En résumé, nous pouvons définir la fonction de planification de la structure discursive comme une stratégie rhétorique qu'emploie l'auteur de l'article de recherche pour annoncer au lecteur la composition logique des parties textuelles de son texte. Nous avons identifié deux routines à trois structures syntaxiques qui permettent de remplir cette fonction.

Ces deux routines sont très présentes dans l'introduction des articles de recherche, et présentent des caractéristiques syntaxiques et sémantiques à la fois régulières et variées. Voilà la modélisation de ces routines :

Routine N° 1 de planification :

Description sémantique	<Une 'partie textuelle' 'numérotée' 'traite' un sujet dédié>			
Exemples	<i>La première partie de cet article présente...</i> <i>La dernière section conclut l'article.</i> <i>... sont présentés dans la section 2.</i>			
Type de phrase	Phrase déclarative au présent, elle peut être à la voix passive.			
Structure syntaxique	1. [nom de SECTION] + { [chiffre] [adjectif d'ÉNUMÉRATION] } + (de l'/cet article) + { [verbe d'ANALYSE_INFO] } + SN 2. SN + être + { [verbe d'ANALYSE_INFO] (pp) } + dans + [nom de SECTION] + { [chiffre] [adjectif d'ÉNUMÉRATION] }			
Réalisations lexicales	1. La première/deuxième/troisième/dernière partie/section ou 1,2,3...	(de l'/cet article/contribution)	analyse/examine/présente/est consacré à/caractérise/propose/décrit/expose/considère...	SN
	2. SN	être	étudié/décrit/développé/résumé/donné	dans la troisième section ou (1, 2, 3...)
Éléments périphériques	Complément de nom : de l'/cet article/contribution Marqueurs discursifs : <i>enfin</i> Une phrase qui annonce le plan et marque le début de la planification de la structure textuelle, comme : <i>L'article se présente de la manière suivante.</i>			
Paramètres énonciatifs	Cette routine s'adresse implicitement au lecteur pour donner des informations sur l'organisation du texte. Elle relève donc du dialogisme interlocutif .			
Position textuelle	Dans la partie introductive.			
Fonctions	Planification de la structure textuelle des écrits scientifiques			

Tableau 5-1 : Routine N° 1 de planification de la structure discursive

Routine N° 2 de planification :

Description sémantique	'L' 'auteur' 'traite' un sous-thème dans une 'partie textuelle' 'numérotée'			
Exemples	<i>Dans la section 2, nous présentons...</i> <i>on formule en conclusion...</i>			
Type de phrase	Phrase déclarative au présent, elle peut être à la voix passive.			
Structure syntaxique	[pronom AUTEUR] + (vouloir) + { [verbe d'ANALYSE_INFO] } + {SN} + { [dans la] + [en] } + [nom de SECTION] + { [chiffre] [adjectif d'ÉNUMÉRATION] }			
Réalisations lexicales	Nous/ on/ Je (vouloir)	analyser/aborder/analyser/examiner/ étudier/considérer/proposer/exposer/ discuter/développer...	SN/ que	(dans la première/deuxième/troisième ou (1,2,3,I,II, III...)) section/ partie/ conclusion)
Éléments périphériques	Complément de nom : de l'/cet article/contribution Marqueurs discursifs : <i>tout d'abord, enfin, d'une part, d'autre part</i> Une phrase qui annonce le plan et marque le début de la planification de la structure textuelle, comme : <i>L'article se présente de la manière suivante.</i>			
Paramètres énonciatifs	Cette routine s'adresse implicitement au lecteur pour donner des informations sur l'organisation du texte. Elle relève donc du dialogisme interlocutif .			
Position textuelle	Dans la partie introductive et le début d'une section locale.			
Fonctions	Annonce du plan textuel global et local des écrits scientifiques, avec parfois une fonction de transition entre des sections.			

Tableau 5-2 : Routine N° 2 de planification de la structure discursive

5.3.1.1. *Caractéristiques syntaxiques, lexicales et sémantiques de ces routines*

Ces deux routines sont des phrases déclaratives au présent⁸⁷. La routine N° 1 peut aussi se réaliser par la voix passive comme le montre l'exemple 2.

- [1] **La troisième section expose et caractérise** les différentes situations migratoires régionales, ainsi que les principaux changements intervenus durant les quinze dernières années. (Géographie) (routine N° 1)
- [2] Les données utilisées pour son estimation **sont décrites dans la section 3**. (Économie) (routine N° 1)
- [3] **Dans la troisième partie, nous exposons** les résultats d'une recherche dans une entreprise industrielle d'envergure mondiale réalisée sur le métier de manager de proximité. (Psychologie) (routine N° 2)

Comme nous l'avons expliqué dans le chapitre 5, la description sémantique d'une routine comporte une glose qui permet de mieux comprendre à la fois la fonction, la structure syntaxique, le contenu sémantique, ainsi que les actes discursifs des écrits scientifiques qui jouent dans une routine. Les étiquettes conceptuelles fréquentes sont '**partie textuelle**', '**numérotée**', '**traite**' et '**auteur**' dont les trois premières sont partagées par les deux routines. '**Partie textuelle**', '**numérotée**' renvoient ensemble aux sections principales de l'article de recherche (*section 1* ou *la troisième partie*).

'**Traite**' renvoie à l'acte discursif fondamental dans les écrits scientifiques : traiter un thème ou une problématique. Le sous-thème est un élément obligatoire pour construire une phrase complète lors de la mise en œuvre du discours.

Dans les deux variations syntaxiques, le verbe d'ANALYSE_INFO (par exemple *analyser, examiner, décrire, préciser...*) occupe la place de prédicat verbal et désigne l'acte '**traiter**', le chiffre ou l'adjectif numéral correspondant à '**numérotée**' qualifie le nom de section, tandis que le nom de SECTION associé à une '**partie textuelle**' peut exercer des fonctions syntaxiques différentes. Dans la première structure (exemple 1), il s'agit du sujet de la phrase. Dans la deuxième structure (exemple 2), il constitue avec la préposition *dans* et le déterminant défini *le* un complément circonstanciel de lieu. Le syntagme nominal, un élément obligatoire, mais pas fixe devient le sujet de la routine. Les éléments dans les parenthèses (*de l'/cet article*) sont facultatifs. Leur existence n'a pas d'impact sur la fonction et le sens de la routine, mais peut en renforcer l'inscription énonciative. Les structures de l'exemple 1 et de l'exemple 2 ont été fusionnées dans la même routine N° 1.

⁸⁷Contrairement à nos attentes, il y a très peu de futur (2 occurrences pour 115 au total).

Les étiquettes sémantiques attribuées aux parties fixes sont reprises des classes sémantiques du LST. Dans les routines de planification, on voit que les noms de SECTION, les verbes d'ANALYSE_INFO et de COMMUNICATION et les adjectifs d'ÉNUMÉRATION sont des paradigmes qui apparaissent habituellement ensemble, avec comme fonction d'annoncer le plan des écrits scientifiques. Plus précisément, quatre noms sont retenus dans le corpus pour le nom de SECTION : *section*, *partie*, *annexe* et *conclusion*, parce que notre corpus n'est composé que d'articles de recherche. On peut supposer que les écrits scientifiques comme la thèse ou le mémoire peuvent intégrer d'autres composants textuels tels que *chapitre*, etc.

Les verbes d'ANALYSE_INFO renvoient aux actions scientifiques typiquement fondamentales dans le domaine scientifique qui visent à « présente[r] les démarches scientifiques et le processus de construction des connaissances dans la discipline » (Yan, 2017, p. 126). Ils comportent 14 sous-classes comme la description des données, le constat d'un fait scientifique, la démonstration de la vérité, la comparaison, l'évaluation, la conclusion, etc.⁸⁸ Les verbes identifiés appartiennent notamment aux quelques sous-classes spécifiques comme les verbes d'examen pour « étudier en détail, faire l'analyse de quelque chose pour en avoir une connaissance approfondie, juger de sa valeur » (cité du site LST : <http://lst.demarre-shs.fr/classification>) comme *aborder*, *analyser*, *examiner*, *étudier*, *approfondir*... ou les verbes de description destinés à représenter en détail les caractéristiques d'un objet d'étude (*ibid.*) comme *présenter*, *caractériser*, *décrire*, *illustrer*, *représenter*.

En parallèle, certains verbes destinés à communiquer et transférer des informations et des idées dans les écrits scientifiques comme *insister (sur)*, *proposer*, *suggérer*, *exposer*, *résumer* rentrent aussi dans la construction des deux routines. Au lieu de renvoyer à un acte scientifique, ces verbes de COMMUNICATION renvoient plutôt aux actes communicatifs liés à la recherche. Un écrit scientifique est avant tout un lieu de communication sur les travaux scientifiques. L'emploi de ces verbes permet à la routine de s'inscrire dans le déroulement énonciatif du discours, et d'accomplir l'argumentation du scénario scientifique avec les actes scientifiques.

En dehors de ces deux grandes classes de verbes, ces routines incluent certains verbes plus généraux qui ne relèvent pas spécifiquement du domaine scientifique, comme *fournir*, *donner*, *offrir*, *conduire*. D'autres verbes sont répartis dans d'autres champs sémantiques (*introduire*, *développer*, *conclure*, *contenir*) ou n'ont pas été inclus dans le LST comme

⁸⁸ Voir la définition proposée par le site LST : <http://lst.demarre-shs.fr/classification>

réintroduire, faire l'objet de. Certaines expressions polylexicales peuvent également participer à la construction de cette routine et occupe la place du prédicat comme : *faire l'objet de, mettre en parallèle, être consacré à, être dévolu à, avoir pour but de, s'attacher à.*

- [4] La totalité des démonstrations **est contenue dans une annexe.** (Économie)
- [5] L'examen des comportements des jeunes **fait l'objet de la deuxième section.** (Géographie)
- [6] **La section 2 est consacrée à** la présentation de notre modèle théorique. (Économie).
- [7] **Nous représenterons dans la section suivante** les sous-espaces pertinents dans le plan (M, u) pour un vecteur de paramètres fixés. (Économie) (dans le corps de l'article)
- [8] **Nous illustrons** le travail d'implémentation **dans la section 4.** (Linguistique)

Nous pouvons constater que l'emploi des verbes est assez flexible, souvent lié à la problématique ou au thème de la section cible. En observant de plus près les verbes liés à ces routines, les plus récurrents sont ceux qui renvoient aux actes scientifiques (verbe d'ANALYSE_INFO) et communicatifs (verbe de COMMUNICATION).

Le chiffre et l'adjectif d'énumération qualifient une partie textuelle et précisent son ordre dans l'ensemble des parties d'un texte (*1, 2, 3...* ou adjectifs numéraux *premier, deuxième...*). En résumé, les routines de planification comportent trois champs fondamentaux : deux champs pour indiquer une section précise du texte, et le troisième pour marquer l'acte scientifique ou énonciatif qui permet d'annoncer le contenu à prévenir.

Sur le plan sémantique, les deux routines de planification sont très proches avec trois éléments fixes communs ('**partie textuelle**', '**numérotée**', '**traite**') qui sont aussi essentiels pour construire des routines de planification. La seule différence réside dans l'emploi du pronom auteur, ce qui permet de mettre en scène ou non le sujet auteur. Elles sont par essence des expressions équivalentes du point de vue sémantique et fonctionnel. Pour cette raison, l'emploi des deux routines de planification dans les articles de recherches peut apparaître en alternance dans un même article. En réalité, elles s'enchaînent fréquemment et apparaissent, dans la majorité des cas, au moins 3 fois dans un paragraphe, comme le montre l'exemple 9. L'enchaînement des routines de planification construit une ossature discursive qui intègre toutes les informations récapitulatives essentiellement pour aborder la structure des sections textuelles du discours scientifique.

Par ailleurs, l'auteur peut recourir à des phrases de prélude. *L'article s'organise de façon suivante* dans l'exemple 9 (souligné en rouge) se place devant toutes les routines de

planification et anticipe sur la structure du texte pour le lecteur. Ces phrases de prélude sont de deux types : 1) comme l'exemple 9, elles précisent le mode d'organisation ; 2) comme en (10), elles annoncent directement le nombre de sections du texte.

[9] **L'article s'organise de la façon suivante.** L'enquête utilisée **est présentée dans la section 2 (routine N° 1)**, ainsi que l'échantillon sur lequel porte l'analyse économétrique. **Dans la troisième section, nous explicitons (routine N° 2)** le modèle économétrique sur lequel s'appuient les analyses empiriques. **Dans les sections suivantes, nous présentons et analysons (routine N° 2) les résultats.** (Économie)

[10] **Notre article est divisé en cinq sections.** **La première section propose** un bref rappel de la littérature consacrée aux modèles de prospection et indique ses développements récents NNN8. **La deuxième section présente les hypothèses** de notre modèle et caractérise le concept d'équilibre et le critère de bien-être social. L'effet de liquidité de la monnaie et le seuil optimal de la quantité de monnaie sont étudiés dans la troisième section. **La quatrième section réintroduit** l'effet d'éviction et endogénéise la contrainte postulée par KIYOTAKI et WRIGHT [1991, 1993] portant sur le stock d'encaisses des agents. **La dernière section conclut l'article.** (Économie)

En observant les SN suivant les verbes d'ANALYSE_INFO, nous avons remarqué des collocations régulières. Ainsi, *présenter* forme souvent des collocations avec résultat comme *présenter les résultats* (exemple 9, souligné en italique), *présenter les hypothèses* (exemple 10), *présenter le modèle* (exemple 11).

[11] **La première section présente le modèle** et calcule le revenu et le taux d'inflation d'équilibre dans les n pays de l'Union, qui dépendent à la fois de la politique monétaire et des politiques budgétaires NNN2... (Économie)

En bref, les routines de planification s'enchaînent pour former des routines d'ossature plus large comme *sentence frames* de Lewis (1996) (cf. section 1.2.3.). En outre, les phrases de prélude indiquant le début d'une annonce du plan textuel sont également cooccurrentes avec ces routines, ce qui renforce la volonté de l'auteur de mieux accompagner la lecture.

5.3.1.2. *Éléments périphériques et paramètres énonciatifs*

Deux éléments périphériques participent souvent à la construction de cette routine : le complément du nom *de l'article* ou *de cet article* (12, 14) qui précise que la section abordée appartient bien au discours en cours, mais pas à un autre discours. Cet élément n'est pas obligatoire et n'influence ni le sens fondamental ni la fonction de la routine. Les marqueurs discursifs des étapes de l'énonciation comme le commencement *tout d'abord* (14), des marqueurs de structuration *d'une part, d'autre part* (16), de conclusion *enfin* (13) permettent de structurer les étapes et les énumérations du discours académique.

- [12] **La première partie de l'article propose** un constat : animés par des objectifs de prévention des risques professionnels et de promotion de la santé, la plupart des acteurs pose la question du « psychosocial » à partir de la question de l'exposition à des risques dans les milieux de travail. (Psychologie)
- [13] **Enfin, la Section 7 présente** quelques remarques de conclusion et replace notre article dans la perspective plus générale de la théorie des incitations. (Économie)
- [14] **Dans la dernière partie de l'article, nous analysons** les effets sur le bien-être des diverses modalités de financement de l'audit. (Économie)
- [15] **Dans la Section 3, nous discutons tout d'abord** le cas où les préférences des consommateurs sont plus complexes que dans le modèle canonique (Économie)
- [16] **Nous montrons en Section 4, d'une part que** la politique de vente en primeur rend possible la discrimination intertemporelle en prix, pourvu que la quantité vendue en primeur soit rationnée, **d'autre part que** cette politique rapporte au monopole un profit plus grand que la politique de prix unique. (Économie)

Ces routines intègrent une dimension interlocutive du dialogisme, mais de façon implicite, dans la mesure où la mention du plan a pour objectif de rendre le texte plus clair pour le lecteur. L'auteur anticipe l'attente de son lecteur (pour les articles de recherches, ce sont souvent les pairs) et utilise ces expressions récurrentes afin de clarifier la composition du texte. Mais comme ces routines n'évoquent pas directement le lecteur par un marqueur linguistique spécifique, elles restent dans un dialogisme interlocutif implicite.

5.3.1.3. *Fonctionnement rhétorique*

La fonction de planification de la structure textuelle est fondamentale pour construire un article académique clair. Rappelons que dans l'approche linguistique textuelle, Adam (2005) accorde une place importante au plan du texte. En ce qui concerne le discours scientifique, Swales (1990, p. 141) propose un modèle « Create a Research space » (voir la figure 5-1) afin de considérer les mouvements rhétoriques dans l'introduction des articles académiques toutes disciplines confondues. Ce modèle propose une relation d'analogie entre les mouvements rhétoriques observés dans les introductions des articles de recherche et le système écologique. « Écrire une introduction » signifie créer et occuper une place dans cet écosystème. Trois mouvements avec des étapes obligatoires ou facultatives sont proposés : « établir un territoire », « établir une niche » et « occuper la niche »⁸⁹ (voir la figure 5-1).

La planification de la structure concerne le troisième mouvement après les deux premiers qui contextualisent et problématisent la recherche. Dans le mouvement 3 « occuper la

⁸⁹ C'est notre traduction.

niche » (souligné par nous dans la figure 5-1), l'étape 3 consiste à indiquer la structure des articles de recherche. Cette étape se trouve souvent à la fin de l'introduction et cherche à expliquer les détails de l'organisation du contenu propositionnel d'un article de recherche. Elle peut commencer par une phrase de préluce comme nous l'avons observé dans la section 5.3.1.1 : *L'article s'organise de la façon suivante...* et puis s'enchaîne avec les routines de planification qui explicite les étapes énonciatives du discours en question (voir les exemples 9, 10 cité plus haut). Nous pouvons voir aussi que les autres étapes, comme la mise en évidence des objectifs et l'annonce de la présente recherche occupent une place essentielle dans le mouvement 3. Leur objectif consiste à « increasing explicitness », autrement dit, assurer le bon suivi du côté du lecteur. Nous y revenons pour les routines d'annonce de l'objectif/objet.

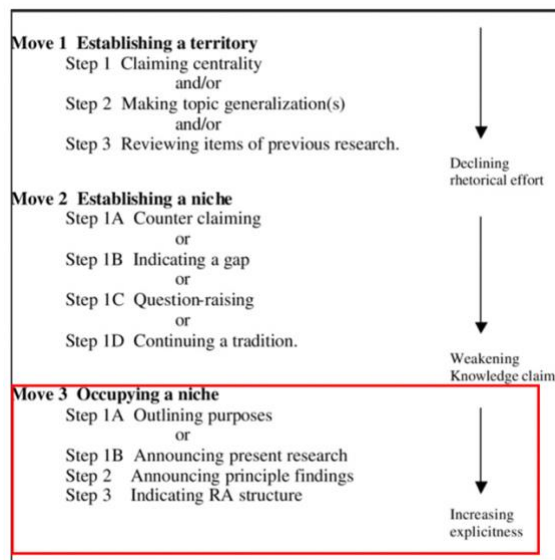


Figure 5-1 : Modèle CARS pour l'introduction des articles de recherche en anglais (Swales, 1990, p.141)

Bien que la plupart des routines de planification se trouvent dans la partie introductive du discours, elles peuvent aussi remplir une fonction secondaire de transition entre des sections du texte en dehors de la planification du plan général, quand elles se positionnent dans le corps principal du texte. Dans un discours plus long comme une thèse ou un mémoire, ces routines peuvent se trouver dans les introductions de chapitre. La transition qui annonce le plan local d'une partie textuelle précise, permet de rappeler au cours de la lecture la progression des contenus, ce qui aide le lecteur à se repérer dans l'activité de lecture (exemple 17).

- [17] **Notre discussion s'organise en deux points. La section 2 est consacrée à la construction et à la résolution d'un modèle de pollution cumulative aménagé pour étudier les conséquences d'une catastrophe climatique irréversible. La dernière section s'attache à interpréter et à approfondir les résultats de la section 2. (Économie, dans la discussion).**

La mise en œuvre de ces routines permet à l'auteur de montrer d'une part, sa capacité à organiser discursivement la recherche menée, d'autre part, sa préoccupation de prendre en charge les besoins de son lecteur virtuel.

5.3.2. Routines d'annonce de l'objectif/l'objet d'étude

L'article de recherche se construit toujours autour d'une problématique de recherche à traiter. Identifier une problématique de recherche signifie « circonscrire et décrire de manière pertinente l'objet de recherche, le (les) objectif(s) de recherche, le (les) problème(s) ou la (les) question(s) de recherche » (Class & Schneider, 2015, p. 53) dès le début du texte. Cette délimitation de l'objet/l'objectif de recherche constitue un facteur clé qui permet à l'auteur de planifier la structure discursive en permettant au lecteur de comprendre tout de suite l'intention de communication et de situer l'étude de façon plus aisée dans le domaine concerné (*ibid.*).

Cette fonction se rapproche de la fonction d'introduction du topique (*In the course of this essay, we shall...*) du modèle d'Ädel (2006, 2010) et la prédiction de l'objectif d'étude (*desire to, in this chapter*) chez Hyland (2005) (voir section 3.4.1). Dans notre typologie, nous nommons cette fonction « annonce de l'objectif/l'objet d'étude ». Il s'agit ici d'exposer au lecteur le but que l'auteur veut atteindre ou le thème que l'auteur souhaite aborder dans son article.

Les routines d'annonce de l'objectif/objet sont plus fréquentes et plus nombreuses dans notre corpus que celles d'annonce du plan. Nous avons retenu 6 routines dans cette sous-fonction. Ces routines ont parfois une structure syntaxique et sémantique analogue à celles d'annonce du plan (*nous proposons dans cet article de présenter...*), mais leur fonction est réservée à l'objectif ou l'objet d'étude. Voilà les 6 routines catégorisées dans cette fonction d'annonce de l'objectif/l'objet d'étude.

Routine N° 1 d'annonce de l'objectif/l'objet d'étude :

Description sémantique	<Le 'but' du 'discours en cours' 'traite' un thème>				
Exemples	<i>L'objectif de cet article est d'analyser...</i> <i>Notre objectif consiste à évaluer...</i> <i>L'objectif est ici de voir...</i>				
Type de phrase	Phrase déclarative au présent, au futur. Elle peut être au passé et à la forme négative.				
Structure syntaxique	1. $\left\{ \begin{matrix} le \\ l' \\ les \end{matrix} \right\} + [\text{nom de FINALITÉ}] + (de \ l' / \ cet + [\text{nom de DOCUMENT}]) + \left\{ \begin{matrix} est \ de \\ consiste \ à \\ vise \ à \end{matrix} \right\} + \left\{ \begin{matrix} [\text{verbe d'ANALYSE_INFO}] \\ [\text{verbe de PROCESSUS}] \end{matrix} \right\} + \text{SN}$				
	2. $\left\{ \begin{matrix} notre \\ mon \end{matrix} \right\} + [\text{nom de FINALITÉ}] + \left\{ \begin{matrix} est \ de \\ consiste \ à \end{matrix} \right\} + \left\{ \begin{matrix} [\text{verbe d'ANALYSE_INFO}] \\ [\text{verbe de PROCESSUS}] \end{matrix} \right\} + \text{SN}$				
Réalisations lexicales	<i>L'objectif/</i> <i>l'objet/</i> <i>le but</i>	<i>(de \ l'(la)/cet \ (te)</i> <i>article/papier/contribution/</i> <i>recherche)</i>	<i>est de/</i> <i>consiste à/vise à</i>	<i>analyser/examiner/étudier/</i> <i>montrer...</i>	SN
	<i>Notre/mon</i>	<i>objectif/objet/but</i>	<i>est de/</i> <i>consiste à/vise à</i>	<i>montrer/caractériser/</i> <i>élucider/développer...</i>	SN
Éléments périphériques	Marqueurs discursifs : de déduction : <i>ainsi, donc, alors</i> ; de réexamen : <i>en effet</i> ; de clarification : <i>plus précisément, précisément</i> ; Marqueurs de localisation : <i>ici</i> et marqueurs de manière : <i>justement</i> .				
Paramètres énonciatifs	Cette routine s'adresse implicitement au lecteur pour annoncer l'objectif d'étude. Elle relève donc du dialogisme interlocutif .				
Position textuelle	Dans la partie introductive (la majorité), le corps de l'article et la conclusion.				
Fonctions	Fonction principale : annonce de l'objectif d'étude ; fonctions secondaires : rappel de l'objectif d'étude, précision de l'objectif d'étude.				

Tableau 5-3 : Routine N° 1 d'annonce de l'objectif/l'objet d'étude

Routine N° 2 d'annonce de l'objectif/l'objet d'étude :

Description sémantique	Le 'discours en cours' ('vise à') 'traite (r)' un thème				
Exemples	<i>L'article s'attache à montrer...</i> <i>Cette contribution se proposera d'examiner...</i>				
Type de phrase	Phrase déclarative au présent, au futur. Elle peut être au passé.				
Structure syntaxique	déterminant+[nom de DOCUMENT]+([verbe de FINALITÉ])+[verbe d'ANALYSE_INFO] +SN				
Réalisations lexicales	<i>le/l'/</i> <i>ce/cet/cette/</i> <i>notre/mon (a)</i>	<i>article/texte/</i> <i>contribution/</i> <i>travail/recherche</i>	<i>s'attache à/tente de/</i> <i>(se) propose de/s'efforce de/</i> <i>vise à/chercher à</i>	<i>analyser/examiner/</i> <i>étudier/montrer...</i>	SN
Éléments périphériques	Très peu de marqueurs discursifs (1 occurrence de l'adverbe de déduction <i>donc</i> , 2 occurrences avec l'adverbe de clarification <i>précisément</i> , 1 occurrence de l'adverbe de manière <i>plus modestement</i>), un adjectif qualificatif <i>présent</i> du sujet.				
Paramètres énonciatifs	Cette routine s'adresse implicitement au lecteur pour annoncer l'objectif d'étude. Elle relève donc du dialogisme interlocutif . Déterminant en tant qu'indice énonciatif.				
Position textuelle	Dans la partie introductive (la majorité), la conclusion.				
Fonctions	Fonction principale : annonce de l'objectif d'étude ; fonctions secondaires : rappel de l'objectif d'étude.				

Tableau 5-4 : Routine N° 2 d'annonce de l'objectif/l'objet d'étude

Routine N° 3 d'annonce de l'objectif/l'objet d'étude :

Description sémantique	Le 'discours en cours' 'communiqué' une 'étude'			
Exemples	<i>L'article propose une analyse...</i> <i>Cet article propose une étude...</i>			
Type de phrase	Phrase déclarative au présent.			
Structure syntaxique	<i>d</i> déterminant + [nom de DOCUMENT] + [verbe de PROPOSITION] + $\left\{ \begin{array}{l} \text{[nom d'EXAMEN]} \\ \text{[nom d'INTERPRÉTATION]} \end{array} \right\}$			
Réalisations lexicales	<i>le/l'/ce/cet/cette/</i>	<i>article</i>	<i>propose</i>	<i>une étude/une analyse/une interprétation...</i>
Éléments périphériques	Pas d'élément périphérique.			
Paramètres énonciatifs	Cette routine s'adresse implicitement au lecteur pour annoncer l'objectif d'étude. Elle relève donc du dialogisme interlocutif . Déterminant en tant qu'indice énonciatif.			
Position textuelle	Dans la partie introductive (la majorité), la conclusion.			
Fonctions	Fonction principale : annonce de l'objectif d'étude ; fonctions secondaires : rappel de l'objectif d'étude.			

Tableau 5-5 : Routine N° 3 d'annonce de l'objectif/l'objet d'étude

Routine N° 4 d'annonce de l'objectif/l'objet d'étude :

Description sémantique	Le 'discours en cours' 'se focalise sur' un thème			
Exemples	<i>L'article s'intéresse à...</i> <i>Cet article (s') interroge (sur)...</i>			
Type de phrase	Phrase déclarative au présent.			
Structure syntaxique	<i>d</i> déterminant + [nom de DOCUMENT] + [verbe de THÈME] +			
Réalisations lexicales	<i>le/l'/ce/et/notre</i>	<i>article/étude/papier/travail</i>	<i>s'intéresse à/se focalise sur/(s')interroge (sur)...</i>	SN
Éléments périphériques	Marqueurs discursifs : de déduction : <i>donc, ainsi</i> ; de conclusion : en conclusion. Marqueurs de clarification : <i>plus spécifiquement, spécifiquement</i> . Marqueur d'intensité : <i>plus particulièrement</i> , et l'adjectif <i>présent</i> pour qualifier le nom de DOCUMENT.			
Paramètres énonciatifs	Cette routine s'adresse implicitement au lecteur pour annoncer l'objectif d'étude. Elle relève donc du dialogisme interlocutif . Déterminant en tant qu'indice énonciatif.			
Position textuelle	Dans le résumé, l'introduction, le corps de l'article.			
Fonctions	Fonction principale : annonce de l'objectif d'étude ; fonctions secondaires : précision de l'objectif d'étude.			

Tableau 5-6 : Routine N° 4 d'annonce de l'objectif/l'objet d'étude

Routine N° 5 d'annonce de l'objectif/l'objet d'étude :

Description sémantique	<L' 'auteur' 'vise à' 'traiter' un thème (dans le 'discours en cours' ou dans une 'partie textuelle')>					
Exemples	<i>Nous proposons dans cet article d'analyser...</i> <i>Nous proposons ici d'étudier</i>					
Type de phrase	Phrase déclarative au présent.					
Structure syntaxique	[pronom AUTEUR]+ verbe $\subseteq \left\{ \begin{array}{l} \text{[aller vouloir]} \\ \text{[Verbe de FINALITÉ]} \\ \text{[verbe de PROPOSITION]} \end{array} \right\} + [\text{verbe d'ANALYSE_INFO}] + (\text{dans} + \left\{ \begin{array}{l} \text{nom de DOCUMENT} \\ \text{nom de SECTION} \end{array} \right\}) + \text{SN}$					
Réalisations lexicales	<i>nous/on</i>	<i>(aller/vouloir)</i>	<i>se proposer/ tenter de/ proposer...</i>	<i>analyser/présenter/ clarifier...</i>	<i>dans cet/ce papier/ article...</i>	SN
Éléments périphériques	Marqueurs discursifs : de déduction : <i>donc</i> ; de clarification : <i>précisément</i> ; d'opposition : <i>en revanche</i> . Marqueur de localisation : <i>ici</i> . Marqueurs de temporalité : <i>à présent, dans un premier temps, ensuite</i> . Marqueurs de modalité : <i>principalement</i> , et l'adjectif <i>présent</i> pour qualifier le nom de DOCUMENT.					
Paramètres énonciatifs	Cette routine s'adresse implicitement au lecteur pour annoncer l'objectif d'étude. Elle relève donc du dialogisme interlocutif .					
Position textuelle	Dans le résumé, l'introduction, le corps de l'article et la conclusion.					
Fonctions	Fonction principale : annonce de l'objectif global ou partiel d'étude ; fonctions secondaires : rappel de l'objectif d'étude					

Tableau 5-7 : Routine N° 5 d'annonce de l'objectif/l'objet d'étude

Routine N° 6 d'annonce de l'objectif/l'objet d'étude :

Description sémantique	<L' 'auteur' 'se focalise sur' un thème (dans le 'discours en cours')>					
Exemples	<i>Dans cet article, nous nous intéressons à...</i> <i>Je me centre sur...</i> <i>On s'interroge ici sur...</i>					
Type de phrase	Phrase déclarative au présent au futur					
Structure syntaxique	[pronom AUTEUR]+(aller vouloir pouvoir choisir de)+[verbe de THÈME]+(dans+nom de DOCUMENT)					
Réalisations lexicales	<i>nous/on/jé</i>	<i>(aller/vouloir/ pouvoir/choisir de)</i>	<i>s'intéresse à/se focalise sur/ (s')interroge (sur)/se centrer/concentrer...</i>	<i>Dans cet/ce/article/papier/ la suite...</i>		SN
Éléments périphériques	Marqueurs discursifs : de déduction : <i>donc</i> ; d'exemplification : <i>notamment</i> ; de précision : <i>uniquement, essentiellement, spécifiquement</i> ; de concession : <i>cependant</i> ; de réexamen : <i>en effet</i> ; Marqueur de localisation : <i>ici</i> . Marqueur de temporalité : <i>à présent</i> . Marqueur de modalité : <i>principalement, plutôt</i> . Marqueurs d'intensité : <i>davantage, particulièrement, plus particulièrement</i> .					
Paramètres énonciatifs	Cette routine s'adresse implicitement au lecteur pour annoncer l'objectif d'étude. Elle relève donc du dialogisme interlocutif .					
Position textuelle	Dans le résumé, l'introduction, le corps de l'article, la note					
Fonctions	Fonction principale : annonce de l'objectif d'étude ; fonctions secondaires : écartement, focus ou précision de l'objectif d'étude.					

Tableau 5-8 : Routine N° 6 d'annonce de l'objectif/l'objet d'étude

5.3.2.1. Caractéristiques syntaxiques, lexicales et sémantiques de ces routines

Les routines d'annonce de l'objectif/l'objet d'étude partagent avec les routines de planification une partie d'étiquettes conceptuelles comme 'auteur', 'traite(r)', 'partie textuelle', mais ont aussi recours à d'autres concepts plus typiques : 'but', 'discours en cours', 'étude', 'vise(r) à', 'communique(r)', 'se focalise(r) sur'. L'étiquette 'but' représente la

visée discursive du texte. Il s'agit de l'objectif ou l'objet central d'étude que l'auteur veut aborder dans son texte. Elle n'apparaît que dans la routine N° 1 d'annonce de l'objectif avec des noms de FINALITÉ comme *objectif, but*.

[18] **L'objectif de cet article est d'analyser** ces pratiques de certification de la qualité, en privilégiant l'aspect contrôle par tierce partie, et ce à l'aide d'un modèle de réputation individuelle. (Économie)

[19] **L'objectif de cet article sera de dégager** les principes de variation du sens de l'adjectif grand à partir d'une hypothèse sur ce qui constitue son identité. (Linguistique)

En parallèle, l'étiquette '**discours en cours**' fait référence au discours qui est en train de se produire. Cette étiquette renvoie aux noms qui désignent les supports de communication. Dans le cas des écrits scientifiques, il désigne davantage des noms de DOCUMENTS écrits comme : *article, texte, papier, travail, etc.* Ils sont le produit des actes de communications scientifiques et véhiculent les connaissances scientifiques que l'auteur veut transmettre.

Les étiquettes verbales '**vise(r) à**', '**communique(r)**', '**se focalise(r) sur**' peuvent toutes prendre '**discours en cours**' ou '**auteur**' comme agent d'action. '**Vise(r) à**' est réalisé souvent par les verbes de FINALITÉ. Il s'agit d'une sous-classe des verbes d'ANCRAGE_CONSTRUCTION qui servent à construire un modèle d'analyse dans la recherche comme formuler les problématiques, les hypothèses, former des concepts, donner des définitions, des références et mener des réflexions, comme *adopter, concevoir, définir, ancrer, postuler, contraindre, centrer...* La famille de FINALITÉ concerne spécifiquement les actions mises en œuvre pour atteindre une finalité explicite d'une recherche, comme *s'attacher à, (se) proposer, viser à, s'efforcer de, tenter de* (exemple 20), *chercher à*, ou les collocations hors du LST : *se donner pour objectif de* (exemple 21), *avoir pour ambition de, avoir pour but de, être une tentative pour*.

[20] **Ce texte tente de saisir** les enjeux musicaux, sociaux et mémoriels de cette pratique vocale, dont elle semble être à la fois le symptôme et la revendication. (Anthropologie)

[21] **Le présent article se donne précisément pour objectif de déterminer** si le pluralisme de l'information est, dans les faits, favorisé ou non avec la multiplication des sites web. (Science de l'informatique)

'**Communique(r)**' signifie une action de communication précise de l'auteur. Dans les routines d'annonce, cette action est effectuée par les verbes de propositions comme *proposer* (exemple 22).

[22] **Cet article propose une analyse** empirique des transformations de l'espace régional européen du point de vue de l'agglomération des activités économiques et de recherche, et des spécialisations sectorielles dans les régions dites « centrales » et « périphériques ». (Géographie)

Le concept '**se focalise sur**' signifie « avoir pour objectif de faire quelque chose », réalisé par les verbes de thème relevant de la catégorie d'ANCRAGE_CONSTRUCTION. Les verbes de thème les plus fréquents dans les routines sont *s'intéresser* (*s'interroger*), *se concentrer*, (*se centrer*), *se focaliser*.

[23] **Cet article s'intéresse à** l'émergence, à l'institutionnalisation et à la transformation de la « formule magique », cette répartition des postes (2 + 2 + 2 + 1) entre les quatre principaux partis politiques, qui a prévalu au sein du gouvernement suisse entre 1959 et 2003. (Sciences politiques)

[24] **Nous nous centrons dans cette étude sur** les comportements non verbaux des élèves en période de jeu libre, sans prendre en compte ceux du professeur des écoles qui feront l'objet d'une prochaine publication. (Science de l'éducation)

Enfin, le concept '**étude**' renvoie aux observations issues des activités cognitives dans le domaine scientifique, comme les noms classés régulièrement dans deux sous-classes des noms PROCESSUS_COGNITIF qui permettent de « structurer [ou] analyser les observables et/ou objets construits » dans les écrits scientifiques (LST) : la classe d'EXAMEN comme *analyse* (exemple 22), *étude*, celle d'INTERPRÉTATION comme *interprétation* (exemple 25), *revue de littérature*.

[25] **L'article propose une interprétation** renouvelée de la politique scolaire et du développement de la scolarisation post-obligatoire sous la Quatrième République. (Histoire)

Les routines d'annonce d'objectif/objet d'étude sont la plupart du temps des phrases déclaratives au présent ou au futur, mais elles peuvent avoir quelques réalisations au passé composé qui exercent plutôt une fonction de rappel ou de récapitulation de l'objectif d'étude (exemple 26). Dans ce cas, une telle routine se situe plutôt dans le corps ou la conclusion d'un article académique.

[26] **Dans le corps de cet article, nous avons tenté de montrer que** la police en tant que corps autonome était largement en mesure de fabriquer de l'ordre social en dehors de toute injonction des pouvoirs publics. (Sciences politiques)

Au niveau syntaxique, nous trouvons certaines occurrences dans la position subordonnée ou dans la structure clivée comme montre les exemples 27-29 :

[27] La principale rectification à opérer s'avère à la fois simple, en référence aux données empiriques, et compliquées, en référence à des arrière-plans

normatifs **que nous tenterons ensuite de mettre en lumière.** (Science de l'éducation)

[28] C'est à la mise en place de cette configuration **que nous nous intéressons ici.** (Linguistique)

[29] C'est ce niveau de représentation, rarement envisagé dans les travaux portant sur le texte, en compréhension comme en production (Coirier, Gaonac'h et Passerault, 1996 ; Fayol, 1997), **que nous proposons d'étudier dans cet article.** (Psychologie)

Ces deux structures syntaxiques mettent en relief l'objectif/l'objet d'étude. La position subordonnée attribuée aux routines d'annonce de l'objectif/l'objet le statut « complément du nom antécédent ». Il s'agit ici d'une formulation très courante dans les écrits scientifiques qui permet à la fois d'annoncer le contenu suivant, et d'intégrer de façon plus flexible cette annonce dans le processus d'argumentation du texte scientifique. Dans l'exemple 27, *des arrièrè-plans normatifs* est donc été mis en premier plan, alors que l'annonce est mise en second plan. Cette manière qui diminue le poids de l'annonce permet de focaliser l'attention du lecteur sur le contenu propositionnel tout en guidant sa lecture. Par contre, la structure clivée (28, 29) met en évidence directement l'objectif/l'objet à annoncer.

5.3.2.2. *Éléments périphériques et paramètres énonciatifs*

Ces routines relèvent de la dimension interlocutive implicite. Elles comportent ainsi plusieurs déictiques : l'article défini (*le, l'*) ou le déterminant démonstratif (*ce, cet*), le déterminant possessif (*mon, notre*), le pronom auteur (*je, nous, on*) qui sont mis en scène en tant que scripteur responsable des écrits scientifiques.

La participation des marqueurs discursifs est plutôt productive dans l'ensemble des routines d'annonce de l'objectif/l'objet. Nous pouvons relever des marqueurs de CLARIFICATION (*précisément, plus précisément, concrètement*), de DÉDUCTION (*donc, ainsi*), de concession (*en revanche*), etc. Ces marqueurs discursifs contribuent à la construction de la progression discursive avec la routine d'annonce. Il existe aussi des adverbes de MANIÈRE qui marque la façon de faire une activité ou processus scientifique (*explicitement, résolument, justement*), d'intensité (*plutôt, plus particulièrement*) et de temporalité (*à présent*).

[30] **L'objectif de cet article est de montrer concrètement**, à partir du témoignage direct d'un fermier du Hainaut autrichien, la diffusion et l'application des méthodes culturelles flamandes dans les anciens Pays-Bas méridionaux au xviiiè siècle. (Histoire)

[32] **Dans cet article, nous nous intéresserons donc plus spécifiquement** à l'émergence de nouvelles modalités d'écriture et de publication de la science, développées dans le cadre d'une revue interactive, ACP, de manière

à étudier leur portée sur les modalités de production et de validation des connaissances en aérologie. (Science de l'information)

[33] **Cet article s'intéressera plus particulièrement aux** controverses déclenchées par la prétention de cet homme d'affaires à s'investir simultanément dans les sphères de l'économie, de la science et de la politique. (Sciences politiques)

En parallèle, le marqueur d'ESPACE_LOCALISATION (*ici*) est très fréquent dans les routines d'annonce. Il peut désigner « dans ce discours » ou « dans la partie particulière de ce discours », permettant de préciser la préoccupation de l'auteur dans le discours en cours. Plus précisément, sur le plan sémantique, son emploi précise l'objectif/l'objet en question. Il s'agit ici d'un objectif plus détaillé qui pourrait ne pas être l'objectif global d'étude, mais peut-être un but partiel. Sur le plan énonciatif, *ici* est un indice d'énonciation et sert à situer l'objectif par rapport aux autres contenus dans le texte.

[34] **L'objectif est ici de voir** comment se structurent scalairement des surfaces délimitées par le réseau urbain à différentes échelles. (Géographie)

Tous ces marqueurs se répètent régulièrement dans la routine, assurant la cohésion et la cohérence du discours scientifique.

5.3.2.3. *Fonctionnement rhétorique*

La fonction principale de cette routine a pour intention de signaler au lecteur l'objectif de l'étude. Si nous revenons maintenant au modèle CARS du Swales (1990, p. 141) (*cf.* figure 5-1 dans la section 5.3.1.3), nous observons que le troisième mouvement « occuper la niche » de ce modèle consiste à montrer au lecteur l'objectif d'étude (le thème de recherche) et le plan d'un article académique. L'étape 1 « présenter les objectifs » (Outlining purposes) ou « annoncer la recherche en cours » (announcing present research) occupe une place fondamentale, voir obligatoire (*ibid.*, p. 159). « présenter les objectifs » permet à l'auteur d'indiquer ses objectifs principaux, « annoncer la recherche en cours » décrit les caractéristiques principales de sa recherche (*ibid.*). Dans notre étude, la fonction d'annonce de l'objectif/l'objet d'étude est la fonction principale de ces routines. Elle vise à faire part au lecteur, dès le début de l'article, de la finalité scientifique de la recherche, ce qui lui permet d'avoir tout de suite une idée claire du thème principal du texte.

En parallèle, l'auteur peut mettre en place d'autres stratégies en variant l'emploi de ces routines.

- Utiliser cette routine dans le corps ou la conclusion du texte pour une fonction de rappel d'une information importante (dans notre cas, l'objectif d'étude) ou une fonction de précision d'un objectif local (avec l'emploi du déictique *ici*,

exemple 34) ;

[35] **Notre objectif principal consistait à observer** le poids de la connaissance du nom des lettres dans la construction précoce des connexions oral - écrit à l'œuvre dans une tâche d'écriture. (Science de l'éducation) (dans la discussion)

[36] **Notre objectif était de tester** la validité de ce modèle en fonction d'un autre contenu ; nous pensons l'avoir atteint. (Psychologie) (dans la conclusion)

- Utiliser cette routine dans une phrase négative avec le marqueur de CONCESSION *mais*, pour éviter la mauvaise anticipation du lecteur (exemple 37) ou de préciser l'objectif d'étude en question par rapport aux autres informations (38).

[37] **L'objectif de cet article n'est pas d'exposer** l'intégralité du processus de construction sociale de cette politique publique, **mais bien de comprendre** les modalités de pénétration de la logique gestionnaire dans le dispositif de préservation de la nature. (Sciences politiques)

[38] À travers la conceptualisation de la relation pédagogique que nous proposons ici, **notre but n'est pas de montrer** une "bonne" manière d'établir une relation pédagogique, **mais de rendre compte de** multiples façons d'entrer en relation avec un ou des élèves. (Science de l'éducation)

Nous remarquons, outre la fonction principale d'annonce de l'objectif/objet global ou partiel du discours, que pour les occurrences contenant le marqueur *ici*, la routine peut remplir aussi une deuxième fonction de focus (sous-catégorie des routines de topicalisation que nous allons détailler ultérieurement dans le chapitre 7) qui souligne le thème principal ou local en question.

En outre, l'emploi négatif des routines écarte le thème, considéré comme non essentiel, qui pourrait focaliser à tort l'attention du lecteur. Dans l'exemple 39 en bas, l'auteur annonce au lecteur un thème qu'il ne va pas traiter, afin qu'il se concentre sur le thème central du discours.

[39] **Nous ne nous intéressons pas, ici,** à la question des fonds propres ni à celle des réserves. (Économique)

En résumé, les routines d'annonce de l'objectif/l'objet d'étude peuvent remplir :

1. Une fonction principale : annonce de l'objectif d'étude du discours en cours (au présent et au futur).
2. Des fonctions secondaires comme le rappel de l'objectif d'étude du discours en cours (au passé, dans le corps ou dans la conclusion d'un article), la précision de l'objectif ou le thème à traiter en cours (dans les phrases négatives ou de concession).

5.3.3. Routines de navigation textuelle

La fonction de navigation textuelle vise soit au renvoi intratextuel (*comme nous l'avons vu auparavant...*), soit au renvoi intertextuel (*voir aussi l'article de qn*). Les termes de « renvois intratextuels » et « renvois intertextuels » sont issus du travail de Tutin (2018 a, p. 86–87) portant sur les expressions polylexicales transdisciplinaires dans les articles de recherche en sciences humaines. Ils ont été observés comme des fonctions textuelles avec la structuration du texte et de la topicalisation dans le travail de Tutin (2018 a, p. 86–87). Au lieu de dire que ces deux fonctions sont simplement textuelles, nous les classons comme une sous-fonction « navigation textuelle » de la structuration textuelle.

Nous préférons le terme « navigation textuelle » qui fait allusion au « voyage » du lecteur dans le texte en cours ou vers d'autres textes. Par conséquent, toutes les routines de fonction de navigation textuelle doivent endosser la mission de renvoyer le lecteur vers un lieu textuel, soit pour rappeler un point déjà évoqué précédemment ou qui va être évoqué dans la suite du discours, soit pour renvoyer le lecteur aux autres discours. Selon la nature du renvoi, nous subdivisons encore la fonction de navigation textuelle en « renvoi intratextuel », visant à faire voyager le lecteur dans le discours en cours, et en par « renvoi intertextuel » invitant à se référer aux autres discours⁹⁰. À l'intérieur de chaque sous-classe, il faut encore préciser quelques points flous.

La navigation intratextuelle est proche de plusieurs fonctions décrites dans les modèles métadiscursifs (Hyland, 2005, p. 51 ; Ädel, 2010, p. 86) : les marqueurs endophoriques, le rappel, la prévision. Hyland et Ädel ont des visions différentes pour ces fonctions. Ädel (2010, p. 86) définit la fonction endophorique comme des marqueurs visant à indiquer un endroit précis du texte pour expliquer le point courant d'un texte (*From these map points, we see that ... as we can see in (5)*). Le rappel (*X above; As I mentioned earlier*) et l'anticipation (*we'll be coming to that*) sont deux fonctions indépendantes de la fonction endophorique. La première pointe vers l'arrière dans le discours, rappelle au lecteur les points déjà discutés. La deuxième prévient le lecteur de ce qui va se passer après, dans le discours. De son côté, Hyland (2005, p. 51) regroupe le rappel et la prévision dans les marqueurs endophoriques sans faire de différence sur le lieu textuel (avant, arrière ou lieu précis). Notre sous-catégorie « renvoi intratextuel » se rapproche de la vision d'Hyland, qui regroupe toutes les routines renvoyant à

⁹⁰Nous préférons ici le terme les « autres discours » au lieu du « discours d'autrui », parce que l'auteur peut se référer à ses propres discours qui ne sont pas le discours en cours. Il s'agit de l'autoréférence qui est une pratique assez courante dans le domaine scientifique.

n'importe quelle partie textuelle.

Ici, il ne faut pas confondre la fonction d'anticipation avec la fonction d'annonce de l'objectif/l'objet définie plus haut (*cf.* section 5.3.2). D'abord, leurs portées textuelles ne sont pas identiques. L'annonce de l'objet/objectif d'étude dispose d'une portée globale sur le texte entier ou sur une section, alors que la navigation intratextuelle annonce le thème suivant et a donc une portée plutôt limitée, qui ne touche pas le discours dans sa globalité. Ensuite, la position textuelle où se trouvent les deux types de routines est aussi différente : les routines d'anticipation apparaissent dans les parties centrales de l'article, tandis que les routines d'annonce apparaissent souvent dans l'introduction ou le début d'une section textuelle.

Quand le renvoi intratextuel préoccupe la navigation au sein du discours en cours, le renvoi intertextuel se charge de la mission d'envoyer le lecteur à d'autres discours qui présentent en tant que références de preuve ou sources d'informations. La fonction de renvoi intertextuel se distingue de la fonction de preuve (*(to) cite X*) de la catégorie de Hyland (2005, p. 51) pour informer la source d'un positionnement. La fonction de preuve est définie comme « une représentation métalinguistique d'une idée provenant de quelqu'un autre (Thomas and Hawes, 1994, p. 129) qui guide l'interprétation du lecteur et établit un commentaire de l'auteur sur cette ressource » (notre traduction) (2005, p. 51). Autrement dit, elle est plus proche de la citation et permet d'indiquer l'attribution de la responsabilité d'une idée, ce qui offre des preuves pour l'argumentation. Dans notre étude, le renvoi intertextuel est un terme pour désigner la fonction des routines qui renvoie explicitement le lecteur à un autre discours sans entrer dans les détails de ce discours. Cette fonction est donc plutôt un renvoi qu'une référence. Pour mieux expliquer notre positionnement, nous nous référons ici à un type de modes de référence au discours d'autrui dans le discours scientifique proposé par Boch et Grossmann (2002, p. 43) appelé « l'évocation ». L'évocation consiste en une façon d'intégrer le dit d'autrui : elle « [met] en arrière-plan les connaissances partagées, ou les éléments non essentiels au propos tout en inscrivant la recherche dans un espace épistémique identifiable » (*ibid.*), comme l'exemple (proposition soulignée) donné par les deux linguistes :

En linguistique, on peut citer les études qui ont marqué la fin des années quatre-vingt, Cruse (1986), Winston, Chaffin et Hermann (1987), Evens (1988), Iris, Litowitz et Evens (1988), Chaffin et Hermann (1988).

C'est dans cette idée de ne pas entrer dans les détails des références, mais les mettre en arrière-plan que le renvoi intertextuel est utilisé. Le renvoi intertextuel peut ainsi être défini comme ce qui permet à l'auteur de renvoyer le lecteur aux discours d'autrui pour une meilleure

interprétation du discours en production. Il permet au lecteur de situer le discours en question par rapport aux discours déjà produits dans les domaines concernés.

En résumé, la navigation textuelle comporte deux orientations complémentaires : vers les matériels propositionnels au sein du discours en cours et vers d'autres discours. En fonction de nos définitions, nous avons identifié 8 routines, dont 6 routines de renvoi intratextuel et 2 routines de renvoi intertextuel. Voici la modélisation de ces routines.

Routine N° 1 de navigation textuelle (renvoi intratextuel) :

Description sémantique	<Une information est 'présentée' dans un 'lieu textuel'> ⁹¹				
Exemples	<i>Les résultats sont exposés dans le tableau 1 ci-dessous.</i> <i>Ces 28 catégories sont présentées en annexe I.</i> <i>[3] Leur trajectoire et profil respectifs seront présentés dans la section IV.</i>				
Type de phrase	Phrase déclarative au présent ou au futur à la voix passive, syntagme adjectif				
Structure syntaxique	(X)+(être)+{[verbe de DESCRIPTION]}+dans/en+{[nom de SECTION]}+({[chiffre]}{[adjectif d'ÉNUMÉRATION]})				
Réalisations lexicales	(X)	(est/sont)	présenté)/exposé/décrit/rep orté...	dans/en tableau/annexe/section/ graphique/page	(suivant/précédent/1, 2/a, b...)
Éléments périphériques	Pas d'élément périphérique.				
Paramètres énonciatifs	Cette routine s'adresse implicitement au lecteur pour indiquer le lieu textuel ou péri-textuel d'une information. Elle relève donc du dialogisme interlocutif .				
Position textuelle	Dans le corps principal du discours.				
Fonctions	Renvoi intratextuel, évidentialité.				

Tableau 5-9 : Routine N° 1 de navigation textuelle

⁹¹ Cette routine a une configuration syntaxique très proche de la structure 2 de la routine N°01 de planification (cf. section 5.3.1), mais elle se trouve souvent dans le corps principal du discours au lieu de l'introduction. Les variations lexicales présentent aussi des différences, par exemple, la routine de planification utilise plus des noms de SECTION comme *section*, *partie*, tandis que la routine de navigation montre une préférence pour des noms de GRAPHIQUE comme *tableau*, *figure*, des noms de section comme *annexe*, *introduction*.

Routine N° 2 de navigation textuelle (renvoi intratextuel) :

Description sémantique	< L'‘auteur’ renvoie le lecteur à ‘voir’ une information (dans un ‘lieu textuel’) >						
Exemples	<i>comme nous allons le voir ; on va le voir ; nous présentons en annexe... Comme nous l'avons vu dans la section précédente ; On l'a vu</i>						
Type de phrase	Phrase déclarative						
Structure syntaxique	$(comme) + [\text{pronom AUTEUR}] + (le) + \{ [\text{verbe de CONSTAT}] \} + (dans/en + \{ [\text{nom de SECTION}] \} + \{ [\text{chiffre}] \} + \{ [\text{nom de GRAPHIQUE}] \} + \{ [\text{adjectif d' ÉNUMÉRATION}] \})$						
Réalisations lexicales	<i>(comme)</i>	<i>nous /on/ je</i>	<i>(le/l')</i>	<i>(avons/a/allons/va /venons de/vient de)</i>	<i>vu/noté/constaté/mentionné/dit/voir/noter/constater/verrons/verra ...</i>	<i>(dans la (le)/en section/annexe/liste)</i>	<i>précédemment/ci-dessous, plus haut...</i>
Éléments périphériques	Marqueurs discursifs : de concession : <i>certes, mais, pourtant, en revanche</i> ; d'addition : <i>par ailleurs</i> ; d'exemplification : <i>par exemple</i> . Marqueurs de localisation : <i>plus haut, plus loin, ci-dessus, à la fin de la section, en première partie, en annexe, sur le tableau...</i> Marqueurs de temporalité : <i>déjà, au départ, préalablement, précédemment, auparavant, à présent...</i>						
Paramètres énonciatifs	Cette routine s'adresse implicitement au lecteur pour indiquer le lieu textuel ou péri-textuel d'une information. Elle relève donc du dialogisme interlocutif .						
Position textuelle	Dans le corps principal du discours, parfois entre parenthèses.						
Fonctions	Renvoi intratextuel, rappel, prévision, constat						

Tableau 5-10 : Routine N° 2 de navigation textuelle

Routine N° 3 de navigation textuelle (renvoi intratextuel) :

Description sémantique	< Un ‘lieu textuel’ ‘révèle’ une information >				
Exemples	<i>Comme l'indique le tableau 5</i>				
Type de phrase	Phrase déclarative				
Structure syntaxique	$comme + le + [\text{verbe de RÉVÉLATION}] + [\text{nom de GRAPHIQUE}]$				
Réalisations lexicales	<i>comme</i>	<i>le (l')</i>	<i>indique/montre</i>	<i>le tableau/la figure</i>	<i>1, 2, 3...</i>
Éléments périphériques	Marqueurs discursifs de déduction <i>ainsi</i> (1 occurrence).				
Paramètres énonciatifs	Cette routine s'adresse implicitement au lecteur pour indiquer le lieu péri-textuel d'une information. Elle relève donc du dialogisme interlocutif .				
Position textuelle	Dans le corps principal du discours.				
Fonctions	Renvoi intratextuel, évidence				

Tableau 5-11 : Routine N° 3 de navigation textuelle

Routine N° 4 de navigation textuelle (renvoi intratextuel) :

Description sémantique	< L' 'auteur' 'anticipe ou rappelle' un point ou une information >					
Exemples	<i>Je reviendrai sur ce point.</i> <i>Nous y reviendrons.</i> <i>Nous voudrions, avant de conclure, revenir sur...</i> <i>Je passe maintenant à...</i>					
Type de phrase	Phrase déclarative					
Structure syntaxique	[pronom AUTEUR]+ (y pouvoir/aller)+[verbe de PASSAGE]+ (dans en+[nom de SECTION])					
Réalisations lexicales	<i>Nous/on</i> <i>/je</i>	<i>(peut/pouvons</i> <i>/allons)</i>	<i>(y)</i>	<i>passer</i> <i>(maintenant/à présent) à/</i> <i>reviendrons (ai/a) sur</i>	<i>SN/</i> <i>(ce point)</i>	<i>(dans</i> <i>la</i> <i>discussion/conclusion</i> <i>/section 1, 2, 3...)</i>
Éléments périphériques	Marqueurs discursifs : de concession : <i>toutefois</i> , de finalité : <i>pour ce faire</i> ; de temporalité : <i>maintenant, tout d'abord, d'abord, brièvement</i> ; Marqueurs de localisation : <i>plus loin</i> .					
Paramètres énonciatifs	Cette routine s'adresse implicitement au lecteur pour rappeler une information dans un lieu textuel. Elle relève donc du dialogisme interlocutif .					
Position textuelle	Dans le corps principal du discours, parfois entre parenthèses.					
Fonctions	Renvoi intratextuel, prévision un point ou le retour d'un point.					

Tableau 5-12 : Routine N° 4 de navigation textuelle

Routine N° 5 de navigation textuelle (renvoi intratextuel) :

Description sémantique	< L'auteur juge qu'il est 'nécessaire' de 'rappeler' une information >			
Exemples	<i>Il convient de rappeler...</i> <i>Il faut revenir sur...</i>			
Type de phrase	Phrase impersonnelle			
Structure syntaxique	[il]+ [faut convient de] + [verbe de PASSAGE] + $\left\{ \begin{array}{l} \text{SN} \\ \text{complétive que (le fait que)} \end{array} \right\}$			
Réalisations lexicales	<i>Il</i>	<i>faut/convient de</i>	<i>rappeler/revenir sur (à)</i>	<i>SN/le fait que...</i>
Éléments périphériques	Marqueurs de localisation : <i>en conclusion</i> . Marqueurs de temporalité : <i>maintenant</i> .			
Paramètres énonciatifs	Cette routine s'adresse implicitement au lecteur pour rappeler une information. Elle relève donc du dialogisme interlocutif .			
Position textuelle	Dans le corps principal du discours, souvent entre parenthèses.			
Fonctions	Renvoi intratextuel. Mise en évidence.			

Tableau 5-13 : Routine N° 5 de navigation textuelle

Routine N° 8 de navigation textuelle (renvoi intertextuel) :

Description sémantique	< L' 'auteur' 'se réfère' à un 'autre discours' pour une information plus précise > ⁹²				
Exemples	<i>On peut aussi mentionner la récente thèse en sociologie de Juliette Verdière...</i> <i>Je renvoie le lecteur à ce livre...</i> <i>On peut se référer notamment aux travaux de S. Charbonneau (2006)...</i>				
Type de phrase	Phrase déclarative				
Structure syntaxique	[pronom AUTEUR]+ (pouvoir)+ [verbe de RÉFÉRENCE]+ ([nom de DOCUMENT] + (de qqn) quelqu'un)				
Réalisations lexicales	<i>On/nous</i>	<i>(peut/ pouvons)</i>	<i>cite (r) (ons)/se réfère (r) (ons)/note (r) (ons)/mentionne (r) (ons)/se (nous) reporte (r) (ons) à...</i>	<i>(l'article/l'étude/l e travail...)</i>	<i>(de) quelqu'un</i>
Éléments périphériques	Marqueurs discursifs ; de déduction : <i>ainsi</i> ; d'exemplification : <i>notamment, par exemple, en particulier</i> ; d'addition : <i>également, aussi</i> ; de rapport : <i>respectivement</i> .				
Paramètres énonciatifs	Cette routine s'adresse explicitement au lecteur pour indiquer une référence une source d'une information dans un autre discours. Elle relève donc du dialogisme interdiscursif .				
Position textuelle	Dans le corps principal du discours.				
Fonctions	Renvoi intertextuel.				

Tableau 5-16 : Routine N° 8 de navigation textuelle

5.3.3.1. Caractéristiques syntaxiques, lexicales et sémantiques de ces routines

En observant les tableaux ci-dessus, nous pouvons constater que les routines de navigation textuelle possèdent des variations syntaxiques et lexicales assez importantes. En revanche, les champs sémantiques sont assez homogènes dans chaque sous-fonction de « renvoi intratextuel » et de « renvoi intertextuel ».

Les routines de renvoi intratextuel mobilisent des champs sémantiques comme 'présenté (er)', 'voir', 'révèle (er)' qui se réalisent par des verbes de DESCRIPTION (*décrire, présenter*), des verbes de CONSTAT (*voir*), des verbes de DIFFUSION (*dire, illustrer, préciser*), verbe de RÉVÉLATION (*indiquer*). En parallèle, les noms de SECTION (*annexe, introduction*) et de GRAPHIQUE (*tableau, figure*) sont également employés dans aussi les routines de renvoi intratextuel. Voici quelques exemples pour montrer leurs points communs.

[40] Les estimations de ce modèle augmenté **sont reportées dans le tableau 4**. (Économie)

[41] **Comme nous l'avons présenté en introduction**, la notion de contrainte a suscité de nombreux travaux, aussi bien en psychologie qu'en IA. (Psychologie)

[42] Dans l'ensemble, **comme le montre le tableau en annexe 1**, le terme même de préposition apparaît rarement, seul, au premier niveau dans ces tables, sauf dans la grammaire de Wagner et Pinchon, où les pages sur le sujet représentent presque un dixième du total (63/687). (Linguistique)

⁹² Nous avons identifié aussi une occurrence très spéciale qui interpelle directement le lecteur : *le lecteur peut se référer à*.

[43] Une importante partie du travail consiste (ensuite) à regrouper les quelque 1500 professions observées en 10 catégories socioprofessionnelles (**voir tableaux en annexe 2**). (Science de l'éducation)

Il faut noter que l'emploi des noms de GRAPHIQUE a souvent une fonction un peu différente de celui des noms de SECTION. Quand une routine de navigation fait référence à une partie textuelle, elle a souvent pour objectif de 1) rappeler une information déjà évoquée dans cette partie textuelle (exemple 41), et 2) d'informer le lecteur de l'emplacement textuel d'une information souvent éloignée (exemple 42, 43). Par contre, une routine invitant à regarder immédiatement un tableau ou une figure numérotés dans le corps du texte, manifeste soit une « fonction énonciative seconde », soit une « fonction évidentielle » (Grossmann & Tutin, 2010, p. 6) qui permet de montrer la preuve ou l'illustration pour une argumentation (exemple 40).

Nous avons aussi remarqué quelques singularités dans certaines routines de renvoi intratextuel. La routine N° 2 < L' **'auteur'** renvoie le lecteur à **'voir'** une information (dans un **'lieu textuel'**)> et la routine N° 4 < L' **'auteur'** **'anticipe ou rappelle'** un point ou une information > présentent dans notre corpus plus de variétés syntaxiques, lexicales, voire fonctionnelles.

La routine < L' **'auteur'** renvoie le lecteur à **'voir'** une information (dans un **'lieu textuel'**)> possède des configurations syntaxiques très complexes. Nous identifions :

3. des propositions simples qui ne comportent pas de conjonction *comme*, ni de noms marquant les parties textuelles comme *on l'a vu, nous allons le voir* ;
4. des propositions plus complexes qui intègrent plus d'éléments facultatifs comme *comme nous l'avons présenté en introduction, comme nous l'avons vu dans la section précédente*.

Même si ces variations syntaxiques remplissent toutes une fonction de navigation dans le texte, les propositions simples réalisent une navigation sans préciser où l'auteur a donné ou va donner une telle information (représenté par le pronom neutre *le*). Au contraire, les propositions plus complexes montrent une direction textuelle très claire (*en introduction, en annexe, etc.*). Que la routine indique un endroit bien précis ou non, sa présence est indispensable pour assurer la cohésion du discours scientifique.

À l'instar de la routine supra, la routine < L' **'auteur'** **'anticipe ou rappelle'** un point ou une information > contribue à la cohésion du discours scientifique. L'étiquette conceptuelle **'anticipe (r) ou rappelle (r)'** correspond à deux éléments divergents : **'anticiper'** prépare le lecteur au point à aborder (exemple 44) ou sur lequel il reviendra plus tard (45, 46), **'rappeler'**

prévient tout de suite du sujet sur lequel il va revenir immédiatement (47). Nous pouvons également les appeler « éléments transitoires » qui assurent le passage entre des paragraphes et des idées.

[44] (3) □ **Nous passons maintenant à** la description des flux sur le marché du travail. (Économie)

[45] **Je reviendrai plus loin sur** cette question. (Anthropologie)

[46] L'hétérogénéité des emplois et des valeurs supportées par tirer sur permet d'aborder une série de questions d'ordre théorique, **sur lesquelles nous reviendrons en conclusion.** (Linguistique)

[47] Pour résumer ces travaux, **nous pouvons revenir au** modèle de la Section 4.2.1. (Économie)

Si les routines de renvoi intratextuel sont nombreuses et plus complexes dans le corpus, les routines de renvoi intertextuel s'avèrent plus homogènes. Deux routines ont été retenues qui représentent deux façons classiques de renvoyer le lecteur à un '**autre discours**' :

– Utiliser des phrases infinitives injonctives

[48] [9] **Voir l'article de** Liu Xiaojing, op. (Sociologie)

[49] [14] Pour une définition complète de l'érosion et de la dilatation, **voir le chapitre « Outils de base » dans** Voiron (1995, p.32-35). (Géographie)

[50] 2. **Se référer à** OGAWA et FUJITA [1980] pour une discussion approfondie sur ce point. (Économie)

– Utiliser des phrases déclaratives

[51] **On peut cependant mentionner les récents travaux d'**Entman (2010) centrés sur le traitement médiatique du réchauffement climatique. (Science de l'informatique)

[52] Pour une présentation de l'histoire de Westminster Abbey, **on se reportera à l'ouvrage** pratique et bien illustré **de** John Field, Kingdom, Power and Glory. (Histoire)

Le discours d'autrui est exprimé par 1) des noms de DOCUMENT (*l'article, le travail, l'ouvrage* de quelqu'un (48, 51, 52) ; 2) directement le nom d'auteur cité dans la mesure où le nom manifeste une fonction métonymique (50), 3) des noms de SECTION dans un autre discours (49). Ces constructions propositionnelles dans les articles de recherches reflètent la pratique courante de la communauté scientifique pour se référer aux autres discours sans entrer dans les détails.

D'un point de vue global, les routines de navigation textuelles appellent souvent des verbes de navigation (*voir, se reporter, se référer, passer, revenir, etc.*), des noms de SECTION

(*note, annexe...*) ou de DOCUMENT (*article, papier...*). D'ailleurs, l'ajout des éléments périphériques comme des marqueurs discursifs est aussi remarqué pour certaines routines. Nous précisons ce point dans la section suivante.

5.3.3.2. *Éléments périphériques et paramètres énonciatifs*

Toutes les routines ne peuvent pas accepter des marqueurs discursifs ou de localisation. En général, les routines N° 2 et N° 4 de renvoi intratextuel et les deux routines de renvoi intertextuel ont tendance à utiliser plus de marqueurs discursifs (*cependant* dans l'exemple 51), de marqueurs temporels (*maintenant* 44), ou de LOCALISATION (*plus loin* 45), de TOPICALISATION (*sur ce point* dans l'exemple 50). D'ailleurs, les routines de renvoi intertextuel peuvent intégrer des syntagmes prépositionnels indiquant la raison de se référer comme *pour une définition complète, voir...* (49), *pour une présentation..., on se reportera à l'ouvrage de...* (52). La cooccurrence de ces éléments renforce la cohérence du discours avec des routines de navigation textuelle dans le discours.

En ce qui concerne les dimensions dialogiques, les routines de navigation sont capables d'entrer dans plusieurs dimensions. Elles peuvent être incluses d'abord dans le cadre du dialogisme interlocutif, privilégiant le besoin du lecteur de situer les informations dans le discours de l'auteur, et puis dans le dialogisme interdiscursif par le renvoi vers le discours d'autrui. Les routines manifestent de façon implicite ou explicite la dimension interlocutive. Par exemple, les routines de nature injonctive (N° 5, N° 6) s'adressent directement au lecteur (48, 49, 50). La routine N° 7 peut renvoyer explicitement au lecteur en employant le mot *lecteur*, comme l'illustre l'exemple ci-dessous. Ces routines manifestent explicitement une orientation vers l'interlocuteur et nous parlons ici de « dialogisme interlocutif explicite ».

[53] **Nous** ne développerons pas ici toute la procédure relative à l'estimation par le maximum de vraisemblance exact des processus ARFIMA et **renvoyons le lecteur aux travaux de** DAHLHAUS [1989] et SOWELL [1992a]. (Anthropologie)

Outre les phrases injonctives ou le renvoi métalinguistique au lecteur, l'emploi des pronoms auteur *nous, on* inclusif (*on l'a vu, comme nous avons déjà vu*) dans certaines routines peut aussi montrer un témoignage collectif du lecteur et de l'auteur, créant ainsi un espace dialogique apparent.

Les routines de renvoi intertextuel peuvent relever à la fois du dialogisme interlocutif dans le sens où l'on favorise le suivi du discours pour le lecteur, et du dialogisme interdiscursif dans le sens où l'on fait référence aux discours des pairs. La routine N° 6 (*voir l'article de qn*)

manifeste explicitement les deux dimensions, la routine N° 7 (*on peut se référer à qn*) est explicitement interdiscursive mais implicitement interlocutive.

5.3.3.3. *Fonctionnement rhétorique*

Les écrits scientifiques, notamment des articles de recherches représentent un ensemble du produit de la mise en œuvre d'un acte argumentatif et discursif. Les routines de navigation textuelle montrent la prise en considération du lecteur potentiel pour leur besoin de suivre aisément la progression du discours. Même si l'annonce de l'objectif et le plan du discours sont établis préalablement au début du discours, la préoccupation de faciliter la lecture reste présente tout au long du discours.

5.3.4. **Routines de conclusion**

Cette fonction est envisagée comme un moyen de transition logique dans le modèle interpersonnel du métadiscours. Le modèle réflexif d'Ädel (*cf.* section 3.4.2) la considère comme un élément métatextuel. Nous rappelons que Hyland (2005, p. 50) définit la conséquence comme « formuler une conclusion (*in conclusion*) ou indiquer une déduction (*thus, therefore*), ou un contre-argument (*nevertheless, anyway*). Sa définition souligne le raisonnement logique de la conclusion sur le plan argumentatif. Cependant, Ädel (2006, p. 60) ne précise pas si la conclusion clôture une chaîne d'argumentation ou une partie textuelle, mais la définit généralement comme un élément métadiscursif qui conclut un sujet ou un thème. Nous prenons en compte les deux définitions de la conclusion, parce que, premièrement, mettre fin à une argumentation facilite la compréhension et est essentiel pour guider la lecture, et la conclusion indique intrinsèquement la conséquence logique d'un sujet ; deuxièmement, les routines de conclusion dans notre corpus mélangent les deux types de conclusions. Mais contrairement à Hyland (2005, p. 50), nous envisageons comme Ädel (2006, p. 60) la fonction de conclusion comme une fonction de structuration textuelle, non comme une transition logique. La conclusion est ainsi définie comme un acte discursif qui met fin à un sujet ou un thème dont l'auteur propose une synthèse, ou un acte scientifique qui tire une conséquence d'un fait, d'une série d'argumentation, d'une observation scientifique.

Nous remarquons que, dans notre corpus, les routines de conclusion sont relativement rares et présentent peu de variétés syntaxiques et lexicales par rapport aux autres routines de structuration. Ce phénomène est probablement dû à la présence abondante et courante des marqueurs discursifs comme *pour conclure, en conclusion, pour finir* dans le discours qui peuvent remplir facilement la fonction de conclusion, sans compter des marqueurs de synthèse

comme *en résumé, en un mot* (Tran, 2014, p. 94) qui peuvent aussi remplir cette fonction dans une certaine mesure. Deux routines ont été retenues et modélisées ainsi :

Routine N° 1 de conclusion :

Description sémantique	< L' 'auteur' 'tire une conclusion' d'un thème abordé >			
Exemples	<i>On peut conclure que...</i> <i>Nous pouvons déduire...</i>			
Type de phrase	Phrase déclarative			
Structure syntaxique	[pronom AUTEUR]+(pouvoir/tenter de)+(en)+ [verbe d'INFÉRENCE]+ que sur à			
Réalisations lexicales	<i>Nous/on</i>	<i>(pourrait/serait tenté de)</i>	<i>(en) conclure/déduire</i>	<i>que/sur/à</i>
Éléments périphériques	Des marqueurs discursifs : <i>donc, en particulier, seulement, en somme, ici, finalement, etc.</i>			
Paramètres énonciatifs	Cette routine s'adresse au lecteur pour indiquer la conclusion d'une partie textuelle ou d'une chaîne d'argumentation. Elle relève donc du dialogisme interlocutif .			
Position textuelle	Pas dans la partie introduction.			
Fonctions	Conclusion totale ou partielle.			

Tableau 5-17 : Routine N° 1 de conclusion

Routine N° 2 de conclusion :

Description sémantique	< Une 'conclusion' 'importante' est présentée >					
Exemples	<i>Un deuxième résultat important est que...</i> <i>Le premier résultat important est que...</i> <i>Le résultat principal de notre étude est de montrer...</i>					
Type de phrase	Phrase déclarative					
Structure syntaxique	Déterminant + [nom d'IMPLICATION]+[Adjectif d'IMPORTANTANCE]+ être/montre + SN/que					
Réalisations lexicales	<i>Un/le/</i>	<i>(deuxième/premier)</i>	<i>résultat</i>	<i>important/principal</i>	<i>est/montre</i>	<i>SN/que</i>
Éléments périphériques	Des marqueurs discursifs : <i>donc, ici</i>					
Paramètres énonciatifs	Cette routine s'adresse au lecteur pour indiquer la conclusion d'une partie textuelle ou d'une chaîne d'argumentation. Elle relève donc du dialogisme interlocutif .					
Position textuelle	Pas dans la partie introduction.					
Fonctions	La fonction principale est le guidage du lecteur dans la présentation d'une conclusion, la fonction seconde consiste en la mise en relief de cette conclusion.					

Tableau 5-18 : Routine N° 2 de conclusion

5.3.4.1. Caractéristiques syntaxiques, lexicales et sémantiques de ces routines

Ces deux routines possèdent ensemble quatre étiquettes conceptuelles. La première routine possède une structure syntaxique avec le pronom AUTEUR comme sujet, ce qui met en avant la présence de l'auteur en tant qu'énonciateur et chercheur en même temps. Cette routine peut formuler des conclusions ou des déductions globales ou partielles.

[54]Nous pouvons maintenant conclure notre démonstration : sur trois générations (celles qui entourent Ego), la logique de la terminologie dravidienne est isomorphe à un système de sections. (Anthropologie)

[55] **On pourrait en conclure que** les auteurs intéressés par le langage intérieur dans l'œuvre de Vygotskij considèrent le langage dans la diversité de ses fonctions surtout comme moyen de médiation des processus cognitifs. (Linguistique)

[56] **On en déduit que** la plupart des situations d'interaction sont connues et répertoriées par les conducteurs. (Psychologie)

L'étiquette '**tire (r) une conclusion**' désigne un acte scientifique, accompli par les verbes d'INFÉRENCE (*conclure, déduire*) qui renvoient à l'acte scientifique de mettre en évidence la conséquence logique d'un fait, d'une argumentation. La présence de verbe de modalité *pouvoir* aide à nuancer les formulations de l'auteur devant son lecteur en adoucissant d'abord l'affirmation et créant un espace dialogique en laissant au lecteur la liberté de confirmer ou de réfuter la conclusion (Barbet, 2012, p. 53).

Par rapport à la première routine, la deuxième routine a une fonction de conclusion expliquant plutôt des observations, des analyses réalisées. L'étiquette '**conclusion**' apparaît avec les noms d'IMPLICATION, notamment *résultat*. Simultanément, les adjectifs d'IMPORTANCE (*principal, essentiel, fondamental...*) accordent une plus haute valeur à une conclusion vis-à-vis des autres. La routine remplit ainsi une fonction secondaire de focus (sous-classe de la topicalisation, nous y reviendrons dans le chapitre 8).

[57] **Le premier résultat important est qu'**il est socialement préférable d'associer une politique de retrait des terres à une politique de prix garanti, comparativement à une politique caractérisée par un seul des deux instruments. (Économie)

[58] **Le résultat important est donc ici** la stabilité d'une distribution équilibrée des deux préférences plutôt que la variation de quatre points des opinions exprimées. (Sciences politiques)

Sur le plan syntaxique, les routines de conclusion présentent des configurations assez régulières à travers deux structures :

- 1) Routine + complétive que (ou *sur le fait que*) + phrase subordonnée (exemples 55, 56, 57, 60, 61, 62, 63, 64)
- 2) Routine + Syntagmes nominal (exemples 54, 58) ou verbal (routine N° 2 de conclusion, fréquence très rare, exemple 59)

[59] **Le résultat principal de notre étude est de montrer que**, bien que les conducteurs âgés aient un risque majeur d'être impliqués dans un accident mortel dans ces situations de traversée de voie(s), ils sont néanmoins moins dangereux pour les autres impliqués par rapport à des conducteurs plus jeunes, lorsqu'on évalue la dangerosité par le nombre d'avp. (Psychologie)

5.3.4.2. *Éléments périphériques et paramètres énonciatifs*

Si les structures ont assez peu de variantes, les éléments accessoires s'avèrent plus diversifiés, par exemple les marqueurs discursifs (*cependant, immédiatement, difficilement, donc, ainsi, en particulier, en somme, enfin, finalement*), de modalité (*seulement, sans doute*), d'espace (*ici*), de temporalité (*à nouveau, maintenant, dès lors, immédiatement*) cooccurrence de façon dynamique avec la première routine.

[60] **On peut donc conclure qu'**au cours de cette seconde période, il existe une diffusion des scolarités prolongées, principalement par l'enseignement public, et plus encore dans les cours complémentaires que dans les établissements secondaires [40]. (Histoire)

[61] **En somme, nous pouvons conclure que** lors d'une tâche portant sur la catégorie, la couleur caractéristique d'un objet ou "couleur de mémoire" (par ex., Dumaurier, 1992 ; Goldstein, 2002) est traitée sans que cela ne soit explicitement demandé. (Psychologie)

[62] **On peut sans doute en conclure que** cette série est véritablement caractérisée par une mémoire de long terme, même si son degré de mémoire longue n'est pas le plus élevé. (Économie)

Nous identifions aussi des marqueurs qui précisent la manière de conclure comme *provisoirement, légitimement, raisonnablement* et qui ne sont pas intégrés dans le LST.

[63] **Nous devons ici conclure provisoirement sur le fait que** notre corpus, pour journalistique qu'il soit, ne reflète pas la totalité des emplois existants en situation contemporaine. (Linguistique)

Des marqueurs énumératifs apparaissent plus activement dans la deuxième routine, accordant donc à la routine de conclusion une fonction secondaire d'énumération.

[64] **Le premier résultat notable est que** le taux d'investissement optimal est identique au cas non contraint : la distorsion fiscale n'a pas d'impact sur le taux d'investissement public optimal, qui dépend uniquement de la technologie et des préférences. **Le second résultat notable montre que** lorsque le gouvernement a comme objectif de maximiser la croissance de l'économie, il va choisir de sur-accumuler le capital public, et va donc réduire le taux d'actualisation qu'il annonce aux décideurs publics d'investissement. (Économie)

Ces routines s'adressent implicitement au lecteur pour donner des conclusions ou déductions faites par l'auteur. Elles relèvent donc du dialogisme interlocutif.

5.3.4.3. *Fonctionnement rhétorique*

La fonction principale de ces routines est de formuler des conclusions ou des déductions sur le sujet développé dans l'article. Dans le modèle de Swales (1990, p. 170), la conclusion pour interpréter les résultats est très présente dans les parties *Résultats, Discussion* ou

Conclusion qui peuvent être mélangées parfois. Cette fonction est essentielle dans les articles de recherche. Elle fait écho aux routines d'annonce de l'objectif/l'objet d'étude dans la mesure où elle cherche à répondre aux questions posées ou aux objectifs fixés dans l'introduction. La coopération entre l'annonce de l'objectif/l'objet et la conclusion permet à l'auteur de créer un écosystème complet du discours académique en termes argumentatifs et discursifs.

5.4. Synthèse

Dans ce chapitre, nous nous sommes focalisée sur les routines de structuration qui servent à mieux organiser le matériau propositionnel des articles de recherche. Quatre sous-fonctions avec 18 routines ont été analysées.

La fonction de structuration discursive a été redéfinie en adoptant une optique fonctionnelle basée sur les observations des données métadiscursives. Elle a pour objectif l'organisation du matériau propositionnel du discours scientifique. Nous avons élaboré un classement des routines à quatre sous-fonctions rhétoriques adapté aux données recueillies : la planification de la structure textuelle, l'annonce de l'objectif/objet d'étude, la navigation textuelle et la conclusion. À travers leurs définitions, une analyse qualitative en donnant des modélisations des routines a été proposée.

La planification de la structure discursive permet d'annoncer au lecteur le plan de l'article, ce qui permet au lecteur de comprendre au préalable la composition des sections du contenu de l'article. Nous avons retenu 2 routines dans cette sous-catégorie, qui sont largement présentes en économie, discipline dans laquelle, dont l'organisation de la structure est très affirmée (*cf.* section 8.3). La deuxième sous-catégorie vise à annoncer explicitement au lecteur l'objectif ou l'objet d'étude du discours scientifique. Cet objectif peut être celui de l'article principal, mais aussi celui d'une section plus précise. Ces routines (6 routines) précisent la problématique traitée dans un texte scientifique et le but à atteindre à la fin du texte. C'est la classe navigation textuelle qui comporte dans notre corpus le plus de routines (8 routines) avec des fréquences importantes. Cette fonction est centrale pour garantir le bon suivi de la progression du texte, en signalant au lecteur les informations dans les parties textuelles du discours en cours ou dans d'autres discours. Finalement, les routines de conclusion s'avèrent moins nombreuses avec deux routines identifiées. Leur présence indique la fin d'une chaîne d'argumentation ou de l'article total, permettant également au lecteur de se situer dans la progression discursive des écrits scientifiques.

Nous remarquons que dans les routines de structuration, les classes sémantiques initiales proposées par le LST peuvent être exploitées directement sans adaptation pour décrire des routines⁹³. En parallèle, l'emplacement de ces routines manifeste des régularités. On peut rappeler que les routines de planification de la structure et celle d'annonce d'objectif/l'objet d'étude se trouvent largement dans l'introduction, les routines de conclusion/déduction dans le corps principal et la conclusion, alors que celles de navigation textuelle peuvent se placer dans le corps principal, la conclusion et les notes.

Notre analyse montre aussi que sur le plan linguistique, les routines sont des propositions qui présentent une grande variabilité aux plans syntagmatique et paradigmatic. Pour chaque sous-catégorie, les champs sémantiques des mots (verbes, noms) en jeu présentent également des régularités.

⁹³ Ce n'est pas le même cas pour les autres routines.

Chapitre 6. Routines à fonction métalinguistique : définitions, modélisations et analyses

Les routines à fonction métalinguistique sont des routines commentant l'usage de la langue du discours scientifique afin d'éclairer, soit un contenu propositionnel énoncé précédemment pour éviter le malentendu ou l'incompréhension éventuellement générés, soit l'utilisation spécifique d'un mot, d'un terme ou d'une expression. Les fonctions métalinguistiques renvoient à l'autocontrôle de l'auteur sur l'emploi de la langue mise en énonciation. Dans cette partie, nous allons préciser d'abord ce que nous appelons « fonctions métalinguistiques » en rappelant certaines notions souvent évoquées, mais qui suscitent parfois une certaine controverse. Ensuite, nous expliquerons les fonctions que nous avons classées comme métalinguistiques ainsi que leurs définitions précises. Le reste du chapitre suivra la même progression que le chapitre précédent pour la modélisation et l'analyse des routines métalinguistiques.

6.1. Précisions sur le terme « métalinguistique »

Le terme « métalangage » est un terme incontournable dans l'étude des fonctions métalinguistiques. Dans le domaine de la linguistique, le concept de métalangage comporte une certaine ambiguïté, en particulier avec les termes proches de « métadiscours » et de « méta-énonciation », d'autant plus que ces deux derniers termes ne font pas consensus. Nous allons clarifier ce que nous entendons par « fonction métalinguistique » en précisant les termes « métalangage », « métadiscours » et « méta-énonciation ».

La notion de métalangage repose sur la capacité des langues à décrire et à représenter les langues elles-mêmes (Jaworski, Coupland, & Galasinski, 2012, p. 3). Le métalangage a ainsi été considéré comme exerçant « la fonction métalinguistique d'une langue donnée » (Rey-Debove, 1997, p. 21). En ce sens, la fonction métalinguistique s'applique en quelque sorte à la description, à la production ou à la compréhension d'une langue, comme l'indique Gombert (1990, p. 27).

Métalangage ou activité métalinguistique [...] concerne le langage et son utilisation, autrement dit comprenant : 1 — les activités de réflexion sur le langage et son utilisation, 2 — les capacités du sujet à contrôler et à planifier ses propres processus de traitement linguistique (en compréhension ou en production). Ces activités et ces capacités peuvent concerner tout aspect du langage qu'il soit phonologique (on parlera alors d'activités métaphonologiques), syntaxique (activités métasyntaxiques), sémantique (activités métasémantiques) ou pragmatique (activités métapragmatiques).

Les études sur le métalangage et la fonction métalinguistique s'inscrivent souvent dans une perspective didactique, en particulier l'acquisition d'une langue⁹⁴. Par exemple, dans une classe de langue française, le professeur de langue peut expliquer aux apprenants la langue française par la langue française ou par une autre langue. Le système de cette langue explicative employée par le professeur est bien un métalangage. Cette mise en évidence sur la langue et son utilisation rejoint l'idée de Rey-Debove (1997, p. 22) sur le premier niveau d'application du métalangage. Il s'agit, dans le domaine de la linguistique ou de la lexicographie, de l'ensemble des signes qu'on utilise pour décrire une langue naturelle donnée, ce qui correspond au langage qu'emploient les linguistes ou les didacticiens pour décrire ou enseigner la langue. Cet emploi pour parler de l'aspect linguistique d'une langue met en évidence la nature du métalangage en tant qu'instrument vis-à-vis de la langue-objet.

Cependant, Rey-Debove (*ibid.*) relève un deuxième niveau d'utilisation du métalangage dans une pratique courante où il implique le discours qu'on utilise pour commenter ce qu'on a déjà dit. Il reflète une réflexion de l'énonciateur sur l'expression ou le contenu déjà énoncé. Ainsi, le métalangage peut désigner les énoncés scientifiques ou didactiques dans des domaines autres que la langue comme l'économie, l'histoire, ou ludiques pour mieux communiquer. En étendant l'utilisation du métalangage des usages spécialisés aux usages plus courants, le deuxième emploi du métalangage n'a plus pour objectif la vérification d'un aspect linguistique ou grammatical du langage, mais principalement le commentaire sur le code ou le message. C'est principalement cette fonction, également présente dans la définition classique de Jakobson (1985, p. 116), qui nous intéressera ici. On parle également de fonction « épilinguistique » pour ce deuxième type d'usage à la suite de Culioli.

L'activité épilinguistique a été définie comme « une activité métalinguistique non consciente » (Culioli, 1968, p. 108). Elle est opposée à l'activité métalinguistique qui désigne plutôt « une activité métalinguistique consciente ». L'activité épilinguistique concerne ainsi plutôt « une maîtrise fonctionnelle de règles d'organisation ou d'usage de la langue » (Gombert, 1990, p. 27). Le discours ainsi produit après des activités épilinguistiques est appelé également « gloses spontanées » (Culioli, 1999, p. 74). Ce sont « les textes qu'un sujet produit lorsque, de façon spontanée ou en réponse à une sollicitation, il commente un texte précédent » (Culioli, 1981 cité par Berthoud, 1982, p. 159). Ces gloses renvoient « à la pratique langagière du sujet énonciateur » (*ibid.*). Dans notre étude, nous conserverons le terme classique de

⁹⁴ À titre indicatif, l'article de Huot & Schmidt (1996) sur Conscience et activité métalinguistique. L'ouvrage de (Brédart & Rondal, 1997) sur les activités métalinguistiques chez des enfants.

« métalinguistique », car « l'activité métalinguistique – qu'elle soit de type “méta” ou “épi” – est constitutive de l'activité langagière, conçue comme activité de production et d'interprétation des énoncés » (Boutet, Gauthier, & Saint-Pierre, 1983, p. 214). Une autre difficulté pourrait être que les deux fonctions sont présentes dans l'écrit scientifique et qu'il est parfois difficile de les distinguer.

À côté de ces deux grandes acceptions du terme « métalangage » (métalinguistique/épilinguistique), les notions de « métadiscours » ou de « méta-énonciation » intègrent aussi, dans certaines études, la fonction métalinguistique. Dans l'approche énonciative où s'inscrivent de nombreuses études de langue française, Rey-Debove (1997, p. 20) envisage le métadiscours comme la mise en œuvre de la métalangue naturelle dans le discours. À cet effet, le métadiscours est classé en tant que constituant du métalangage, et les énoncés métadiscursifs remplissent la fonction métalinguistique. Authier-Revuz (1995, p. 18) évoque la notion de métaénonciation en la comparant avec le métalangage et le métadiscours : la fonction d'explication de la langue, du sens des propositions et le renvoi au dit de l'énonciateur peuvent être caractérisés comme métalinguistique, alors que la fonction qui commente l'énonciation est vue comme métadiscursive.

Rappelons que dans le monde anglophone, où les études sur l'anglais académique adoptent souvent une approche fonctionnelle, par exemple, la théorie de la grammaire systémique fonctionnelle (SFG) (Halliday, 1973), on parle de métadiscours pour indiquer tout élément discursif qu'utilise l'auteur pour organiser le texte, pour indiquer son positionnement, pour commenter ou évaluer le contenu et la langue du discours en cours (*cf.*, Chapitre 3). Dans cette approche, la fonction métalinguistique n'est qu'une des fonctions du métadiscours.

Comme nous l'avons expliqué dans le chapitre 3, nous nous intéressons spécifiquement à la fonction métalinguistique qu'un sous-ensemble d'expressions métadiscursives remplit dans le discours, quand celles-ci commentent la langue en production. Les fonctions métalinguistiques en ce sens appartiennent aux fonctions métadiscursives dans le discours scientifique. Elles ont pour objectif de vérifier la bonne compréhension de l'expression utilisée (un terme, un mot, une expression, etc.) ou du contenu exprimé.

6.2. Fonctions métalinguistiques dans le métadiscours

Reconnaître les fonctions comme métalinguistiques n'est pas facile, car il existe peu d'inventaires complets qui les abordent directement, d'autant plus que des études portant sur les fonctions métalinguistiques dans le discours ont parfois tendance à s'intégrer dans l'étude

des fonctions textuelles. Par exemple, dans l'étude de Rahman (2004) sur les dispositifs métalinguistiques favorisant la lecture dans le discours scientifique, l'auteur adopte une perspective plus large pour classer les dispositifs métalinguistiques, en incluant les marqueurs de fonctions textuelles dans les modèles du métadiscours. À la différence de ce classement, nous séparons les fonctions métalinguistiques des fonctions de structuration textuelle. Pour nous, les fonctions textuelles dans les modèles métadiscursifs ont pour objectif d'organiser le contenu d'un texte, comme ce que nous avons évoqué dans le Chapitre 5 sur le plan du texte, l'annonce de l'objectif/l'objet d'étude, la navigation textuelle et la conclusion. En ce qui concerne les fonctions métalinguistiques, nous nous focalisons sur leur capacité de commenter la langue utilisée et le contenu communiqué.

La taxonomie des fonctions discursives du métadiscours dans l'anglais académique écrit d'Ädel (2006, p. 60) et parlé (Ädel, 2010, p. 83) présente un relevé intéressant des fonctions qui peuvent être étiquetées comme métalinguistiques. Ces deux études regroupent 7 fonctions métalinguistiques : la définition, l'énonciation, la réparation, la reformulation, les commentaires sur la forme ou le sens linguistiques, la clarification et la gestion terminologique.

Dans l'étude du discours scientifique écrit, Ädel (2006, p. 62) a identifié deux fonctions qui portent sur la clarté du code : la définition et le commentaire métaénonciatif. La définition est centrale pour le discours scientifique qui manifeste une exigence forte de précision sur le sens des termes ou des concepts essentiels employés par l'auteur (Ädel, 2006, p. 62). Elle concerne à la fois la fonction de définition d'un terme (*what do we mean by... X is what we call Y*), et le commentaire sur la compréhension d'un terme (*I do not quite understand what is meant by...*)(*ibid.*). Le commentaire métaénonciatif concerne les informations à communiquer dans le discours en cours, en ayant recours aux verbes d'énonciation, tels que *say, speak, talk (dire, parler, discuter)*. Elle est actualisée par des expressions construites autour de ces verbes comme *What I am saying is... I am talking about... I can mention that ... (ibid.)*.

Dans une étude plus récente sur le discours scientifique écrit et parlé, Ädel (2010) a listé cinq fonctions métalinguistiques, à savoir la réparation, la reformulation, les commentaires sur la forme ou le sens linguistiques, la clarification, la gestion terminologique.

La réparation et la reformulation ont toutes affaire à l'énoncé source. Plus précisément, la réparation consiste en un acte de correction ou d'annulation concernant l'énoncé source comme (*maybe I should've said, I don't mean to say that out loud*), tandis que la reformulation vise à ajouter une valeur en proposant une explication alternative (*if you'll allow me just, rephrase it a little...*) à l'énoncé source (Ädel, 2010, p. 84). Les commentaires sur la forme ou

le sens linguistique se focalisent sur la forme linguistique, le choix des mots et/ou le sens (*to put it in French language, we can therefore say that...*)(*ibid.*). Cette fonction assure le contrôle du code utilisé. La clarification présente la vision anticipatrice de l'auteur pour éviter une interprétation erronée chez le lecteur (*I'm by no means trying to ...; I wish simply to...*). Ces trois fonctions ressemblent à la fonction remplie par les marqueurs de *code glosses* de Hyland (2005, p. 52) qui a pour objectif d'exemplifier et de clarifier les messages transmis (*such as; in other words...*). La gestion terminologique prend en charge la fonction de définition des termes (*I define _ as _*), ou la fonction de dénomination (appelée étiquetage) (*what I shall call...*) (Ädel, 2010, p. 84), et inclut donc la fonction de définition discutée plus haut.

À partir de ces classes, nous avons élaboré notre propre typologie des routines métalinguistiques en les adaptant aux données de notre corpus. Nous récapitulons ici, les deux plans sur lesquels se focalise la classe des fonctions métalinguistiques : 1) la précision de l'usage de la langue en ce qui concerne le code, le terme, l'expression, 2) la précision du sens véhiculé dans une phrase pour une meilleure explication du contenu propositionnel. Nous avons repéré 13 routines métalinguistiques remplissant 4 fonctions métalinguistiques.

6.3. Routines métalinguistiques

Nous avons identifié quatre sous-fonctions métalinguistiques : la reformulation, la réparation, la précision terminologique et l'emprunt terminologique. La reformulation et la réparation concernent le plan sémantique du code, veillant à la clarté du sens. La précision et l'emprunt terminologiques se focalisent sur l'utilisation du code, notamment sur les termes, les mots et les expressions disciplinaires. La précision terminologique comporte encore trois fonctions plus précises : la définition, la dénomination et la désignation, comme l'indique la figure 6-1 :

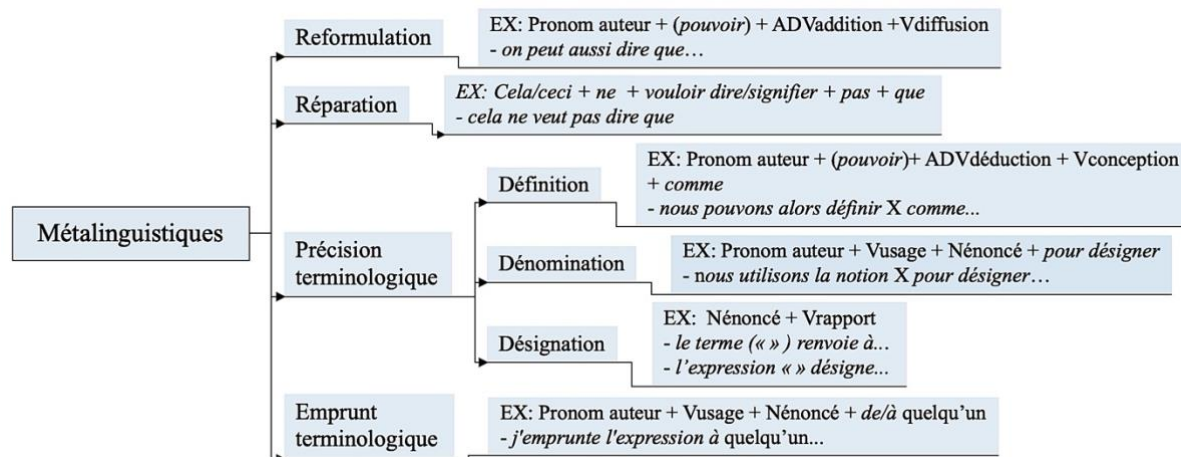


Figure 6-1 : Les sous-fonctions métalinguistiques

Dans cette section, nous présenterons ces sous-classes de fonctions métalinguistiques ainsi que les définitions que nous leur avons attribuées et les observations sur les caractéristiques linguistiques des routines identifiées.

6.3.1. Routines de reformulation

Nous retenons d'abord la fonction métalinguistique de reformulation, qui a beaucoup été étudiée dans le domaine de l'interaction verbale depuis 40 ans. La définition la plus classique est formulée par Gülich & Kotschi (1987, cité par Eshkol-Taravella & Grabar, 2018, p. 5) : la reformulation est « [une] opération linguistique qui établit une relation d'équivalence entre un énoncé source (ES) et un énoncé reformulateur (ER) liés par un marqueur de reformulation (MR) ». Cette définition met en évidence le critère d'équivalence sémantique entre l'ES et l'ER. Cette opération qui cherche à garder un sens identique entre deux énoncés est également appelée « la reformulation paraphrastique » ou plus simplement « la paraphrase ». Il s'agit d'une nouvelle façon de dire le même sens que l'énoncé source. La reformulation de ce type est souvent introduite par des marqueurs comme *en d'autres termes* ou *c'est-à-dire*. À côté de la reformulation paraphrastique, Rossari (1993, p. 17) relève, la reformulation non paraphrastique. Cette dernière « opère une réinterprétation du point de vue » et introduit « une nouvelle perspective énonciative annoncée par les instructions sémantico-paraphrastiques du marqueur [de reformulation] » (*ibid.*). Dans le cas de la reformulation non paraphrastique, la liaison entre l'ES et l'ER est souvent réalisée par des marqueurs comme *en somme* ou *bref* (Rossari 1993, p. 17).

À l'instar de Rossari (1993), Steuckardt (2009, p. 172), dans une étude sur les marqueurs de reformulation, distingue trois fonctions essentielles dans une perspective rhétorique dont (a) est proche de la reformulation paraphrastique et (b) et (c) peuvent être considérés comme non paraphrastiques :

- (a) fabriquer un énoncé Y qui a un sens dénotatif identique à l'énoncé X ;
- (b) changer de façon pragmatique la portée de l'énoncé X par la formulation de l'énoncé Y ;
- (c) changer de façon pragmatique l'orientation argumentative de l'énoncé X par l'énoncé Y.

Comme nous le voyons, la reformulation dénote donc plus qu'une relation d'équivalence entre deux énoncés, puisqu'elle peut introduire une variation sémantique et argumentative. Dans les écrits scientifiques, la reformulation consiste généralement en une stratégie pratique

de l’auteur pour clarifier le sens exprimé (*c’est-à-dire, autrement dit*), pour reprendre des informations (*en somme, en un mot*), pour reconsidérer un point de vue (*en définitive, au total*), afin de garantir une meilleure compréhension de l’information et du savoir chez le lecteur (Tran, 2014, p. 96).

Les études sur la reformulation portent souvent sur les marqueurs de reformulation, ou sur les « marqueurs de glose », c’est-à-dire des marqueurs qui permettent d’introduire une phrase explicative pour expliciter une phrase précédente (Langhans, 2005, p. 5). Par exemple, Steuckardt (2018, p. 20) identifie des marqueurs de glose formulés autour des mots *dire*, comme *c’est-à-dire, comme on dit, pour mieux dire*. On peut trouver dans les écrits scientifiques, des marqueurs comme *c’est-à-dire, en d’autres termes, autrement dit, à savoir, en somme, en résumé, en un mot, en définitive, au total, somme toute, au fond, en fin de compte* (Tran, 2014, p. 104).

En bref, par l’acte de formuler d’une façon différente des énoncés sources, la reformulation rend des informations éventuellement difficiles à comprendre plus accessibles pour le lecteur. Derrière cette fonction métalinguistique se cache aussi la stratégie rhétorique de l’auteur qui prend en compte son lecteur et anticipe ses attentes. À partir de l’étude de corpus, nous avons identifié trois routines de reformulation.

Routine N° 1 de reformulation.

Description sémantique	<L’ auteur ‘ modifie la façon de dire ’ un énoncé source>				
Exemples	<i>On peut même dire que...</i> <i>On peut également dire que...</i> <i>On parle aussi de...</i>				
Type de phrase	Phrase déclarative				
Structure syntaxique	[pronom AUTEUR]+ (<i>pouvoir</i>)+ $\left(\begin{array}{c} \text{[adverbe d'ADDITION]} \\ \text{[adverbe de CONCESSION]} \end{array} \right)$ + [verbe de DIFFUSION] + SN/ <i>que</i>				
Réalisations lexicales	<i>On/nous</i>	<i>(peut/pouvons)</i>	<i>même/aussi/également/de même/au moins</i>	<i>dire/parler</i>	<i>SN/que</i>
Éléments périphériques	Marqueurs discursifs, comme les adverbes d’addition <i>aussi, également, de même</i> , les adverbes de concession <i>même, cependant, au moins</i> . Ces éléments permettent d’élargir ou restreindre la portée pragmatique ou argumentative.				
Paramètres énonciatifs	Dialogismes intralocutif et interlocutif. Cette routine s’inscrit d’abord dans la dimension intralocutive, car elle consiste à préciser les propres énoncés de l’auteur. En parallèle, comme toutes les routines de guidage, elle peut être envisagée dans la dimension interlocutive, on parle également d’interlocutif-anticipatif.				
Position textuelle	Dans le corps principal du discours (pas de position spécifique).				
Fonctions	Reformulation d’un énoncé source par un énoncé reformulé, avec souvent un changement d’orientation pragmatique et argumentative.				

Tableau 6-1 : Routine N° 1 de reformulation

Routine N° 2 de reformulation.

Description sémantique	< 'P' 'énoncé source' est 'reformulé' >		
Exemples	<i>Cela veut dire que... ce qui veut dire que...</i> <i>Cela signifie que...</i>		
Type de phrase	Phrase déclarative		
Structure syntaxique	Cela ce qui + { <i>veut dire</i> / <i>signifie</i> } + que		
Réalisations lexicales	<i>Cela/ce qui (pouvoir)</i>	<i>veut dire/signifie</i>	<i>que</i>
Éléments périphériques	Marqueurs discursifs, comme les adverbes d'addition : <i>aussi, également</i> ; les adverbes d'exemplification : <i>en particulier, surtout, notamment</i> ; les adverbes de réexamen : <i>en fait</i> ; les adverbes de concession : <i>en tout cas, néanmoins</i> ; les adverbes de clarification : <i>autrement dit, concrètement, en d'autres termes</i> , les adverbes de conclusion ou de déduction : <i>donc</i> ; les adverbes de modalité : <i>simplement, seulement</i> .		
Paramètres énonciatifs	Dialogismes intralocutif et interlocutif. Cette routine s'inscrit d'abord dans la dimension intralocutive, car elle consiste à préciser les propres énoncés de l'auteur. En parallèle, comme toutes les routines de guidage, elle peut être envisagée dans la dimension interlocutive, on parle également d'interlocutif-anticipatif.		
Position textuelle	Dans le corps principal du discours (pas de position spécifique). Elle se trouve souvent en tête d'une phrase.		
Fonctions	Reformulation d'un énoncé source par un énoncé reformulé, avec souvent un changement d'orientation pragmatique et argumentative.		

Tableau 6-2 : Routine N° 2 de reformulation

Routine N° 3 de reformulation.

Description sémantique	< l'énoncé source peut être 'formulé d'une façon différente' >			
Exemples	<i>Pour le dire autrement</i>			
Type de phrase	Phrase déclarative, proposition infinitive			
Structure syntaxique	pour + (le) + dire + { <i>autrement d'un mot</i> / <i>en peu de mot</i> / <i>simplement</i> }, cette routine ne comporte pas de classes sémantiques.			
Réalisations lexicales	<i>Pour</i>	<i>(le)</i>	<i>dire</i>	<i>autrement/d'un mot/en peu de mot/simplement</i>
Éléments périphériques	Pas d'élément périphérique.			
Paramètres énonciatifs	Dialogismes intralocutif et interlocutif. Cette routine s'inscrit d'abord dans la dimension intralocutive, car elle consiste à préciser les propres énoncés de l'auteur. En parallèle, comme toutes les routines de guidage, elle peut être envisagée dans la dimension interlocutive, on parle également d'interlocutif-anticipatif.			
Position textuelle	Dans le corps principal du discours (pas de position spécifique). Elle se trouve souvent en tête d'une phrase, est une proposition subordonnée.			
Fonctions	Reformulation d'un énoncé source par un énoncé reformulé, avec souvent un changement d'orientation pragmatique et argumentative.			

Tableau 6-3 : Routine N° 3 de reformulation

6.3.1.1 Caractéristiques syntaxiques, lexicales et sémantiques de ces routines

Nous pouvons constater à partir de la modélisation linguistique que les routines de reformulation possèdent des structures syntaxiques assez figées par comparaison aux routines de structuration, car elles sont accompagnées de réalisations lexicales très limitées. Par conséquent, les étiquettes conceptuelles ne correspondent pas toujours à la plupart des mots

dans une classe sémantique du LST, mais à un ou deux mots précis. La routine N° 1 <L' **'auteur'** **'modifie la façon de dire'** un énoncé > dont l'étiquette **'modifie la façon de dire'** correspond seulement au verbe *dire* et *parler* (exemples 1 et 2). Les étiquettes **'reformulé(er)'** et **'formulé d'une façon différente'** dans la routine N° 2 et N° 3 sont réalisées respectivement par deux verbes *vouloir dire/signifier* et l'adverbe *autrement* (exemples 3, 4). Parmi les trois routines, la routine N° 3 est une proposition infinitive qui est la plus figée (exemple 5). Dans les exemples suivants, les routines apparaissent en caractère gras, l'énoncé source (ES) et l'énoncé reformulé (ER) sont en italique.

- [1] Il était donc aligné avec un axe naturel de l'environnement lorsque les planches étaient présentées verticalement, ou avec *un axe du dispositif lorsqu'elles étaient présentées horizontalement* (dans ce dernier cas, **nous parlerons aussi d'axe vertical** puisque l'orientation 0° correspond à une projection de l'axe vertical de l'environnement sur un plan horizontal, Shepard et Hurwitz, 1984). (Psychologie) (ça peut aussi être une routine de nomination, on nomme ES par ER)
- [2] Au contraire, *la connaissance scientifique est relative, restreinte, posant elle-même les limites de son domaine de validité*, et **on peut même dire** en un certain sens **qu'elle est plus approximative**. (Sciences de l'éducation)
- [3] La démonstration de Vygotskij est convaincante ; *il voit dans le fait que le sujet utilise les significations du mot d'une manière fonctionnelle le fondement et la réalité du langage intérieur*. **Cela veut dire qu'on ne peut pas localiser la pensée dans le foyer intérieur du sujet, car c'est en utilisant le langage que le sujet donne naissance à la pensée**. (Linguistique)
- [4] De façon symétrique, *une lecture « animiste » se heurte cette fois à la double distance des intériorités et des physicalités qui caractérise le régime intermédiaire des relations homme/animal*. Concrètement, **cela signifie que la communication entre eux est rompue sans que pour autant leur discontinuité soit totale et définitive**. (Anthropologie)
- [5] L'article s'aligne sur l'implicite idéologique et épistémologique selon lequel l'essentiel de l'activité humaine serait réductible à des "faits de langage", ou plus précisément, réduit l'ensemble des processus d'information-communication organisationnels à *des faits langagiers* ou, **pour le dire autrement à des formes sémiotiques**. (Science de l'information)

Du point de vue syntaxique, les routines de reformulation ont des portées bien différentes qui peuvent varier en fonction du besoin. Elles peuvent porter sur un énoncé source (ES) entier (portée globale), ou sur un constituant de l'ES (portée locale) (Tran, 2014, p. 113). Par exemple, la routine *nous parlons aussi de* dans l'exemple 1 établit un rapport synonymique entre deux objets scientifiques *un axe du dispositif* et *un axe vertical*, agissant donc une portée syntaxique locale. *On peut même dire que* dans l'exemple 2 introduit avec l'énoncé reformulé

(ER) une reformulation de portée globale sur l'ES entier. La routine N° 2 fonctionne généralement sur le sens global de l'ES. La routine N° 3 dispose d'une construction syntaxique proche d'une locution adverbiale. Sa position syntaxique peut se situer souvent à la tête d'un énoncé entre l'ES et l'ER, mais elle peut aussi se trouver au milieu d'une phrase (exemple 5).

Du point de vue sémantique, les routines de reformulation ne se limitent pas à une simple paraphrase du sens. La routine N° 1, dans le cas où elle se réalise avec le verbe *dire* comme *on peut dire que* permet, hors de la reformulation de l'ES, de 1) résumer les informations dans l'ES et donne une formulation plus claire pour interpréter l'ES (exemple 2) ; 2) de conclure ou de déduire une information à partir de l'ES (exemple 6).

[6] *Si l'on considère que p est un bon indicateur d'inégalité de la population face au revenu, **alors on peut dire que** non seulement la société dépollue parce que la pollution est un bien indésirable, mais aussi parce que c'est le seul moyen de redistribuer de la richesse, dans notre modèle.* (Économie)

La routine N° 2 permet de réexaminer le contenu de l'ES et de changer également l'orientation argumentative (3) ou d'induire un changement de la portée sémantique de l'ES (4). La dernière routine peut créer soit un rapport synonymique avec une partie d'un ES (portée locale) (5), soit un rapport métonymique avec la globalité de l'ES (4).

Nous souhaitons souligner ici le faible nombre des routines de reformulation et leur caractère peu varié qui peuvent s'expliquer par deux raisons :

1. Les écrits scientifiques comportent beaucoup de locutions adverbiales reformulatives comme *en d'autres termes, c'est-à-dire, à savoir*, etc. qui sont des outils très commodes pour formuler des configurations reformulatives avec le schéma <ES + marqueur reformulatif + ER>.
2. L'objectif de la reformulation consiste à rendre l'interprétation de l'ES plus claire pour le lecteur. Dans ce cas, l'utilisation des propositions de reformulation pourrait rendre la phrase ER plus lourde, et peut entraver la compréhension. Mais l'utilisation de ces routines peut introduire une autre fonction argumentative comme la déduction ou l'addition.

6.3.1.2 *Éléments périphériques et paramètres énonciatifs*

Nous trouvons systématiquement dans les routines N° 1 et N° 2 des marqueurs discursifs. La routine N° 1 présente avec ceux dans les paradigmes d'ADDITION *aussi, également, de même* et de CONCESSION *même, cependant, au moins*. Les paradigmes d'adverbes discursifs apparaissant avec la routine N° 2 sont plus variés. En dehors des deux classes citées plus haut, on trouve encore des adverbes de RÉEXAMEN : *en fait* ; de CONCESSION avec plus de

variation : *en tout cas, néanmoins* ; de CLARIFICATION : *autrement dit, concrètement, en d'autres termes*, de CONCLUSION/DÉDUCTION : *donc* et aussi des adverbes de modalité : *simplement, seulement*. Ces marqueurs discursifs jouent un rôle primordial. Comme nous l'avons vu au début de cette section, la reformulation ne consiste pas seulement à créer une relation équivalente entre l'ES et l'ER, mais également à introduire une nouvelle perspective argumentative, ou à élargir ou restreindre la portée sémantique de l'ES par l'ER. La présence des adverbes permet de mettre en évidence ces éléments supplémentaires. La routine N° 3 de reformulation ne cooccurt pas avec les marqueurs discursifs, car elle est très lexicalisée et s'apparente quasiment à un adverbe.

Les routines de reformulation s'inscrivent fortement dans la dimension intralocutive (aussi interlocutive-anticipative) qui prête attention aux énoncés propres de l'auteur. Le pronom AUTEUR dans la routine N° 1 (*on/nous*) désigne l'auteur des articles de recherche. C'est lui qui exerce l'acte de reformulation en proposant un ER. Dans les routines N° 2 et N° 3, la présence de l'auteur n'est pas aussi explicite que dans la routine N° 1, mais l'action de reformulation est mise en avant par les verbes métalinguistiques *dire* et *signifier*. L'actualisation énonciative se réalise à travers les pronoms démonstratifs *cela/ce qui*, le pronom neutre *le*, autrement dit, les déictiques qui peuvent dénoter une référence anaphorique. Si on examine plus minutieusement le corpus, on peut constater que ces routines peuvent aussi impliquer le dialogisme interdiscursif lorsqu'il s'agit de paraphraser le discours d'autrui.

[7] *En paraphrasant Éric Macé, qui écrit « ce ne sont pas les races qui créent le racisme, mais le racisme qui invente les races » (2006 : 147), on peut dire que ce ne sont pas les Mandé qui pratiquent la polygamie, mais la polygamie qui invente les Mandé. (Anthropologie)*

Dans l'exemple 7, l'ES est un énoncé d'autrui, et l'auteur le reformule par son propre énoncé. On a donc la transformation de l'énoncé d'autrui en un énoncé de l'auteur. Ce procédé inscrit d'une façon plus concrète l'énoncé d'autrui dans la problématique abordée dans l'article.

6.3.1.3 *Fonctionnement rhétorique*

En général, les routines établissent un rapport synonymique entre l'ES et l'ER en gardant la possibilité d'introduire une nouvelle perspective pour interpréter l'ES pour le lecteur. C'est pourquoi elles peuvent remplir des fonctions secondaires comme la récapitulation ou la conclusion comme vu plus haut. Ces expressions permettent à l'auteur une meilleure explication et clarification du sens, donnant lieu à une meilleure compréhension du texte scientifique chez le lecteur.

6.3.2. Routines de réparation

Ensuite, nous retenons également un autre type de routines proche de la fonction de reformulation, mais qui a un objectif réparateur et que nous nommons « routines de réparation ». Cette classe a été inspirée par la classe « réparation » (*repairing* en anglais) de la typologie d'Ädel, (2010, p. 84) pour annuler la contribution précédente comme *I'm sorry bound to the state, maybe I should've said* dans le discours scientifique à l'oral.

Cette fonction vise à corriger ou annuler la contribution précédente. Elle consiste en une reformulation corrective. Nous pouvons qualifier la réparation comme « reformulation négative ». Plus précisément, au lieu de mettre en place une équivalence entre deux énoncés, la fonction de réparation cherche à corriger, voire annuler ce qui était énoncé, ce qui permet à l'auteur de changer son positionnement scientifique.

Bien qu'Ädel n'ait pas identifié de données pour cette fonction à l'écrit, nous avons remarqué que les routines de reformulation N° 1 et N° 2 discutées plus haut exercent bien une fonction réparatrice quand elles s'actualisent à la forme négative. Au lieu de classer ces routines de reformulation négative simplement dans les routines de reformulation, nous avons décidé de les mettre dans une nouvelle classe, cela en raison de deux considérations : 1) ces routines de réparation sont toutes aussi nombreuses ; 2) l'annulation ou la correction de l'ES ne représentent pas la même stratégie rhétorique que la reformulation. Elle peut parfois réparer une fausse interprétation éventuelle chez le lecteur. Les routines de réparation qui correspondent aux routines de reformulation à forme négative fonctionnent de façon spécifique. Comme les routines de reformulation, nous avons identifié deux routines de réparation que nous modélisons ci-dessous.

Routine N° 1 de réparation.

Description sémantique	<L' 'auteur' 'répare' un énoncé source >				
Exemples	<i>On ne peut pas dire...</i>				
Type de phrase	Phrase déclarative à forme négative				
Structure syntaxique	[pronom AUTEUR]+ <i>ne pas</i> + (<i>pouvoir</i>)+ [verbe de DIFFUSION] + SN/ <i>que</i>				
Réalisations lexicales	<i>On/nous</i>	<i>ne (peut/pouvons) pas/plus</i>	<i>donc/réellement/en mot/nullement</i>	<i>un</i>	<i>dire/parler de SN/que</i>
Éléments périphériques	Marqueurs discursifs ou modaux, comme l'adverbe de conclusion/déduction : <i>donc</i> ; de récapitulation : <i>en un mot</i> . Des adverbes de modalité : <i>réellement</i> ; de certitude : <i>véritablement</i> ; de négation : <i>nullement</i> .				
Paramètres énonciatifs	Dialogismes intralocutif et interlocutif. Cette routine s'inscrit d'abord dans la dimension intralocutive, car elle consiste à corriger les propres énoncés de l'auteur. En parallèle, comme toutes les routines de guidage, elle peut être envisagée dans la dimension interlocutive, on parle également d'interlocutif-anticipatif.				
Position textuelle	Dans le corps principal du discours (pas de position spécifique).				
Fonctions	Réparation d'un énoncé source par un énoncé corrigé, il s'agit également de l'anticipation de la réaction du lecteur.				

Tableau 6-4 : Routine N° 1 de réparation

Routine N° 2 de réparation.

Description sémantique	< 'l'énoncé source' est 'réparé' >				
Exemples	<i>Cela ne veut pas dire que... ce qui ne veut pas dire que... Cela ne signifie pas que...</i>				
Type de phrase	Phrase déclarative				
Structure syntaxique	Cela ce qui + ne + { veut dire } + { ^{pas} nullement } + que				
Réalisations lexicales	Mais/en effet	<i>cela/ce qui</i>	<i>ne</i>	<i>veut</i>	<i>dire/signifie que</i>
Éléments périphériques	Marqueurs discursifs et de négation. Les adverbes de concession : <i>mais, néanmoins</i> , l'adverbe de négation : <i>nullement</i> .				
Paramètres énonciatifs	Dialogismes intralocutif et interlocutif. Cette routine s'inscrit d'abord dans la dimension intralocutive, car elle consiste à corriger les propres énoncés de l'auteur. En parallèle, comme toutes les routines de guidage, elle peut être envisagée dans la dimension interlocutive, on parle également d'interlocutif-anticipatif.				
Position textuelle	Dans le corps principal du discours (pas de position spécifique). Elle se trouve souvent en tête d'une phrase.				
Fonctions	Réparation d'un énoncé source par un énoncé corrigé, il s'agit également de l'anticipation de la réaction du lecteur.				

Tableau 6-5 : Routine N° 2 de réparation

6.3.2.1 Caractéristiques syntaxiques, lexicales et sémantiques de ces routines

Les routines de réparation sont des formes négatives de deux routines de reformulation. Deux structures ont été identifiées (dans les exemples, l'ES et l'ER sont en italique).

a. Routines de construction agentive : Pronom auteur + *ne pas pouvoir* + *dire que* :

[8] On peut donc conclure que, à l'exception de la variable « secteur », les matrices de covariance observées sont égales à travers les groupes. *Enfin, en ce qui concerne les tests de Shapiro - Wilk, tous les résultats sont significatifs. Nous ne*

pouvons donc pas dire que nos données se distribuent normalement.
(Linguistique)

- b. Routines de construction pronominale démonstrative : [*cela/ce qui/ceci*] + [*ne veut dire/signifie nullement/néanmoins/pas*] + (*que*) :

[9] On oublie souvent que Spinoza avait précisé dans un texte de jeunesse que s'il n'y a pas plusieurs substances, **cela ne signifie pas que** la substance soit « une ».
(Anthropologie)

Ces routines de réparation peuvent véhiculer une négation plus forte quand elles ne se réalisent avec l'adverbe de négation *nullement* qui exprime le sens « en aucune manière, pas du tout », comme l'indique l'exemple 10.

[10] [6] **Je ne veux nullement dire par là que** tous les États africains précoloniaux étaient des États-nations, mais certains d'entre eux pouvaient être caractérisés ainsi. (Histoire)

6.3.2.2 *Éléments périphériques et paramètres énonciatifs*

Les routines de réparation sont employées avec des adverbes de concession (*mais, néanmoins*) ou de déduction (*donc*). Les adverbes de déduction marquent une conséquence tirée de l'ES. Il s'agit d'une déduction négative que l'auteur signale à son lecteur. *Nous ne pouvons donc pas dire que nos données se distribuent normalement* ici peut être paraphrasé par *nous pouvons donc dire que nos données ne se distribuent pas normalement*.

Quand on emploie un adverbe de concession pour réparer ce qui a été énoncé, il peut créer une surprise par l'ER chez le lecteur, parce que la concession tisse une relation d'opposition entre l'énoncé précédent et l'énoncé suivant. La routine de réparation avec un adverbe de concession permet au lecteur de réfléchir en dehors de l'ES et de changer de point de vue. Dans l'exemple 11, l'ER introduit un effet de surprise en signalant que la relation de comparaison entre une grande maison et une moyenne maison n'était pas tout le temps valide.

[11] *Par exemple une grande maison est toujours supérieure en volume à une maison qui serait plus moyenne, **mais cela ne veut pas dire que** quand on parle d'une grande maison, il soit toujours question de la comparer avec quelque maison qui serait moyenne.* (Linguistique)

En un mot, les marqueurs de concession et de déduction renforcent la volonté de réparer ou contester la contribution précédente ou la déduction anticipée du lecteur.

Comme les routines de reformulation, les routines de réparation constituent aussi un lieu de dialogisme intralocutif, encore plus caractéristiques du dialogisme interlocutif-anticipatif cette fois. Comme nous l'avons expliqué au début du chapitre, les activités métalinguistiques sont parfois implicites. Nous constatons que la réparation s'avère en revanche plus explicite.

Elle consiste à corriger l'énoncé source et introduit une précision différente d'après l'anticipation de la réaction du lecteur. L'auteur prend soin de ne pas être mal compris par son lecteur. L'exemple 10, à travers la routine *je ne veux nullement dire par là que*, explique le souci de surveiller le point de vue de l'auteur pour éviter le malentendu.

6.3.2.3 *Fonctionnement rhétorique*

Ces routines de réparation mettent en place une double fonction de réparation et de déduction. Elles manifestent aussi une fonction interactionnelle d'anticipation des objections du lecteur. Elles reflètent fortement la dimension dialogique interlocutive du discours scientifique que l'on retrouve avec d'autres marqueurs comme ceux de surprise (*il est surprenant de constater*) (Tutin, 2017). En introduisant ces routines, l'auteur anticipe la déduction du lecteur, et suscite une surprise en signalant une déduction contraire aux attentes de celui-ci.

6.3.3. Routines de précision terminologique

La classe de fonction de précision terminologique comporte toutes les routines qui ont pour fonction d'expliquer l'usage ou l'interprétation d'un mot, d'un terme, ou encore d'une expression utilisée dans le texte scientifique. Cette classe regroupe donc la fonction de gestion terminologique et la fonction des commentaires sur la langue (Ädel, 2010, p. 84). Nous la divisons en trois sous-classes : la **définition**, la **dénomination**, la **désignation**.

La **définition** est une activité cognitive indispensable de l'auteur dans les écrits scientifiques qui consiste à attribuer un sens explicite aux mots, termes, expressions, notions ou concepts par des énoncés définitoires. Ces derniers peuvent s'avérer explicitement métalinguistiques lorsqu'ils comportent des verbes métalinguistiques comme *définir* ou *signifier*. Prenons l'exemple des énoncés définitoires comme *nous pouvons définir X comme...*, *X est défini comme...*, etc. Dans d'autres cas, ils sont non explicitement métalinguistiques comme la forme copulative⁹⁵ avec le verbe *être*, comme *X est Y*.

La **dénomination** recense des routines qui visent à « produire des dénominations, c'est-à-dire des mots qui s'interprètent comme la dénomination d'une chose » (Riegel, 1990, p. 99). Au point de vue sémiotique, la dénomination et la désignation (que nous verrons plus loin) instaurent une relation de référence entre un signe linguistique et un élément de la réalité (Kleiber, 2001, p. 23). Mais le sens du processus opérationnel est tout à fait inverse : la

⁹⁵ Ces énoncés ont été exclus de nos données de routines, car ils ne comportent pas au moins 3 mots comme l'un des critères fixés dans la section 4.5.

dénomination constitue un processus cognitif de l’auteur des écrits scientifiques pour créer une expression linguistique qui permet de renvoyer, soit à un phénomène spécifique ou une chose concrète de la réalité, soit à un concept abstrait relevant du domaine scientifique, c’est-à-dire, une relation chose → signe. Ce processus se fonde sur « un acte de dénomination préalable » qui crée une relation référentielle entre une chose extralinguistique et un signe linguistique (Kleiber, 1984, p. 79). Il est nécessaire d’isoler la dénomination métalinguistique de la dénomination ordinaire : la première porte sur des concepts et établit « une association référentielle durable ou stable » entre un signe linguistique et un concept pour institutionnaliser la représentation conceptuelle de ce signe linguistique dans le texte. La deuxième porte surtout sur les noms propres comme *ce garçon s’appelle Nicolas* (Kleiber, 2001, p. 23), et nous l’excluons bien entendu de notre étude.

La **désignation** est un processus inverse de la dénomination, cherchant à qualifier une chose au travers d’une notion ou un terme déjà existant. Il s’agit, par conséquent, d’une relation signe → chose. Contrairement à la dénomination, elle n’exige pas un acte de dénomination préalable, peu importe que la relation référentielle entre le signe linguistique et la chose extralinguistique soit stable ou non. Elle vise en théorie à créer un rapport entre ce signe linguistique (déjà existant dans la discipline, ou déjà créé dans les parties précédentes du discours de l’auteur) et la chose extralinguistique. Par exemple : *le mot X renvoie à Y*.

Dans cette catégorie, 6 routines ont été identifiées. Voilà leurs modélisations.

Routine N° 1 de précision terminologique - définition.

Description sémantique	< Un ‘ concept/phénomène/fait scientifique X ’ est expliqué par un ‘ énoncé définitoire Y ’ >				
Exemples	<i>X peut être défini comme...</i> <i>X peut alors se définir comme...</i>				
Type de phrase	Phrase déclarative, au présent.				
Structure syntaxique	X + (pouvoir) + être + (adverbe de DÉDUCTION)+ [Verbe de CONCEPTION] + comme + Y				
Réalisations lexicales	<u>X</u>	(peut)	(donc/alors/ainsi)	<i>est (être) défini/conçu</i> <i>se définit(r)</i>	<i>Comme Y</i>
Éléments périphériques	Marqueurs discursifs comme les adverbes de déduction : <i>donc, alors, ainsi</i> ; les adverbes de concession : <i>au moins</i> ; les adverbes de modalité_ approximation : <i>généralement, de plusieurs manières, de trois manières</i> .				
Paramètres énonciatifs	Dialogismes intralocutif et interlocutif. Cette routine s’inscrit d’abord dans la dimension intralocutive, car elle consiste à préciser le sens d’un terme ou d’un fait scientifique dont discute l’auteur dans son discours. En parallèle, comme toutes les routines de guidage, elle peut être envisagée dans la dimension interlocutive, on parle également d’interlocutif-anticipatif.				
Position textuelle	Dans le corps principal du discours (pas dans la conclusion).				
Fonctions	Donner une définition à un terme, à un fait ou un phénomène scientifiques.				

Tableau 6-6 : Routine N° 1 de précision terminologie (définition)

Routine N° 2 de précision terminologique - définition.

Description sémantique	< l' 'auteur' 'formule' un 'énoncé définitoire Y' pour un 'concept/phénomène/faît scientifique X' >				
Exemples	<i>On définit X comme...</i> <i>Nous pouvons alors définir X comme...</i> <i>Nous pouvons concevoir X comme...</i>				
Type de phrase	Phrase déclarative, souvent au présent, parfois au passé.				
Structure syntaxique	Pronom AUTEUR + (pouvoir) + (le/les)+ (Adverbe de déduction) + [Verbe de CONCEPTION] + X + comme+ Y				
Réalisations lexicales	<i>Nous/on</i>	<i>pouvons/peut (le)</i>	<i>donc/alors/ainsi/e nfin/précisément</i>	<i>définir/conc evoir X</i>	<i>comme Y</i>
Éléments périphériques	Marqueurs discursifs comme les adverbes de déduction : <i>donc, alors, ainsi</i> ; de clarification : <i>précisément, plus précisément</i> ; de conclusion : <i>enfin</i> ; d'exemplification : <i>par exemple</i> ; les adverbes de modalité <i>approximation</i> : <i>grossièrement</i> ;				
Paramètres énonciatifs	Dialogismes intralocutif et interlocutif. Cette routine s'inscrit d'abord dans la dimension intralocutive, car elle consiste à préciser le sens d'un terme ou d'un fait scientifique dont discute l'auteur dans son discours. En parallèle, comme toutes les routines de guidage, elle peut être envisagée dans la dimension interlocutive, on parle également d'interlocutif-anticipatif.				
Position textuelle	Dans le corps principal du discours (pas dans la conclusion).				
Fonctions	Donner une définition à un terme, à un fait ou un phénomène scientifiques.				

Tableau 6-7 : Routine N° 2 de précision terminologie (définition)

Routine N° 3 de précision terminologique - dénomination.

Description sémantique	< L' 'auteur' 'attribue un nom - Y' à un concept/phénomène/faît scientifique X >				
Exemples	<i>On appelle Y...</i> <i>Ce que je nomme Y...</i> <i>... que nous le désignons par le mot...</i>				
Type de phrase	Phrase déclarative, au présent et au futur. Phrase subordonnée relative au présent, au passé.				
Structure syntaxique	pronom AUTEUR + (pouvoir) + [verbe de DÉNOMINATION] +(X) + $\left(\begin{array}{c} \text{adverbes d'ADDITION} \\ \text{adverbes de TEMPORALITÉ_CHRONOLOGIE} \end{array} \right) + Y$				
Réalisations lexicales	<i>On/je/nous (pouvoir)</i>	<i>appeler/nommer/désigner (X)</i>	<i>donc/aussi/également/désormais/jusqu'à maintenant/par la suite/dès lors</i>	<i>Y</i>	
Éléments périphériques	Marqueurs discursifs comme les adverbes de déduction : <i>donc</i> ; les adverbes d'addition : <i>aussi, également</i> . Marqueurs de temps : <i>désormais, par la suite, dès lors, jusqu'à maintenant</i> .				
Paramètres énonciatifs	Dialogismes intralocutif et interlocutif. Cette routine s'inscrit d'abord dans la dimension intralocutive, car elle consiste à préciser le sens d'un terme ou d'un fait scientifique dont discute l'auteur dans son discours. En parallèle, comme toutes les routines de guidage, elle peut être envisagée dans la dimension interlocutive, on parle également d'interlocutif-anticipatif.				
Position textuelle	Dans le corps principal du discours (pas dans la conclusion).				
Fonctions	Donner un nom à un terme, à un fait ou un phénomène scientifiques quand il s'agit d'une phrase déclarative au présent ou au futur (<i>on appelle Y X</i>). Mais elle peut remplir d'autre fonction comme rappeler le terme déjà utilisé auparavant dans le discours : (<i>ce</i>) <i>que nous avons appelé X</i> .				

Tableau 6-8 : Routine N° 3 de précision terminologie (dénomination)

Routine N° 4 de précision terminologique - dénomination.

Description sémantique	< Un 'concept/phénomène/faît scientifique X' reçoit un nom indiqué par un 'terme Y' >		
Exemples	<i>X est désigné par le mot Y</i>		
Type de phrase	Phrase déclarative, au présent.		
Structure syntaxique	X + est désigné par + [nom d'ÉNONCÉ] + Y		
Réalisations lexicales	X	<i>est désigné par</i>	<i>le terme/le mot/la notion Y</i>
Éléments périphériques	Pas d'élément périphérique.		
Paramètres énonciatifs	Dialogismes intralocutif et interlocutif. Cette routine s'inscrit d'abord dans la dimension intralocutive, car elle consiste à préciser le sens d'un terme ou d'un fait scientifique dont discute l'auteur dans son discours. En parallèle, comme toutes les routines de guidage, elle peut être envisagée dans la dimension interlocutive, on parle également d'interlocutif-anticipatif.		
Position textuelle	Dans le corps principal du discours (pas dans la conclusion).		
Fonctions	Donner un nom à un terme, à un fait ou un phénomène scientifiques.		

Tableau 6-9 : Routine N° 4 de précision terminologie (dénomination)

Routine N° 5 de précision terminologique - dénomination.

Description sémantique	< L' 'auteur' 'utilise' un 'terme Y' afin d' 'établir un lien' entre ce terme et un concept/phénomène/faît scientifique X >			
Exemples	<i>J'utilise le terme Y pour désigner...</i> <i>Nous utilisons la notion Y pour désigner...</i>			
Type de phrase	Phrase déclarative, souvent au présent et au futur.			
Structure syntaxique	pronom AUTEUR + { [verbe d'USAGE] + [nom d'ÉNONCÉ] } + Y + pour désigner + X parler de			
Réalisations lexicales	<i>On/je/nous (pouvoir)</i>	<i>utiliser/employer</i> <i>parler de Y</i>	<i>le terme/la notion/l'expression Y</i>	<i>pour désigner</i> X
Éléments périphériques	Très peu de marqueurs : l'adverbe de temporalité : <i>désormais</i> ; de déduction : <i>donc</i> .			
Paramètres énonciatifs	Dialogismes intralocutif et interlocutif. Cette routine s'inscrit d'abord dans la dimension intralocutive, car elle consiste à préciser le sens d'un terme ou d'un fait scientifique dont discute l'auteur dans son discours. En parallèle, comme toutes les routines de guidage, elle peut être envisagée dans la dimension interlocutive, on parle également d'interlocutif-anticipatif.			
Position textuelle	Dans le corps principal du discours (pas dans la conclusion).			
Fonctions	Donner un nom à un terme, à un fait ou un phénomène scientifiques.			

Tableau 6-10 : Routine N° 5 de précision terminologie (dénomination)

Routine N° 6 de précision terminologique - désignation.

Description sémantique	< Un ‘terme Y’ <i>se rapporte à ou indique</i> un concept/phénomène/faît scientifique X>		
Exemples	<i>Le terme « X » renvoie à...</i> <i>Le mot « » désigne...</i> <i>L’expression signifie...</i>		
Type de phrase	Phrase déclarative, au présent.		
Structure syntaxique	[nom d’ÉNONCÉ] + { [verbe de RAPPORT] } _{signifier} + Y		
Réalisations lexicales	<i>Le/l’/la terme/mot/expression/notion</i>	<i>renvoyer à/désigner/signifier</i>	Y
Éléments périphériques	Quelques marqueurs comme l’adverbe d’addition <i>aussi, non seulement, mais aussi</i> . Les adverbes de temporalité_ fréquence : <i>souvent, toujours</i> ; l’adverbe de déduction <i>donc</i> , de réexamen <i>en effet, en ce sens</i> ; d’exemplification <i>par exemple</i> . L’adverbe de modalité_ approximation <i>généralement</i> .		
Paramètres énonciatifs	Dialogismes intralocutif et interlocutif. Cette routine s’inscrit d’abord dans la dimension intralocutive, car elle consiste à préciser le sens d’un terme ou d’un fait scientifique dont discute l’auteur dans son discours. En parallèle, comme toutes les routines de guidage, elle peut être envisagée dans la dimension interlocutive, on parle également d’interlocutif-anticipatif.		
Position textuelle	Dans le corps principal du discours (pas dans la conclusion).		
Fonctions	Donner un nom à un terme, à un fait ou un phénomène scientifiques.		

Tableau 6-11 : Routine N° 6 de précision terminologie (désignation)

6.3.3.1 Caractéristiques syntaxiques, lexicales et sémantiques de ces routines

Les routines de précision terminologique sont très importantes dans les écrits scientifiques. Elles ont pour objet la clarification et le contrôle de l’utilisation des termes. Nous allons présenter à tour de rôle les trois fonctions précises – la définition, la dénomination et la désignation.

Les routines de définition mobilisent principalement deux étiquettes communes : ‘concept/phénomène/faît scientifique X’ et ‘énoncé définitoire Y’. Il s’agit de deux constituants essentiels pour formuler une définition, comme le montrent les exemples ci-dessous (les concepts X à définir et les énoncés définitoires Y sont indiqués en italique et récapitulés dans le tableau, tandis que les expressions qui ont été soulignées correspondent aux marqueurs discursifs) :

[11] Aussi, à ce stade, **nous définissons** *la communication institutionnelle publique (X) comme un ensemble de techniques de légitimation constituant une communication publique politique indirecte (Y)*. (Science de l’information)

[12] Dans une langue comme le français, *la notion syntaxique de sujet (X) que je définis comme le premier argument d’un nom d’action ou d’un verbe (Y)*, en vient souvent à se grammaticaliser comme un possesseur. (Linguistique)

[13] **Nous pouvons alors définir de trois manières au moins** *les proportions démographiques (X) : comme la part d’une classe d’âge dans la population totale (Y), comme la part d’une classe d’âge rapport à la population des 15 à 64 ans (Y) et comme la part dans la population active (Y)*. (Économie)

[14] *Un persona (X) peut ainsi être défini comme « un archétype d'utilisateur, à qui l'on a donné un nom et un visage, et qui est décrit avec attention, en termes de besoins, de buts et de tâches » (Y)* (Blomquist & Arvola, 2002, p. 197). (Psychologie)

[15] *La légalité (X) peut se définir comme ce qui est conforme au droit (Y)*. (Psychologie)

Exemples	X (concepts à définir)	Y(énoncés définitoires)
11	<i>la communication institutionnelle publique</i>	<i>un ensemble de techniques de légitimation constituant une communication publique politique indirecte</i>
12	<i>la notion syntaxique de sujet</i>	<i>le premier argument d'un nom d'action ou d'un verbe</i>
13	<i>les proportions démographiques</i>	1) <i>la part d'une classe d'âge dans la population totale ;</i> 2) <i>la part d'une classe d'âge rapport à la population des 15 à 64 ans ;</i> 3) <i>la part dans la population active</i>
14	<i>Un persona</i>	<i>« un archétype d'utilisateur, à qui l'on a donné un nom et un visage, et qui est décrit avec attention, en termes de besoins, de buts et de tâches »</i>
15	<i>La légalité</i>	<i>ce qui est conforme au droit</i>

Tableau 6-12 : les concepts X à définir et les énoncés définitoires Y dans les exemples 11-15

À l'instar de Rebeyrolle & Tanguy (2000), nous retenons également trois configurations syntaxiques principales : constructions agentive (11, 13), passive (14) et pronominale passive (15), qui correspondent à trois modèles de routines de définition très récurrents. La construction agentive peut aussi subir un statut subordonné dans une proposition comme l'exemple 12. Les verbes mobilisés dans cette classe sont des verbes de conception qui sont liés à l'activité cognitive, dont les plus fréquents sont *définir* et *concevoir*. Les différentes fonctions de ces routines peuvent équivaloir à « précise[r] le sens d'une unité terminologique du domaine » (14), ou encore à « circonscrire un phénomène ou un objet » (Jacques, 2011) comme les exemples 11-13. Ces routines exigent souvent la présence mutuelle de l'objet à définir et de sa définition dans un même énoncé.

Les structures agentives et passives sont toutes très fréquentes dans notre corpus. Leur abondance dans les écrits scientifiques peut s'expliquer par le fait que l'auteur a besoin de recourir à des termes ou des expressions complexes. En ce qui concerne la formulation des routines de définition, deux éléments se présentent mutuellement : l'objet à définir (X) et sa définition (Y).

Nous retenons pour les routines de dénomination 3 modèles qui se construisent souvent à partir de verbes ou de noms métalinguistiques liés aux étiquettes comme '**attribue(r) un nom - Y**', '**terme Y**', '**utilise(r)**' et '**établir un lien**'. L'étiquette '**attribue(r) un nom - Y**' est

réalisée par les verbes de DÉNOMINATION dans la grande catégorie « ANCRAGE-CONSTRUCTION » (cf. section 5.3.2.1) comme *appeler, nommer, désigner*, et celle de ‘**terme Y**’ par les noms d’ÉNONCÉ comme *terme, mot, expression, notion*. L’étiquette ‘**utilise(r)**’ renvoie aux verbes d’USAGE (*employer, utiliser, mobiliser...*). ‘**Établir un lien**’ dans la routine N° 5 < **L**’**auteur**’ ‘**utilise**’ un ‘**terme Y**’ afin d’**établir un lien**’ entre ce terme et un concept/phénomène/fait scientifique X> (l’exemple 17) réfère plutôt à un seul verbe *désigner*. Ce verbe n’appartient pas à la classe des verbes de DÉNOMINATION, mais plutôt aux verbes de RAPPORT qui indiquent « l’état d’avoir un rapport avec quelque chose, avoir pour objet quelque chose » (Yan, 2017, p. 321). Les verbes de RAPPORT sont eux-mêmes classés encore dans la grande catégorie des verbes d’ÉTAT pour renvoyer « à l’état au sens général : une manière d’être, une situation ou condition statique » (Yan, 2017, p. 320) comme *se composer, résider, symboliser, demeurer, figurer, concerner*, etc.⁹⁶ Par conséquent, le verbe *désigner* dans la routine de dénomination N° 5 prend plutôt le sens de ‘renvoyer, référer’ au lieu du sens ‘attribuer un nom’ dans les routines N° 3 < **L**’**auteur**’ ‘**attribue un nom - Y**’ à un concept/phénomène/fait scientifique X> et N° 4 < Un ‘**concept/phénomène/fait scientifique X**’ reçoit un nom indiqué par un ‘**terme Y**’>. C’est aussi pour cette raison que nous avons attribué des étiquettes différentes pour le cas du verbe *désigner* dans les différentes routines.

De plus, on observe une variabilité remarquable au niveau syntaxique (le nom et le terme attribués (Y) et les concepts X sont indiqués en italique, suivis respectivement des notations X et Y, les marqueurs discursifs ont été soulignés.) :

1. Construction agentive (routines N° 3 et N° 5) qui peut apparaître dans la phrase canonique (16-17), la proposition subordonnée (18-19) et la structure clivée (20). L’objet à dénommer peut être un symbole (18), un concept ou un fait scientifique (16, 17, 19) ou un élément concret de la réalité (20).

[16] **On appellera cet âge (X) "âge minimal de départ à taux plein" (Y) par la suite.**
(Économie)

[17] **Nous avons employé le terme « émeutes » (Y) pour désigner les incendies de voitures ou de bâtiments, les affrontements brefs et réitérés avec la police (X).**
(sciences politiques)

[18] - le mode discret : *le X (X) que je nomme clé (Y)* n’a pas d’autre statut que celui de clé. (Linguistique)

[19] Mais entre les prescriptions du système d’obligations matérielles et morales qu’organise *la parenté et la réalité des comportements (X)*, **que j’appelle des**

⁹⁶ Voir plus de détail sur le site LST : <http://lst.demarre-shs.fr/classification> pour les classifications des verbes d’ÉTAT.

« relations de parenté » (Y), il y a un écart sur lequel joue tout un ensemble de facteurs tels que la communication, la fréquence des contacts et leur nature, les échanges de biens et services, la géographie familiale ou encore les relations affectives. (Anthropologie)

[20] *L'expression d'une identité réelle, dans ce cas ethnique, a été la réponse à un problème social : la contradiction était sociale, la mobilisation fut identitaire (X). C'est ce que j'appelle l'anticolonialisme identitaire (Y).* (Histoire)

2. Construction passive (routine N° 4) qui recourt à deux noms d'énoncé *terme* et *mot*, et à un verbe métalinguistique désigné dans le sens 'appeler, signifier, attribuer un nom', qui relève de la classe sémantique « dénomination ».

[21] D'ailleurs, devant le tribunal des chorfa, *le possédé (X) est désigné par le terme khachba (Y)* (litt . : « une planche », « une écorche charnelle ») pour signifier sa pleine absence à lui-même et la présence active de l'esprit qui prend possession de toutes ses facultés vitales.

À l'inverse de la dénomination, la routine N° 6 ayant pour fonction la désignation < Un 'terme Y' 'se rapporte à ou indique' un concept/phénomène/fait scientifique X > se construit autour de verbes de RAPPORT comme *désigner, renvoyer* ou de REPRÉSENTATION *signifier* que nous regroupons tous dans la classe sémantique de verbe de RAPPORT pour des raisons de simplicité. Ces verbes correspondent à l'étiquette conceptuelle 'se rapporte à ou indique'. Un 'terme Y' renvoie comme les routines de dénomination aux noms scientifiques de la classe d'ÉNONCÉ *terme, expression, mot*. Les routines sont en caractère gras, le signe linguistique et sa désignation sont en italique.

[22] **Ce terme signifie** « paysage », dans toute sa dimension sociale et dans le rapport qu'il sous-entend de l'homme au milieu. (Géographie)

[23] Un premier argument en faveur de la thèse ci-dessus est l'existence bien connue de paires du type en N/dans le N, le second terme étant un complément locatif, alors que **le premier terme renvoie à** une action qui fait intervenir N de façon stéréotypique. (Linguistique)

[24] **Le terme zaouïa (Y) désigne** généralement *le lieu de réunion d'une confrérie mystique (X)* et par extension, il signifie aussi la confrérie elle-même. (Anthropologie)

[25] Dans le monde grec, la couverture nuptiale est une pièce d'étoffe, nommée le plus souvent *chlaina (Y)*, **terme qui signifie** à la fois « couverture » et « manteau » (X). (Histoire)

Il est intéressant de noter que la plupart des occurrences dans notre corpus utilisent des noms d'ÉNONCÉ (*terme, expression*) souvent suivis d'un signe linguistique autonome en apposition du sujet comme *ce terme* (22), *le premier terme* (23). En comparant la nature de l'objet direct des verbes dans cette routine, nous remarquons que la routine de désignation peut

créer une relation prédicative signe → signe comme l'exemple 25 : le signe « *chlaina* » est expliqué par deux signes « *couverture, manteau* », ou un rapport prédicatif signe → chose comme l'exemple 24.

6.3.3.2 *Éléments périphériques et paramètres énonciatifs*

Nous identifions plusieurs classes sémantiques d'adverbes de fonctions divers et les récapitulons dans le tableau 6-13⁹⁷ :

Classes générales des adverbess	Sous-Classes sémantiques des adverbess	Routines de définition	Routines de dénomination	Routine de désignation
Discursifs	Déduction	<i>alors, ainsi, donc</i>	<i>donc</i>	<i>donc</i>
	Conclusion	<i>enfin</i>	-	-
	Concession	<i>au moins</i>	-	-
	Clarification	<i>précisément, plus précisément</i>	-	-
	Addition	-	<i>aussi, également</i>	<i>non seulement, mais aussi</i>
	Réexamen	-	-	<i>en effet, en ce sens</i>
	Exemplification	<i>par exemple</i>	-	<i>par exemple</i>
Temporalité	Chronologie	-	<i>désormais, par la suite, dès lors, jusqu'à maintenant</i>	-
	Fréquence	-	-	<i>souvent, toujours</i>
Modalité	Approximation	<i>grossièrement</i>	-	<i>généralement</i>
Manière		<i>de plusieurs manières, de trois manières</i>	-	<i>de la même manière</i>

Tableau 6-13 : les marqueurs identifiés avec les routines de précision terminologique

On constate que les adverbes de DÉDUCTION se retrouvent dans les trois sous-fonctions de routines, sauf la routine N° 4 < Un '**concept/phénomène/faît scientifique X**' reçoit un nom indiqué par un '**terme Y**'> (*X est désigné par le terme Y*). L'adverbe *donc* est notamment fréquent. Il semble que les routines de définition préfèrent les adverbes permettant de marquer la déduction/conclusion et d'éclaircir (*précisément, par exemple*) le sens.

Les routines de définition et les routines de désignation mobilisent plus de types de marqueurs discursifs, ce qui permet de constater qu'elles s'intègrent de manière plus dynamique dans le discours. En général, les fonctions des adverbes discursifs, comme nous l'avons vu dans les routines précédentes, servent d'abord à donner à ces routines une dimension argumentative. Cela montre ensuite le caractère dynamique de l'activité définitoire dans le discours scientifique qui privilégie l'emploi d'un terme ou d'une définition réfléchiée par l'auteur.

⁹⁷ *Grossièrement, de plusieurs manières, de trois manières* ne sont pas inclut dans le LST. C'est nous qui ajoutons après les avoir identifiés dans le corpus.

Il est intéressant de noter ici la présence des adverbes de temporalité dans les routines de dénomination et de désignation :

- pour marquer le moment où l'auteur dénomme un fait/concept scientifique X par un terme Y dans le processus énonciatif : *désormais, par la suite, dès lors, jusqu'à maintenant*. Dans l'exemple 26, l'auteur établit un lien stable et constat entre *la notation et les conjonctions de subordination*, et l'applique dès le moment énonciatif de dénomination (marqué par *désormais*) dans le reste du discours (aussi l'exemple 16 : *on appelle Y X par la suite*). La stabilisation d'emploi d'un concept scientifique à un moment précis dans le texte est une pratique très courante. L'auteur distingue ainsi son point de vue de celui de ses devanciers.

[26] **Nous utiliserons désormais cette notation pour désigner** les conjonctions de subordination. (Linguistique)

- Pour marquer une fréquence d'utilisation d'un terme : *souvent, toujours*.

[27] Aujourd'hui, **le terme désigne toujours** les danseuses, mais aussi les musiciens de mariages [24] (Anthropologie).

Les marqueurs de MODALITÉ et de MANIÈRE explicitent comment l'activité définitoire est entreprise, ou permettent à l'auteur d'atténuer sa responsabilité scientifique. Les adverbes de MANIÈRE comme *de plusieurs manières, de trois manières* fournissent plus d'une possibilité de définir un concept/faît scientifique, *de la même manière* montre le choix identique de l'auteur adopté dans la partie précédente. On suppose que l'auteur scientifique ne s'efforce pas seulement d'intégrer l'activité définitoire dans sa réflexion (marqueurs discursifs), mais aussi de communiquer au lecteur comment il élabore sa définition. Les adverbes d'APPROXIMATION cités dans le tableau 6-13 correspondent aux marqueurs de *Hedges* dans le classement des marqueurs métadiscursif de Hyland (2005) (*cf.* tableau 3-6, Chapitre 3).

[28] **On peut définir grossièrement** une telle activité **comme** le fait de rapporter un événement à un public, au moyen de différentes techniques. (Linguistique)

Les routines de précision terminologique s'inscrivent généralement dans la dimension intralocutive. Elles expriment la réflexion de l'auteur sur les notions disciplinaires. L'auteur mobilise ces routines pour commenter son propre discours en cours, notamment sur le terme ou l'expression, autrement dit, le code langagier du discours. Le contrôle de l'interprétation du code employé constitue un atout contribuant à la réussite d'un écrit scientifique plus intelligible pour le lecteur.

Le contexte nous montre que ces routines peuvent parfois s'inscrire dans la dimension interdiscursive, lorsque les définitions sont reprises des autres chercheurs :

[29] Or le territoire, forestier ou non, **peut être défini comme** un « espace à métrique topographique », par opposition aux réseaux qui relèveraient d'une « métrique topologique [16] ». (Géographie)

[30] D'après ces auteurs, l'autorégulation **peut être définie comme** la capacité de l'enfant à planifier, guider et diriger ses propres comportements en fonction des circonstances. (Psychologie)

Il s'agit ici plutôt de la routine N° 1 < Un '**concept/phénomène/fait scientifique X**' est expliqué par un '**énoncé définitoire Y**' > (*X est défini/conçu comme Y*) qui reprend souvent la définition déjà établie dans la discipline, permettant donc à cette routine d'avoir une propriété interdiscursive.

6.3.3.3 *Fonctionnement rhétorique*

Nous pouvons conclure que les routines de précision terminologique ont pour objectif la création d'un lien définitoire entre deux éléments. Les trois fonctions désignent les trois types de lien définitoire :

- 1) La définition crée un lien entre un concept/fait scientifique et un définissant qui peut être énoncé par l'auteur ou emprunté à quelqu'un d'autre ;
- 2) La dénomination crée un lien métonymique définitoire entre un signe linguistique créé (un terme, un mot, une expression) et un concept/fait scientifique ;
- 3) La désignation crée un lien référentiel entre un concept/fait scientifique et un signe linguistique (un terme, un mot, une expression) déjà existé.

Ces trois fonctions fondamentalement métalinguistiques permettent de créer des relations référentielles stables entre deux éléments dont entre autres, l'expression linguistique. De même que la reformulation et la réparation, on peut aussi constater ici une anticipation de l'auteur. Ce dernier explique des concepts, des faits scientifiques, des termes ou mots éventuellement difficiles à comprendre chez le lecteur. Ces routines, en prêtant attention à la relation entre le signe linguistique utilisé et son référent dans la discipline, permettent de participer à la constitution des « fondements épistémologiques d'une discipline », à côté « des méthodes utilisées, du cadre théorique mobilisé et des institutions garantes du contenu analysé » (Lefebvre, 2006, p. 7).

En outre, nous identifions une fonction de rappel que la routine de dénomination peut remplir dans le contexte, à travers l'actualisation au temps du passé.

[31] Dans ces conditions, **ce que nous avons appelé** la « volatilité d'intervalle », et que Baker nomme « market makers 'bid - ask spreads' », augmente, ce qui tend à faire augmenter la volatilité du prix. (Sociologie)

Ici, l'auteur rappelle le nom « volatilité d'intervalle » dont il a probablement déjà parlé dans la partie précédente du texte, et reprend le concept. Cela permet de créer la cohésion (la reprise d'un terme ici) et la cohérence (lier la phrase avec la logique du texte).

6.3.4. Routines d'emprunt terminologique

Nous ajoutons dans la catégorie métalinguistique une fonction que nous appelons « emprunt terminologique ». Cette fonction porte sur la source d'un terme, d'un mot ou d'une expression énoncée par une personne autre que l'auteur. Elle fait partie de la filiation scientifique. L'auteur s'inscrit ainsi dans une lignée existante basée sur des affinités ou des courants institutionnalisés, lui permettant de « mieux cerner sa propre identité de chercheur » (Grossmann, Tutin, & Silva, 2009, p. 187). Dans les écrits scientifiques, la filiation scientifique reflète « le caractère cumulatif du travail scientifique [en s'appuyant sur] les travaux antérieurs, les critères de reconnaissances et l'autorité des publications reconnues continuent, généralement à bon droit, de justifier l'appui sur les pairs et les prédécesseurs » (*ibid.*, p. 192). Elle couvre une série de pratiques citationnelles telles que la référence à un courant théorique, à un domaine scientifique, à une école de pensée, à un auteur précis et l'emprunt.

L'emprunt terminologique appartient consiste à reprendre des objets scientifiques construits simples⁹⁸ comme des termes, ou des définitions disciplinaires déjà établis dans le domaine scientifique. À l'inverse de la dénomination, où l'auteur crée un nouveau nom et l'attribue à un concept/fait scientifique, l'emprunt d'un terme déjà bien établi dans la discipline permet de s'inscrire dans une filiation scientifique. En outre, l'emploi d'un terme peut économiser l'effort d'interprétation pour le lecteur, notamment pour les articles de recherche dont l'auteur et le lecteur partagent normalement des héritages théoriques ou méthodologiques⁹⁹.

Nous avons identifié deux routines qui peuvent remplir cette fonction, l'une est agentive, l'autre est prépositionnelle.

⁹⁸ Une classe des noms scientifiques.

⁹⁹ Grossmann et *al.* (2009, p. 189) distingue l'utilisation et l'héritage de la filiation scientifique. L'utilisation renvoie au transfert de connaissance pour la traçabilité des références, comme l'emprunt terminologique (*pour reprendre les termes de J. Dubois (1971)*). L'héritage consiste plutôt en l'inscription dans une communauté scientifique, dans ce cas, l'auteur peut communiquer son positionnement, par l'adhésion ou la démarcation, comme *notre modèle s'inscrit dans la lignée de celui proposé par Green et Zhou [1998]. À la différence de ces derniers... (ibid.)*.

Routine N° 1 d'emprunt terminologique.

Description sémantique	< L' 'auteur' 'emprunte' un 'terme/définition' ' à quelqu'un d'autre que l'auteur' >				
Exemples	<i>On reprend le terme de quelqu'un...</i> <i>Nous reprenons la définition de quelqu'un...</i>				
Type de phrase	Phrase déclarative				
Structure syntaxique	pronom AUTEUR + (pouvoir) + [verbe d'USAGE] + [nom d'ÉNONCÉ] + de à + quelqu'un				
Réalisations lexicales	<i>On/je/nous</i>	<i>utiliser/employer</i>	<i>le terme/la notion/l'expression/la définition</i>	<i>de/à</i>	<i>quelqu'un</i>
Éléments périphériques	Peu de marqueurs discursifs (<i>ainsi</i>)				
Paramètres énonciatifs	Dialogismes interdiscursif et intralocutif. Cette routine s'inscrit surtout dans la dimension interdiscursive par le fait qu'elle mobilise un signe linguistique (dénoté à un concept) employé par les pairs. La réutilisation d'un terme établi permet de renforcer le fondement épistémologique du discours de l'auteur dans une discipline. Elle est également intralocutive, car par nature, elle gère le terme employé par l'auteur dans le discours en cours.				
Position textuelle	Dans le corps principal du discours (pas dans la conclusion).				
Fonctions	Réemploi d'un terme, d'un mot, d'expression, d'une définition de quelqu'un d'autre que l'auteur. Fonction secondaire : filiation scientifique.				

Tableau 6-14 : Routine N° 1 d'emprunt terminologie

Routine N° 2 d'emprunt terminologique.

Description sémantique	< un 'terme/définition' 'à quelqu'un d'autre que l'auteur' est 'emprunté(e)' >				
Exemples	<i>Pour reprendre l'expression de quelqu'un...</i> <i>pour reprendre le terme de quelqu'un...</i>				
Type de phrase	Phrase déclarative, proposition prépositionnelle				
Structure syntaxique	pour + reprendre utiliser+ [nom d'ÉNONCÉ] + de à proposé par forgé par + quelqu'un				
Réalisations lexicales	<i>Pour</i>	<i>reprendre</i>	<i>le mot/le terme/l'expression/la notion/la définition</i>	<i>de/à/forgé par/proposé par</i>	<i>quelqu'un</i>
Éléments périphériques	Très peu de marqueurs discursifs (<i>ainsi</i> une fois).				
Paramètres énonciatifs	Dialogismes interdiscursif et intralocutif. Dialogismes interdiscursif et intralocutif. Cette routine s'inscrit surtout dans la dimension interdiscursive par le fait qu'elle mobilise un signe linguistique (dénoté à un concept) employé par les pairs. La réutilisation d'un terme établi permet de renforcer le fondement épistémologique du discours de l'auteur dans une discipline. Elle est également intralocutive, car par nature, elle gère le terme employé par l'auteur dans le discours en cours.				
Position textuelle	Dans le corps principal du discours (pas dans la conclusion).				
Fonctions	Réemploi d'un terme, d'un mot, d'expression, d'une définition de quelqu'un d'autre que l'auteur. Fonction secondaire : filiation scientifique.				

Tableau 6-15 : Routine N° 2 d'emprunt terminologie

6.3.4.1 Caractéristiques syntaxiques, lexicales et sémantiques de ces routines

Ces deux modèles de routines mobilisent des étiquettes conceptuelles presque identiques : 'emprunte(r)', 'terme/définition' et 'à quelqu'un d'autre que l'auteur'. L'étiquette 'emprunte(r)'¹⁰⁰ signifie ici utiliser quelque chose d'autrui, renvoyant aux verbes

¹⁰⁰ L'étiquette 'utilise(r)' dans les routines de précision terminologique, renvoyant aussi aux verbes d'USAGE. Mais les réalisations lexicales pour les deux étiquettes 'emprunte(r)' et 'utiliser' ne sont pas

d'USAGE : *utiliser, mobiliser, emprunter, reprendre*, dont les deux derniers sont les plus fréquents pour les routines d'emprunt terminologique.

Dans les routines, l'étiquette '**terme/définition**' renvoie à la classe de nom d'ÉNONCÉ, mais avec les réalisations lexicales selon les routines. Par rapport aux routines de précision terminologique qui emploient les noms *terme, mot, expression* et qui sont des constituants d'un énoncé et relevant du métalangage grammatical, '**terme/définition**' y ajoute encore deux noms de plus : *définition, notion*. Ces deux noms portent davantage sur l'explication d'un concept. Nous avons donc décidé d'ajouter '**définition**' dans l'étiquette afin de signaler cette différence entre '**terme**' et '**terme/définition**'. La dernière étiquette '**à quelqu'un d'autre que l'auteur**' a pour fonction de préciser la source d'un terme ou d'une définition. Nous soulignons ici le fait que le terme ou la définition ne proviennent pas de l'auteur lui-même, mais de quelqu'un autre. Cette étiquette est donc essentielle pour caractériser l'emprunt.

Du point de vue syntaxique, la première routine d'emprunt terminologique possède une forme agentive (26, 27, 28) et la deuxième routine a une structure infinitive (29) (la routine est mise en gras, la source de l'emprunt est en italique).

[26] **Nous empruntons le terme "cadre" à Goffman** pour dénoter des "schémas d'interprétation" qui permettent aux individus de localiser, de percevoir, d'identifier et de "labelliser" les événements de leur vie et du monde en général. (Sociologie)

[27] **Nous reprenons ici les définitions proposées par K. Lambrecht** pour lequel le topique et le focus sont des relations pragmatiques : "the pragmatic relation between a topic and a proposition is assumed to be predictable or recoverable, the relation between the focus element and the proposition is assumed to be unpredictable or non-recoverable for the addressee at the time of the utterance". (Linguistique)

[28] **Nous utilisons la définition de TOWNSEND [1994]** pour les échelles d'équivalence, mais les résultats ne changent que très peu lorsqu'on utilise d'autres échelles d'équivalence ou simplement le nombre de personnes dans le ménage. (Économie)

[29] L'interrogation identitaire est au cœur des transformations affectant nos sociétés, prises dans le double jeu de la mondialisation libérale des échanges et de « l'unification des marchés symboliques », **pour reprendre l'expression de Patrick Champagne**. (Anthropologie)

Il est intéressant de noter que la première routine peut parfois avoir une réalisation lexicale proche à la routine de dénomination (*on utilise le terme X pour désigner Y*), comme le

identiques, pour indiquer cette différence, nous utilisons deux étiquettes : '**emprunte(r)**' : *utiliser, mobiliser, emprunter, reprendre* ; '**utiliser**' : *utiliser, employer, mobiliser*.

montre l'exemple 26 nous empruntons le terme « cadre » à Goffmann pour dénoter des « schémas d'interprétation ». Cette actualisation lexicale peut exercer en partie une fonction de dénomination. Ici, le terme « cadre » n'est pas un terme créé par l'auteur, mais il provient de quelqu'un d'autre que l'auteur souhaite utiliser pour désigner un fait scientifique.

6.3.4.2 *Éléments périphériques et paramètres énonciatifs*

Contrairement aux routines de précision terminologique, nous trouvons très peu de marqueurs discursifs qui entrent dans les routines d'emprunt terminologique, comme *ainsi* qui marque la déduction. Nous pensons que cette différence est liée au fait que la définition a besoin d'être inscrite dans la réflexion (il s'agit au fond du résultat de la conceptualisation de l'auteur sur un phénomène), alors que l'emprunt d'un terme ou d'une notion est moins attaché à la réflexion de l'auteur.

[30] [26] **On peut ainsi reprendre la définition qu'en donne Alain Lancelot :**
« une disposition relativement persistante à présenter une réaction organisée d'une certaine façon, c'est-à-dire à manifester un certain type de comportement motivé, vis-à-vis d'un objet (ou d'une situation donnée) quand cet objet est en cause » (Alain Lancelot, « L'orientation du comportement politique », dans Madeleine Grawitz, Jean Leca (dir .)) (Sciences politiques)

L'emprunt terminologique relève du dialogisme interdiscursif dans la mesure où l'auteur réutilise les termes ou les notions des pairs. Ces routines renvoient à une caractéristique fondamentale des écrits scientifiques : l'auteur écrit son propre texte en se fondant sur la « mémoire générique » (Amossy, 2005), soit « [l']héritage intellectuel » qui fonctionne comme « un appui terminologique » (Grossmann, Tutin, & Silva, 2009). Ainsi, l'auteur signale son « affiliation à un courant ou à une école, [ce qui l'aide] à définir [son] identité du chercheur, à clarifier l'usage des concepts qu'il utilise » (*ibid.*). Par rapport à la routine de structure infinitive (29), la routine avec un sujet auteur (30) manifeste une intervention plus directe de l'auteur dans le discours scientifique, ce qui lui permet d'indiquer sa filiation scientifique de façon directe et précise. Ici, l'auteur joue explicitement un rôle de chercheur à travers les pronoms personnels (*nous, on, je*).

En parallèle, nous considérons que les routines d'emprunt terminologique peuvent aussi manifester une dimension intralocutive. L'auteur intègre le dit d'autrui dans son discours, afin de l'adapter à son propre dit.

6.3.4.3 *Fonctionnement rhétorique*

Comme nous l'avons remarqué plus haut, la fonction d'emprunt terminologique exerce principalement une fonction métalinguistique en ce qui concerne l'explication d'un terme, d'un

mot ou d'une expression, mais aussi la fonction de la filiation scientifique. Il faut signaler que l'emploi des termes existants dans le domaine scientifique où s'inscrit l'auteur est une pratique très courante. Il permet au lecteur de situer tout de suite le discours de l'autre au sein de son domaine par rapport aux autres recherches déjà effectuées.

6.4. Synthèse

Dans ce chapitre, nous avons observé le fonctionnement et les caractéristiques linguistiques des routines métalinguistiques. Une typologie de 6 types de fonctions métalinguistiques, inspirée d'Ädel (2010), a été établie : fonctions de reformulation, de réparation, de définition, de dénomination, de désignation et d'emprunt terminologique, dont la définition, la dénomination et la désignation sont regroupées dans la catégorie de précision terminologique.

L'analyse des résultats montre que les routines métalinguistiques sont étroitement liées aux propriétés du discours scientifique : l'abondance des termes ou concepts difficiles ou inédits force l'auteur à mobiliser des routines, de définition, de dénomination, de désignation pour mieux expliquer le sens. L'exigence de clarté appelle la reformulation ou la réparation de l'énoncé précédent. Par l'emprunt des concepts et des définitions des pairs, l'auteur annonce sa filiation scientifique et s'intègre dans l'univers discursif de la communauté scientifique visée. Par ailleurs, on observe des structures variées pour ces routines, qui peuvent s'insérer dans plusieurs positions dans une phrase en fonction du besoin. Mais on relève aussi des routines très lexicalisées comme *cela veut dire/signifie que ; pour le dire autrement ; pour reprendre le terme de quelqu'un, etc.*

Bien que les marqueurs discursifs soient moins présents que dans les routines structurantes, on notera tout de même la présence de certains marqueurs de DÉDUCTION, d'ADDITION ou de CONCESSION. Ces marqueurs permettent d'intégrer l'activité métalinguistique dans la dynamique argumentative et montrent que certaines activités métalinguistiques s'inscrivent dans une dimension argumentative très forte et spécifique : ainsi, la définition s'accompagne souvent de marqueurs de DÉDUCTION, ce qui marque cette activité définitoire est bien effectuée grâce à la réflexion de l'auteur.

Parmi les routines métalinguistiques, les routines de précision terminologique sont riches et nombreuses, ce qui montre que la place de la terminologie dans les écrits scientifiques est importante, surtout dans certaines disciplines comme nous le verrons dans le dernier chapitre.

Chapitre 7. Routines de topicalisation : définitions, modélisations et analyses

Les routines de topicalisation visent à mettre en relief les informations essentielles du texte scientifique pour le lecteur. Ces routines manifestent-elles des caractéristiques linguistiques spécifiques par rapport aux routines que nous avons observées dans les chapitres 6 et 7 ? Dans ce chapitre, nous préciserons d'abord la définition de « topicalisation » adoptée dans la présente recherche en examinant des études associées aux marqueurs permettant de parler du thème et de souligner des points essentiels dans un texte. Les sous-fonctions et les routines classées seront analysées par la suite en suivant la même logique que celle des deux chapitres précédents.

7.1. Précision sur le terme « topicalisation »

Dans la tradition française, le concept de topicalisation est plutôt abordé dans une perspective grammaticale ou en linguistique textuelle. Les recherches les plus détaillées et complètes sur ce sujet sont les études de Combettes (1999, 2003, 2006, 2007) de Grobet (2002) et de Prévost (2003). Pour ces linguistes, les marqueurs de topicalisation sont liés au thème ou topique pour évoquer le sujet d'un propos. Ils peuvent être des dispositifs grammaticaux comme des éléments déictiques ou anaphoriques, ou des éléments lexicaux ou phraséologiques comme des locutions prépositionnelles *quant à, à propos de, en ce qui concerne, pour ce qui est de*, ou des structures détachées *ce qui... est de, c'est... qui/que...* (Combettes & Prévost, 2001 ; Grobet, 2002 ; Prévost, 2003). Adam (2005, p. 118–120) évoque le terme de topicalisation pour désigner un type d'organiseurs textuels qui permet de réaliser « le passage d'un objet du discours à un autre [qui] est souligné par des *marqueurs de changement de topicalisation* ». Ces marqueurs comportent des éléments grammaticaux évoqués plus haut, comme *quant à, en ce qui concerne*. Ainsi, dans l'approche de la linguistique textuelle, la topicalisation est envisagée sous l'angle textuel, avec l'objectif de signaler des parties thématiques du texte en collaboration avec d'autres marqueurs textuels, tels que les organisateurs spatiaux, temporels et énumératifs si on reprend les termes d'Adam (*ibid.*)(cf. section 5.1).

De nombreuses études sur le métadiscours en anglais ont aussi abordé la question de la topicalisation. Meyer (1975, cité par Crismore, 1983, p. 9–10) identifie des marqueurs de pointage (*pointer words*) qui ont pour fonction d'informer explicitement le lecteur sur une idée particulière, par exemple, *My first important point is that ERA is not a dead issue*. Ces marqueurs peuvent être employés afin d'attirer l'attention sur l'importance d'une idée ou sur

l'opinion d'un auteur. Williams (1981) considère quant à lui, les marqueurs de topicalisation (*Topicalizers*) en tant que sous-catégorie du métadiscours. Ces marqueurs attirent l'attention sur un sujet particulier dans une proposition, un paragraphe ou une section entière (Crismore, 1983, p. 8). Les *Topicalizers* attirent non seulement l'attention sur le thème, mais aussi sur le changement de thème. Ils comportent donc des expressions comme *in regard to, in the matter of, turning now to*, ou des phrases présentatives comme *there are three reasons why we...* Inspiré par cette description, Vande Kopple (1985) reprend les éléments de cette sous-catégorie dans la classe des Connecteurs de sa propre typologie, en les définissant en tant qu'éléments métadiscursifs qui aident à prêter attention au thème d'un segment textuel, avec des marqueurs comme *with regard to/en ce qui concerne*. Crismore *et al.* (1993) les envisagent aussi comme des connecteurs qui ont pour fonction de signaler le changement du sujet : *now I will discuss/maintenant, je vais discuter*.

Le concept de topicalisation correspond donc à plusieurs définitions et présente un certain flou. Dans cette étude, nous proposons une définition plus restreinte de la topicalisation qui se rapproche davantage du sens de la mise en relief d'une idée ou d'un élément dans un texte, et nous excluons le sens de changement de thème. Plus précisément, nous adoptons un point de vue rhétorique pour éliminer des définitions envisagées dans une perspective grammaticale ou lexicale. En d'autres termes, nous n'intégrons pas des marqueurs qui ont pour fonction d'évoquer ou de changer simplement le thème du propos (*quant à, à propos de, en ce qui concerne*), classés en tant que marqueurs de thème dans les modèles interactifs du métadiscours), ou même des configurations grammaticales de structure clivée qui contribuent à souligner le thème (*c'est... qui/que..., ce qui/que..., c'est...*), étudié par les linguistes françaises comme Combette (2003), Grobet (2002), Carter-Thomas (2009), Carter-Thomas & Jacques (2022). Même si le clivage de la structure d'une phrase peut accorder une « interprétation particulière à une partie de la phrase » (Carter-Thomas, 2009, p. 8), il n'est qu'une stratégie qui propose des alternatives de construction syntaxique d'une phrase¹⁰¹. Nous considérons que les fonctions de topicalisation contribuent à accorder une place plus saillante à un constat, à un fait ou à un argument en direction du lecteur. Pour cette dimension, nous distinguons encore trois sous-fonctions rhétoriques avec sept routines.

¹⁰¹ Dans une certaine mesure, on peut paraphraser une phrase avec le clivage, voire avec le pseudoclivage comme l'observent Carter-Thomas & Jacques (2022) pour le discours académique oral.

7.2. Routines de topicalisation

Nous avons retenu dans cette classe trois sous-fonctions, à savoir la mise en évidence, le (co)constat et le focus, comme le montre la figure 7-1. La mise en évidence permet de souligner une information saillante pour le lecteur. Le (co)constat met l'accent sur le constat fait par l'auteur. Le focus, comme son nom l'indique, se focalise sur le thème essentiel à aborder dans un lieu textuel précis. Même si ces trois sous-fonctions présentent quelques différences au niveau fonctionnel et sémantique, leur noyau fonctionnel demeure toujours la mise en évidence d'une information saillante qui est minutieusement sélectionnée par l'auteur pour le lecteur. En parallèle, les champs lexicaux exploités par les routines présentent également de nombreux points communs.

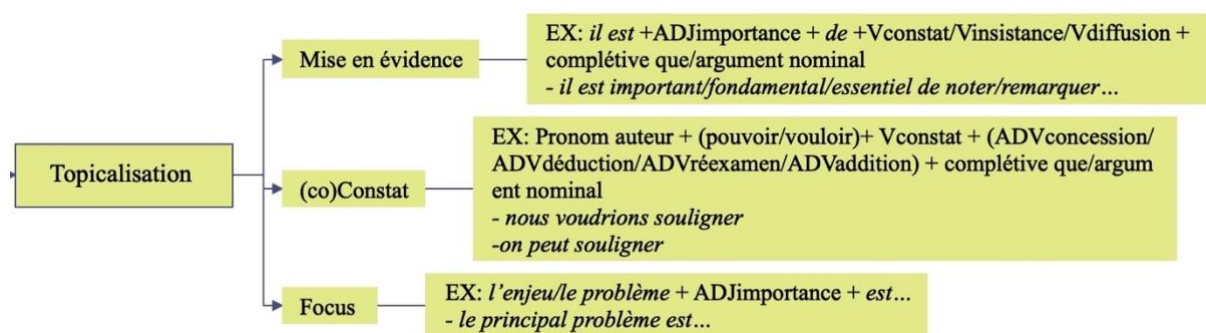


Figure 7-1 : Les sous-fonctions de la topicalisation

Comme les chapitres 5 et 6, les sous-classes de la fonction de topicalisation seront présentées dans la section courante avec leurs définitions et caractéristiques linguistiques.

7.2.1. Routines de mise en évidence

La mise en évidence est la fonction la plus représentative de la dimension de topicalisation. Elle hiérarchise les statuts des éléments dans le texte et sélectionne un élément qui mérite une attention particulière. Cet élément peut être une information, un constat, un fait ou un objet scientifiques. Nous avons identifié trois routines, dont deux avec une structure impersonnelle.

Routine N° 1 de mise en évidence.

Description sémantique	< L' 'auteur' 'met en évidence' une information pour son lecteur >				
Exemples	<i>Je voudrais souligner...</i> <i>Nous pouvons mentionner que...</i>				
Type de phrase	Phrase déclarative				
Structure syntaxique	Pronom AUTEUR + (pouvoir/vouloir) + {verbe de DIFFUSION} + (adverbe discursif) + {SN complétive que}				
Réalisations lexicales	<i>On/je/nous</i>	<i>peut(x)/pouvons /veut(x)/doit</i>	<i>néanmoins/ enfin/</i>	<i>souligner/mentionner/indiquer/insister sur (le fait)/insister sur l'importance</i>	SN/que
Éléments périphériques	<p>Marqueurs discursifs : les adverbes de conclusion : <i>enfin, pour conclure, pour finir</i> ; de déduction : <i>donc, ainsi</i> ; de concession : <i>néanmoins, certes, cependant</i> ; d'opposition : <i>en opposition</i> ; d'addition : <i>également, aussi, d'ailleurs, de plus, de surcroît</i> ; de réexamen : <i>en effet</i> ; de clarification : <i>plus précisément</i>, de topicalisation : <i>concernant, quant à</i>.</p> <p>Marqueurs modaux : les adverbes de restriction : <i>seulement, simplement</i> ; d'approximation : <i>en gros, de la manière générale</i>.</p> <p>Marqueurs de localisation <i>ici, y, là encore</i>.</p>				
Paramètres énonciatifs	Dialogisme interlocutif. L'auteur sélectionne l'information essentielle pour le lecteur.				
Position textuelle	Dans le corps principal du discours (pas de position spécifique).				
Fonctions	Souligner l'importance d'une information pour le lecteur.				

Tableau 7-1 : Routine N° 1 de mise en évidence

Routine N° 2 de mise en évidence.

Description sémantique	< Une information/un constat s'avère 'essentiel(le)' méritant d'être 'mis(e) en évidence' >					
Exemples	<i>Il est important de souligner...</i> <i>Il est fondamental de noter...</i>					
Type de phrase	Phrase déclarative, impersonnelle					
Structure syntaxique	Il est + (adverbe discursif) + [adjectif d'IMPORTANTANCE] + de + {verbe de CONSTAT} + {SN complétive que}					
Réalisations lexicales	<i>Il est</i>	<i>plus/également/alors/particulièrement/donc très...</i>	<i>important/fondamental/ essentiel/intéressant</i>	<i>de</i>	<i>souligner/mentionner/indiquer/noter/constater/remarquer</i>	SN/que
Éléments périphériques	<p>Marqueurs discursifs : les adverbes de conclusion : <i>enfin</i> ; de déduction : <i>donc</i> ; de concession : <i>néanmoins, toutefois</i> ; d'addition : <i>de plus, par ailleurs</i> ; de réexamen : <i> finalement</i> ; d'énumération : <i>ensuite</i>, de localisation : <i>ici</i> ; de topicalisation : <i>à ce propos, à cet égard</i>.</p>					
Paramètres énonciatifs	Dialogisme interlocutif. L'auteur sélectionne l'information essentielle pour le lecteur.					
Position textuelle	Dans le corps principal du discours (pas de position spécifique).					
Fonctions	Souligner l'importance d'une information pour le lecteur.					

Tableau 7-2 : Routine N° 2 de mise en évidence

Routine N° 3 de mise en évidence.

Description sémantique	< Une information/un constat ‘nécessite’ d’être ‘mis(e) en évidence’ >				
Exemples	<i>Il faut souligner...</i> <i>Il convient de noter...</i> <i>Il est à noter...</i> <i>Il importe de remarquer que...</i>				
Type de phrase	Phrase déclarative, impersonnelle				
Structure syntaxique	II + { [verbe de NÉCESSITÉ] } + (adverbe discursif) + { verbe de CONSTAT } + { SN } + { complétive que }				
Réalisations lexicales	<i>Il</i>	<i>faut/convient de/importe de/est à</i>	<i>donc/enfin/en effet/</i>	<i>souligner/mentionner/indiquer/note r/constater/remarquer/mettre en évidence...</i>	<i>SN/que</i>
Éléments périphériques	Marqueurs discursifs : les adverbes de conclusion : <i>enfin, pour conclure, pour finir</i> ; de déduction : <i>donc, ainsi</i> ; de concession : <i>cependant, pourtant, pour autant, néanmoins, toutefois</i> ; d’addition : <i>aussi, également, d’ailleurs, de plus, par ailleurs</i> ; de réexamen : <i> finalement, en effet, effectivement</i> ; d’énumération : d’énumération : <i>d’abord, tout d’abord, ensuite, en premier lieu, en second lieu</i> ; de localisation : <i>ici</i> ; de topicalisation : <i>concernant, à ce propos, à cet égard, à ce titre, sur ce point, dans ce dernier cas, dans ces situations particulièrement, pour ce point.</i> Marqueurs de temporalité : les adverbes de chronologie : <i>à nouveau, dès maintenant, dès le présent.</i>				
Paramètres énonciatifs	Dialogisme interlocutif. L’auteur sélectionne l’information essentielle pour le lecteur.				
Position textuelle	Dans le corps principal du discours (pas de position spécifique).				
Fonctions	Souligner l’importance d’une information pour le lecteur.				

Tableau 7-3 : Routine N° 3 de mise en évidence

7.2.1.1 Caractéristiques syntaxiques, lexicales et sémantiques de ces routines

Ces trois routines partagent l’étiquette ‘**met(tre) en évidence**’ ou ‘**mis(e) en évidence**’ qui est employé pour faire ressortir une information saillante. Cette étiquette renvoie à trois classes de verbes : celles de DIFFUSION et d’INSISTANCE qui appartiennent aux verbes de COMMUNICATION, et celle de CONSTAT qui est intégrée dans la grande catégorie des verbes d’ANALYSE_INFO. La routine N° 1 < L’‘**auteur**’ ‘**met en évidence**’ une information pour son lecteur > manifeste une particularité par rapport aux autres routines en matière des paradigmes lexicaux des verbes qui réalisent l’étiquette ‘**met en évidence**’. Il s’agit ici plutôt des verbes de COMMUNICATION, en particulier les deux classes qui sont en jeu : la classe de diffusion (les verbes retenus sont *indiquer, mentionner, préciser, signaler*) et la classe d’INSISTANCE (*insister, souligner*). Les routines N° 2 et N° 3, avec ces deux sous-classes citées précédemment, incluent aussi la classe des verbes de CONSTAT, comme *constater, noter, observer, relever, repérer, voir, remarquer*.

Les classes de verbes utilisées dans la routine N° 1 diffèrent de celles des autres routines. Les verbes de CONSTAT permettent l’appréciation d’un constat. Ils ne servent pas à rendre plus saillante une information, mais à rendre compte d’une information. Dans les deux structures impersonnelles, c’est grâce aux adjectifs d’IMPORTANCE (représentés par

l'étiquette 'essentiel(le)', routine N° 2), aux **verbes de NÉCESSITÉ** (représentés par l'étiquette 'nécessite(r)', routine N° 3) et à la structure syntaxique *il est à noter* (routine N° 3) que le sens de mise en évidence est réalisé. Néanmoins, dans la routine N° 1, le sens de la mise en évidence est véhiculé et renforcé par les verbes eux-mêmes (*signaler, souligner...*). Par ailleurs, les routines impersonnelles peuvent remplir une seconde fonction de constat quand elles se réalisent avec les verbes de CONSTAT.

Sur le plan syntaxique, les routines de mise en évidence présentent un point commun en ce qui concerne les compléments de la phrase. Elles comportent deux compléments : un syntagme nominal (les exemples 1-3) ou une proposition subordonnée introduite par *que* (4-6). Dans les exemples, les routines sont mises en gras, les compléments de phrases sont en italique, les marqueurs sont soulignés.

- [1] Enfin, **il est important de souligner** *le poids* que pèse, dans le processus adoptif, la germanité prise, ici, dans toutes ses expressions de sexe, d'âge, de biologique et de social. (Anthropologie).
- [2] Concernant les régressions multiples, **il faut remarquer** les variances relativement faibles. (Psychologie).
- [3] Au-delà du caractère anecdotique de cette agression, **nous voudrions souligner** *les injonctions contradictoires* auxquels peuvent être confrontés les agents de sécurité. (Sociologie).
- [4] Finally, **il est important de souligner** *que le nombre de participants à cette étude n'était pas très conséquent (n=181)*. (Psychologie).
- [5] **Toutefois, il est fondamental de noter** *que la partie fractale existe dans la partie non fractale, mais elle ne « s'exprime » pas ; de même, la partie non fractale existe dans la partie fractale, mais elle n'est pas apparente d'où les pointillés sur la figure 1*. (Géologie).
- [6] **Il est également à noter** *que le rite engage ses participants à mobiliser tout à la fois leurs ressources émotionnelles et leurs facultés intellectuelles*, dans la mesure où il se donne à vivre, mais aussi à interpréter, comme tout dispositif symbolique possédant un sens caché derrière le sens apparent. (Science de l'information)

La routine intégrant un complément sous forme de SN est fréquente parmi les occurrences. Elle vise à mettre en évidence la valeur d'importance des idées ou des informations. Dans l'exemple [1], la routine *il est important de souligner* indique l'influence de la germanité - un élément important dans le processus parmi plusieurs éléments caractérisant la parenté comme le sexe, l'âge, les facteurs biologiques et sociaux. De ce fait, avec cette routine, l'auteur sélectionne par souci de clarté une idée fondamentale pour son lecteur.

La construction syntaxique avec la complétive en *que* est également fréquente dans notre corpus. Le fait qu'elle soit suivie d'une phrase complétive facilite la construction d'une idée ou d'une information plus complexe. Dans l'exemple [4], l'auteur met en relief l'information selon laquelle le nombre de participants ne constitue pas un facteur décisif pour influencer le résultat de son étude. Cette précision dissipe le doute éventuellement généré chez le lecteur, ce qui rend la lecture plus intelligible.

Nous identifions également quelques réalisations avec une structure clivée :

[7] D'autre part, aux racines de l'art moderne, **c'est bien une incessante manipulation sur le national et l'international qu'il faut remarquer**, rendant impossible toute approche de l'histoire de l'art en termes d'art national comme international. (Histoire).

La focalisation de l'élément extrait permet de spécifier cet élément et laisse comprendre que cet élément constitue la seule information nouvelle introduite par le clivage. Par conséquent, quand cette routine se forme en structure clivée, elle va renforcer l'important de l'élément extrait. Dans l'exemple, le complément *une incessante manipulation sur le national et l'international* est focalisé.

7.2.1.2 *Éléments périphériques et paramètres énonciatifs*

Il faut signaler que les routines de mise en évidence coexistent souvent avec des marqueurs discursifs, qui sont des éléments indispensables pour la cohésion et la cohérence du texte (Tran, 2014, p. 28), comme le montrent les exemples ci-dessous.

[8] Néanmoins, **il est important de souligner** que cette validation a été réalisée avec des lycéens âgés de 12 à 18 ans. (Psychologie).

[9] Ce que nous avons pu montrer à travers l'étude du texte de Jolicoeur repose sur une situation **qu'il est important de souligner**, en conclusion, pour indiquer en quoi ce texte garde une résonance moderne. (Histoire).

La présence des marqueurs discursifs permet à une routine de topicalisation de remplir certaines fonctions rhétoriques et argumentatives au-delà de la fonction de mise en relief. Dans l'exemple 8, l'auteur utilise *néanmoins, il est important de souligner que* pour introduire une nouvelle information qui mérite aussi une attention dans le texte. Cette information constitue à la fois un argument complémentaire (compléter l'information de l'âge) et un contre-argument (exclure la généralisation de cette validation pour les autres tranches d'âge). Dans l'exemple 5, le connecteur *toutefois* fournit au constat mis en avant par la routine *il est fondamental de noter que* un sens d'opposition. *Enfin* et *finalement* dans l'exemple [1] et l'exemple [4] servent à introduire un élément comme étant le dernier de l'argumentation, et signalent en même temps

le statut important de celui-ci, tandis qu'*en conclusion* dans l'exemple [9] ajoute une fonction de récapitulation à la routine.

Il est intéressant de souligner la présence des marqueurs de topicalisation renvoyant à un sujet ou un thème précis. Dans les exemples 10 et 11, l'auteur se focalise d'abord sur le thème à traiter, puis met en évidence l'information essentielle concernant ce thème.

[10][12]Sur ce point, **il faut noter** une certaine étanchéité entre les différentes activités ou contributions de ce monde agricole. (Science de l'information)

[11]Dans ce dernier cas, **il faut remarquer que**, si les familles des chefs étaient ngunies, la masse de la population était maconde. (Histoire)

Sur le plan énonciatif, la routine N° 1 < L' 'auteur' 'met en évidence' une information pour son lecteur > s'actualise à travers les pronoms personnels *je*, *on* et *nous* et les verbes modaux (*vouloir*, *pouvoir*). Les pronoms dans cette routine renvoient à l'auteur, parce que l'acte de mise en évidence d'un élément dans le discours ne peut qu'être réalisé par l'auteur, ce qui est différent des routines de (co)constat qui implique souvent une inclusion du lecteur (nous reviendrons sur ce point dans la section 8.2.1.1.). L'ajout de verbes modaux (*vouloir*, *pouvoir*) aide à nuancer les formulations de l'auteur devant son lecteur. *Pouvoir* adoucit d'abord l'affirmation et crée un espace dialogique en accordant la liberté au lecteur de confirmer ou de réfuter l'information mise en évidence (Barbet, 2012, p. 53), tandis que *vouloir* renforce l'intention de mise en évidence, tout en évoquant implicitement la demande d'autorisation (Desclés, 2012, p. 51) au lecteur.

Néanmoins, la phrase impersonnelle n'intègre pas de traces énonciatives explicites. L'actualisation dans ce cas se réalise davantage dans la partie étendue des routines, c'est-à-dire au niveau du complément d'objet. Dans les exemples ci-dessous (10 et 11), *la jeune artiste*, *les observations* sont bien des anaphores qui s'actualisent dans le complément d'objet, et permettent d'inscrire toute la phrase dans le processus énonciatif, parallèlement aux marqueurs discursifs, aux locutions adverbiales (*à cet égard*), voire aux déictiques (par exemple *ici*).

[10] **Il est à noter que** *la jeune artiste* a su, parallèlement, se faire aider pour des tâches plus administratives par une association ayant pour mission d'aider les créateurs à monter des projets. (Sociologie)

[11] À cet égard, **il est important de noter que** *les observations* concernant le vent ne répondaient pas à un désir de connaissance scientifique, mais à des considérations pratiques. (Histoire)

Les routines de mise en évidence relèvent clairement du dialogisme interlocutif, en hiérarchisant les informations et sélectionnant les plus importantes pour le lecteur potentiel. Par

ailleurs, les marqueurs peuvent jouer un rôle important pour renforcer l'appartenance à la dimension interlocutive, en dehors de leurs fonctions argumentatives. Par exemple, les adverbes de concession *toutefois*, *cependant*, *néanmoins* constituent un lieu argumentatif. Ils dépassent la simple fonction de mise en évidence, et renforcent le trait interlocutif de la routine, la concession étant basée sur la déduction inverse de celle du lecteur à partir d'« un monde de connaissances, de valeurs et de raisonnements partagés entre les locuteurs » (Marchello-Nizia, 2007). Imaginons dans la relation entre l'information A et l'information B que B est la déduction de l'auteur d'après A, mais B est l'inverse des attentes du lecteur, ce qui crée un effet de surprise (*ibid.*). Dans ce cas précis, l'auteur anticipe cette surprise du lecteur, en attachant sa propre déduction (B) à A afin d'aboutir à son propre constat, auquel il invite le lecteur à adhérer malgré la surprise. Cet effet de surprise a été également observé dans les routines de réparation (*cf.* Chapitre 6).

7.2.1.3 *Fonctionnement rhétorique*

Rappelons que la fonction de mise en évidence consiste à souligner une information essentielle pour faciliter l'interprétation des informations abondantes dans le discours scientifique. Cette fonction a été conçue comme une classe de « métatextèmes qui soulignent l'importance de l'information » par Skubis (2016, p. 91–92), comme *on doit souligner*, *faites attention*, *on doit remarquer*, *ce qui est le plus important*. Au lieu de mettre la fonction de mise en évidence dans la catégorie métatextuelle comme Skubis (2016), nous préférons la placer, comme d'autres fonctions de topicalisation que nous allons voir, dans la catégorie « Topicalisation », étant donné que ces routines ne visent pas à l'organisation textuelle (structuration textuelle) ni au contrôle de la clarté en utilisation de la langue ou les termes (fonctions métalinguistiques), mais visent à attirer l'attention sur une information particulière.

Les routines à structure impersonnelle, au travers de la réalisation par les verbes de CONSTAT, exercent une fonction auxiliaire du constat. Elles mettent en relief le constat après l'activité de l'observation.

7.2.2. Routines de (co)constat

Le (co)constat signifie que l'auteur se rend compte d'un fait ou d'un constat et le présente au lecteur. À la suite de Grossmann et Tutin (2010), nous utilisons le terme (co)constat pour inclure le cas où l'auteur invite son lecteur à constater le fait avec lui pour le rallier à sa cause. Le (co)constat est accompagné d'une valeur évidentielle.

Les routines dans ce type ont souvent l'auteur (rarement l'auteur seul) ou l'auteur/lecteur comme sujet, avec l'emploi du pronom *on*, *nous* (les deux derniers sont généralement inclusifs).

Nous avons retenu une routine de (co)constat qui possèdent une structure syntaxique similaire à la routine N° 1 de mise en évidence < L’**‘auteur’ ‘met en évidence’** une information pour son lecteur>. Nous allons voir dans la section suivante que la différence entre ces routines n’est pas considérable, et que ces deux routines partagent beaucoup de similarités.

Routine de (co)constat.

Description sémantique	< L’ ‘auteur’ ‘fait un constat’ (ou invite le lecteur à faire un constat) après une observation >				
Exemples	<i>J’ai pu constater...</i> <i>Nous pouvons noter que...</i>				
Type de phrase	Phrase déclarative, souvent au présent.				
Structure syntaxique	(comme) + Pronom AUTEUR + (pouvoir vouloir) + [verbe de CONSTAT] + $\left(\begin{array}{c} \text{adverbe discursif} \\ \text{dans le(a/es)+ nom de GRAPHIQUE} \end{array} \right) + \left\{ \begin{array}{c} \text{SN} \\ \text{complétive que} \end{array} \right\}$				
Réalisations lexicales	<i>On/je/nous</i>	<i>peut(x)/pouvons/v eut(x)</i>	<i>néanmoins/ enfin/dans le tableau</i>	<i>noter/constater/observer/relever/ remarquer/repérer/(le) voir</i>	<i>SN/que</i>
Éléments périphériques	Marqueurs discursifs : les adverbes de concession : <i>cependant, pourtant, certes, toutefois, néanmoins, malgré tout</i> ; de déduction : <i>donc, ainsi</i> ; de réexamen : <i>en fait, finalement</i> ; de conclusion : <i>enfin, donc</i> , d’énumération : <i>tout d’abord, ensuite</i> ; d’exemplification : <i>par exemple, en particulier</i> ; de réexamen : <i>effectivement, en effet, en réalité, en fin de compte, finalement</i> . Marqueurs de localisation : <i>dans le tableau, dans la figure...</i>				
Paramètres énonciatifs	Dialogisme interlocutif. L’auteur communique au lecteur un constat qu’il a fait dans le texte et l’inviter à le partager.				
Position textuelle	Dans le corps principal du discours (pas de position spécifique).				
Fonctions	Faire un constat pour le lecteur ; inviter le lecteur à partager un constat ; marquer l’évidentialité d’une inférence.				

Tableau 7-4 : Routine de (co)constat

7.2.2.1 **Caractéristiques syntaxiques, lexicales et sémantiques de ces routines**

Cette routine possède une nouvelle étiquette conceptuelle **‘fait(re) un constat’** qui donne lieu à un constat, après avoir observé ou déduit des informations. La construction syntaxique de la routine de (co)constat est presque identique à la routine N° 1 de mise en évidence. Cependant, cette unité n’est qu’apparente. Le prédicat dans la routine de (co)constat s’actualise par les verbes de **CONSTAT**, une sous-classe des verbes d’**ANALYSE_INFO**. Ces verbes montrent que, en dehors de la présentation des démarches ou des processus, l’auteur de l’écrit scientifique doit aussi « mettre en scène une opération intellectuelle d’analyse de données plus que de simple perception ou d’activité graphique » (Grossmann, 2014a, p. 760).

Les verbes constatifs sont très présents dans notre corpus. Cependant, ils n’apparaissent pas exclusivement dans les routines de topicalisation (la mise en évidence et le (co)constat), mais aussi dans la navigation intratextuelle (cf. section 6.3.3), à titre d’exemple, la routine N° 2 de navigation textuelle < L’**‘auteur’** renvoie le lecteur à **‘voir’** une information (dans un **‘lieu textuel’**)> (*comme nous avons noté, on va voir, on l’a vu*). Dans cette routine de navigation, ils remplissent davantage une fonction primaire métatextuelle et une fonction secondaire

d'évidentialité¹⁰², en particulier avec la participation des marqueurs renvoyant aux graphiques (*comme nous avons vu dans la figure/tableau...*). Dans ce cas, les verbes constatifs sont souvent employés au passé et au futur pour signaler ce qui est déjà constaté ou ce que va l'être (Grossmann, 2014 b, p. 87). Quant aux routines de topicalisation, il s'agit plutôt du « constat asserté », pour reprendre le terme de Grossmann (2014 b, p. 87). Ce constat asserté remplit :

une fonction plus centrale pour l'exposition des résultats que les structures méta-énonciatives en *comme*, en ce qu'ils permettent de mettre en scène le résultat d'une démonstration ou de l'observation d'éléments empiriques, [...], de recherche d'informations. (*ibid*)

En un mot, ce type de routines a souvent pour fonction de fournir un constat après des réflexions intellectuelles sur des activités scientifiques. De plus, l'auteur exprime ici une opération de co-constat : il prend en compte la présence de ses lecteurs en les invitant à faire un constat avec lui, ce qui montre l'intention d'intégrer son lecteur (voir les exemples 12-15).

[12] **Nous pouvons constater** dans ce texte **que** la prescription s'exprime de manière explicite à travers des verbes à valeur exhortative. (Linguistique)

[13] **On peut remarquer que** le taux de taxation total est identique dans les deux pays. (Économie)

[14] Tout d'abord, **on constate** une différenciation plus nette entre la côte ouest (flux majoritaires sud-nord) et la côte est (flux majoritaires nord-sud). (Géographie)

[15] Chez les participants vicariants (n=41), **on peut noter que** 33 (80 %) n'effectuent qu'un changement de stratégie et que 8 (20 %) effectuent deux changements. (Psychologie)

[16] **Notons également que** dans la cosmogonie de la confrérie, les chorfa sont représentés par la couleur blanche et symbolisent les esprits venant de l'Arabie. (Anthropologie)

Cette routine a aussi trois structures : avec un syntagme nominal (14), une phrase subordonnée introduite par *que* (12,13,15) ou une construction impérative (16).

Grossmann (2014 b, p. 89) dans son étude sur les verbes constatifs distingue des énoncés constatifs forts comme *nous constatons une légère remontée de la température, on voit que...*, et des énoncés constatifs faibles comme les incisives (*comme*) *on le voit/observe*. Dans notre étude, nous avons repéré les énoncés forts, mais aussi les énoncés faibles. Les énoncés faibles du type (*comme*) *on le voit/observe* peuvent remplir plusieurs fonctions, telles que la fonction cohésive en répétant un constat déjà énoncé et la fonction d'évidentialité, notamment quand

¹⁰² Nous sommes bien consciente que l'évidentialité est étroitement liée au constat, mais dans les exemples cités, leur fonction primaire consiste à rappeler ou annoncer (classés par nous comme la navigation intratextuelle).

cette phrase fait appel au déictique de lieu, mais nous n'avons retenu que des occurrences dans lesquelles la fonction de constat prime.

[17] **On le voit**, *ces différents ajouts ne constituent pas des acquis définitifs, mais plutôt des enrichissements, des complications de la problématique initiale de l'école dont chaque condition se définit en tension avec les autres. (Science de l'éducation)*

[18] **Comme on le voit**, *la catégorisation des objets " naturels " ne va pas sans problème et nécessite qu'on prenne en considération un contexte d'anticipations sémantiques selon des modalités bien distinctes. (Psychologie)*

[19] **On peut le voir** à différentes propriétés. (Linguistique)

[20] **On l'observe** en particulier dans le contraste entre □ un gros buveur et □ un grand buveur, d'un emploi plus contraint, voire même perçu a priori comme moins naturel. (Linguistique)

Les exemples 17-20 montrent la différence entre la fonction de constat et la fonction cohésive : dans les exemples 17 et 18, les parties en italiques sont bien les constats donnés par l'auteur, tandis que dans les exemples 19 et 20, les compléments circonstanciels à *différentes propriétés, dans le contraste...* ne constituent pas un constat, mais un lieu où l'on peut trouver ce constat déjà énoncé et répété ici par le pronom neutre *le* à une valeur anaphorique.

7.2.2.2 *Éléments périphériques et paramètres énonciatifs*

Nous nous intéressons ici, en particulier, à l'utilisation des pronoms personnels renvoyant à l'auteur dans la routine. Dans les écrits scientifiques, *je* se rapporte seulement à l'auteur tandis que *nous* et *on* peuvent renvoyer à plusieurs référents. Ils peuvent être inclusifs en impliquant l'auteur et le lecteur, mais aussi exclusifs en se rattachant seulement à l'auteur, ou les autres. Dans la routine N° 1 de mise en évidence, *nous* et *on* désignent uniquement l'auteur, mais dans la routine de (co)constat, ces deux pronoms personnels correspondent aux pronoms inclusifs. Le cas des pronoms *on* et *nous* est complexe. Nous reviendrons sur ce point dans le chapitre 8.

Comme les routines de mise en évidence, cette routine apparaît souvent avec des marqueurs discursifs pour les intégrer dans l'argumentation du discours scientifique. Les adverbess de CONCESSION (*cependant, pourtant, certes, toutefois, néanmoins, malgré tout*), de DÉDUCTION (*donc, ainsi*), de RÉEXAMEN (*en fait, finalement, effectivement, en effet, etc.*), de CONCLUSION (*enfin, donc*), d'ÉNUMÉRATION (*tout d'abord, ensuite*), d'EXEMPLIFICATION (par exemple, en particulier) sont les plus fréquents.

[21] Malgré tout, **on peut constater que** deux cités, la marchande et l'industrielle, demeurent potentiellement réconciliables, par la possibilité de développer des

échelles de mesure des états de grandeur, et ce malgré des temporalités très différentes.

[22] **On peut ainsi relever qu'**une très faible proportion de ces jeunes fait partie d'associations ou de structures à caractère militant (2 %), ou encore de mouvements d'éducation populaire (3 %). (Sociologie)

Les marqueurs discursifs placent les routines de (co)constat dans le déroulement argumentatif et permettent de montrer que le constat s'appuie sur une réflexion.

La routine de (co)constat appartient également à la dimension interlocutive du dialogisme.

7.2.2.3 *Fonctionnement rhétorique*

En général, la routine de (co)constat vise à présenter un constat et inviter le lecteur à s'allier à l'auteur. En parallèle, comme nous l'avons indiqué, elle peut aussi remplir une deuxième fonction d'évidentialité, notamment avec la présence des compléments circonstanciels de lieu, autrement dit, les éléments péritextuels comme des tableaux, des figures, des graphiques. Dans ce cas, le (co)constat est associé aux éléments visuels et permet de prendre « à témoin dans l'observation de la preuve, [...] renforcée par l'indication fréquente du lieu d'observation, exprimé par un déictique de lieu (*ici, plus haut*) ou complément nominal locatif (*dans cet exemple, sur ce tableau*) » Grossmann et Tutin (2010, p.7). Dans ce cas précis, ces routines ont aussi une fonction secondaire de renvoi intratextuel. C'est aussi la raison pour laquelle la routine de navigation intratextuelle (*comme nous l'avons constaté, on va le voir*) a une construction syntaxique et sémantique très proche des routines de (co)constat.

[23] **On peut cependant remarquer sur le tableau 6** a que lorsqu'on compare les mêmes points de vue issus des deux situations, on n'obtient alors pas de différence significative. (Psychologie)

Dans l'exemple 23, l'auteur et le lecteur peuvent tous les deux faire un constat en observant le tableau fourni dans le texte, car le contenu s'avère aussi évident pour le lecteur grâce à sa présentation en tableau.

7.2.3. **Routines de focus**

Ädel (2006, p. 67) envisage la fonction de focus comme l'une des fonctions métatextuelles (*cf.* tableau 3-7). Le focus est défini comme la focalisation qui porte à nouveau sur un thème déjà abordé dans le texte (*ibid.*), par exemple, *Now I come to the next idea which I presented in the beginning (je vais passer à l'idée prochaine que j'ai présenté au début)*. De notre côté, nous n'avons pas repris cette définition, mais défini la fonction de focus comme une fonction permettant de souligner le thème en question. Cognitivement, cette fonction met

l'accent sur ce qui mérite d'attirer l'attention. Mais l'élément souligné ici n'est plus un constat ou une information, mais bien le thème qui est abordé. Cette classe ainsi définie ressemble aux marqueurs de topicalisation (*Topicalizers*) de Williams (1981). Nous avons exclu la fonction de changement du thème (*in regard to, turning now to*), mais nous avons retenu la fonction qui consiste à attirer l'attention sur le thème spécifique. Deux routines seulement de ce type ont été relevées, ce qui est dû au fait que le discours scientifique emploie beaucoup de marqueurs de topicalisation courts comme *en termes de, à ce sujet, à cet égard, à propos de*, etc. (Tran, 2014, p. 297).

Routine N° 1 de focus.

Description sémantique	< Le ' point ' ' essentiel ' des informations présentées est focalisé >			
Exemples	<i>L'enjeu principal est...</i> <i>Le problème essentiel est...</i>			
Type de phrase	Phrase déclarative			
Structure syntaxique	[nom de THÈME] + [adjectif d'IMPORTANTANCE] + $\left\{ \begin{array}{l} \textit{est (de)} \\ \textit{réside dans} \\ \textit{reste} \end{array} \right\} + \left\{ \begin{array}{l} \text{SN} \\ \text{SV} \\ \text{Complétive que} \end{array} \right\}$			
Réalisations lexicales	<i>L'enjeu/problème/question/point</i>	<i>principal/fondamental/essentiel/majeur/central</i>	<i>est (de)/reste</i> <i>réside dans</i>	SN/SV(syntaxme verbal)/complétive <i>que</i>
Éléments périphériques	Marqueurs discursifs : les adverbes de réexamen : <i>en effet, en fait</i> ; de topicalisation : <i>dans cette optique</i> , de déduction : <i>ainsi</i> ; de concession : <i>mais</i> .			
Paramètres énonciatifs	Dialogisme interlocutif. L'auteur sélectionne le thème essentiel pour le lecteur.			
Position textuelle	Dans le corps principal du discours (pas de position spécifique), parfois dans une note.			
Fonctions	Souligner le thème en question pour le lecteur.			

Tableau 7-5 : Routine N° 4 de mise en évidence

Routine N° 2 de focus.

Description sémantique	< Un élément constitue le ' thème ' en question >		
Exemples	<i>... dont il est question...</i> <i>Il n'est pas question de...</i>		
Type de phrase	Phrase déclarative, impersonnelle		
Structure syntaxique	(dont) + il est + question (de) + (adverbe discursif)		
Réalisations lexicales	<i>(dont) il</i>	<i>est/(n'est pas) question de</i>	<i>ici, donc, ainsi, bien...</i>
Éléments périphériques	Marqueurs discursifs : les adverbes de déduction : <i>donc, alors</i> ; de réexamen : <i> finalement</i> ; de localisation : <i>ici</i> . Marqueurs de modalité : <i>sans doute, plutôt</i> .		
Paramètres énonciatifs	Dialogisme interlocutif. L'auteur sélectionne le thème essentiel pour le lecteur.		
Position textuelle	Dans le corps principal du discours (pas de position spécifique), parfois dans une note.		
Fonctions	Souligner le thème en question pour le lecteur.		

Tableau 7-6 : Routine N° 2 de focus

7.2.3.1 *Caractéristiques syntaxiques, lexicales et sémantiques de ces routines*

Les descriptions sémantiques des routines de focus recourent à deux nouvelles étiquettes conceptuelles : **‘point’** et **‘thème’**. La première permet d’introduire le concept du thème focalisé et s’opère dans notre corpus par les noms de THÈME qui est une sous-classe de la catégorie des noms d’OBJET_SCIENTIFIQUE, définie comme « observables et objets construits de l’activité scientifique » (cité du site LST¹⁰³). Les noms scientifiques de THÈME comme *enjeu*, *problème*, *question*, etc. Les noms utilisés dans la routine N° 1 sont des *shell nouns* (Schmid, 2000) ou des *signalling nouns* (Flowerdew, 2003). Ces noms, largement utilisés dans les écrits scientifiques, regroupent des noms abstraits qui servent à synthétiser ou présenter les informations ou des concepts (Schmid, 2000). Ils ne possèdent pas de signification spécifique qu’en s’inscrivant dans un contexte précis (Flowerdew, 2003, p. 229). Par exemple, dans la routine N° 1, le lecteur ne peut savoir ce que signifient exactement les noms *question*, *point* ou *problème*. Leur sens est spécifié par l’information donnée dans la proposition subordonnée suivant la routine. Les *shell nouns* sont multifonctionnels dans les écrits scientifiques : ils peuvent caractériser et mettre en perspective des blocs d’information et exercer une fonction textuelle reliant le nom à des phrases ou à d’autres parties textuelles. (Schmid, 2000, p. 14). La proposition [shell noun] + *est* + [*de* + infinitif/Syntagme nominal/*que* + proposition subordonnée] est très fréquente tant en anglais et qu’en français. Une telle construction a l’avantage communicationnel de fournir « dans une seule période, un dispositif complexe de mise en saillance, de connexité, de catégorisation d’un référent discursif, en un « concept » *ad hoc* » (Legallois & Grea, 2006, p. 166). Schmid (2000, p. 337) souligne aussi la fonction de focalisation et de topicalisation de cette construction avec des noms *chose*, *point*, *question*, *problème* (*thing*, *point*, *question*, *problem* en anglais), en précisant qu’elle a un effet emphatique.

La deuxième étiquette **‘thème’** désigne le point essentiel à traiter *hic et nunc* dans le discours en cours. Nous employons cette nouvelle étiquette au lieu de celle de **‘point’**, car la réalisation lexicale n’est pas identique à cette dernière. Le **‘thème’** se réfère dans notre corpus au nom *question* en formant une construction impersonnelle très figée : *il est question de*, (les exemples 23-24).

[23] Peut-être répondra-t-on que c’est évident : dans la distinction entre cousins parallèles et cousins croisés, **il est question de différence** entre les deux sexes,

¹⁰³ Voir le site LST : <http://lst.demarre-shs.fr/classification/cat/3>, consulté le 27 avril, 2022.

et c'est cette différence que l'on étend (en posant FZ F ou MB M).
(Anthropologie)

[24] Toutes les portes..., toutes les énergies..., toutes les facultés : on aura relevé sans peine que *l'éducation rationnelle dont il est question ici* n'est pas sans rapport avec une récusation du parcellaire. (Science de l'éducation)

[25] **Ce dont il est question, clairement, c'est d'en terminer** avec la « traditionnelle invisibilité ethnique dans les statistiques » pour « prendre vraiment à bras-le-corps les défis du développement national » [50]. (Sociologie)

[26] **Il n'est pas question ici de statuer sur sa sincérité, mais de souligner** que choisir de défendre une position en surplomb, non située en termes d'intérêts locaux, n'était pas la seule solution possible et que ce discours a ensuite été abandonné. (Histoire)

Cette routine manifeste deux structures syntaxiques :

- a) En tant que construction propositionnelle principale qui peut être suivie par soit un syntagme verbal (26) soit par un syntagme nominal (23, 24) ;
- b) En tant que proposition subordonnée introduite par le pronom relatif *dont* (24). Dans cette construction, le thème en question se place devant *dont*. Cette construction est très courante dans l'écrit pour alléger la tournure. Elle peut aussi se concrétiser en structure pseudoclivée (25).

La troisième étiquette '**essentiel**' est spécifique à la routine N° 1. Elle accorde une valeur fondamentale à '**point**' afin de capter l'attention du lecteur. Déjà discutée dans la routine N° 2 de mise en évidence, elle est représentée généralement par les adjectifs d'importance comme *fondamental, essentiel, important, intéressant*¹⁰⁴.

[27] **L'enjeu fondamental est donc de savoir si la question des identités collectives**, en particulier culturelles, doit être renvoyée à la seule sphère privée, comme ce fut le cas pour la question religieuse au XVIIe siècle, ou si au contraire la notion de communauté politique doit être repensée de manière à relativiser la valeur des droits universels en fonction du pluralisme culturel. (Sociologie)

[28] **Le problème essentiel est sans doute celui des prépositions souvent dites « vides »** ou, du moins, dont certains emplois peuvent être dits « vides ». (Linguistique)

[29] **Le point crucial est que les agents anticipent correctement la distribution de loyers : les deux loyers R_i et R_i sont égaux à l'état stationnaire.** (Économie)

[30] **La question essentielle qui se pose**, dès lors qu'est employé le terme de compétence, **touche** au fait que cette aptitude discriminante se développe

¹⁰⁴ L'adjectif *intéressant* est classé dans la sous-classe des adjectifs AXIOLOGIQUE_POSITIF. Nous l'avons classé dans la classe d'IMPORTANCE car dans les routines de topicalisation, *intéressant* remplit aussi une fonction de capter l'attention du lecteur sur un constat, une information, un fait ou un thème.

principalement au domicile de ces jeunes, à l'occasion des initiatives de recherche qu'ils prennent, leur permettant d'explorer des thèmes personnels ou d'extrapoler les recherches qui leur sont prescrites. (Science de l'information)

[31] **Le principal problème réside dans** le choix d'une conception très extensive du concept de médiation. (Politique)

À l'instar de la première construction syntaxique de la routine N° 2, les compléments d'objet de la routine N° 1 < Le '**point**' '**essentiel**' en quoi consistent des informations > s'accomplissent par des syntagmes nominaux (28) et verbaux (27) et par une proposition subordonnée (29). Elle peut avoir quelques variations comme *la question essentielle se pose* (30), *la question générale qui est posée...*¹⁰⁵, *le principal problème réside dans...* (31), *la question centrale reste pour nous...* Avec cette routine, l'auteur réalise aussi un constat pour indiquer le problème central au lecteur, qui pourra ainsi comprendre la complexité et la difficulté de la recherche.

7.2.3.2 *Éléments périphériques et paramètre énonciatif*

Il existe moins de marqueurs discursifs dans les routines de focus par rapport aux routines de mise en évidence et de (co)constat. Les marqueurs identifiés (adverbes de DÉDUCTION : *alors, donc* ; de RÉEXAMEN : *finalement* ; de LOCALISATION : *ici*) présentent aussi dans les routines de mise en évidence et de (co)constat. Le sous-emploi des marqueurs discursifs peut être attribué à la fonction de focus qui oriente l'attention du lecteur sur un point précis : il s'agit plus ici d'un acte communicatif que d'un élément argumentatif. Par ailleurs, les occurrences des routines de focus sont moins nombreuses que les routines de mise en évidence et de (co)constat (*cf.* section 8.3.1.3.)

Nous remarquons toutefois l'emploi abondant du déictique *ici* dans la routine N° 2 (*il est question ici de...*) : 23 occurrences parmi 37 identifiés voisinent avec *ici*. *Ici* est un adverbe classé dans la catégorie d'ESPACE_LOCALISATION. Il renvoie dans l'écrit soit à un segment textuel, soit au contenu de ce segment (Vuillaume, 2018, p. 25). Dans le premier cas, *ici* peut être interprété comme « *dans ce que je viens de dire* », et dans le deuxième cas, il est synonyme des locutions « *dans ce sens, en l'occurrence* » (Vuillaume, 2018, p. 26). Ainsi, *ici* remplit une fonction textuelle dans l'écrit qui indique au lecteur le thème en cours d'énonciation. L'emploi abondant de *ici* permet de dire que la localisation est souvent importante pour les routines de topicalisation.

¹⁰⁵ *Poser* une question est une collocation. Donc, le participe passé *posée* voisine souvent avec *question*, ne coexiste pas avec d'autres noms de THÈME comme *problème, point, enjeu*. Note : *poser problème* est différente que *poser une question*.

7.2.3.3 *Fonctionnement rhétorique*

Les routines de focus dans notre corpus ont pour objectif de signaler au lecteur que le thème en question mérite une focalisation. Dans les écrits scientifiques, les informations sont souvent très abondantes avec le changement de thèmes, ce qui peut troubler l'attention du lecteur. Pour remédier à ce problème, l'auteur peut employer ces routines de focus pour indiquer de façon explicite le thème essentiel traité dans un lieu textuel précis. Si, les routines s'employant à la forme affirmative peuvent bien jouer le rôle de focalisation, elles sont capables d'introduire une deuxième fonction à la forme négative, éliminant ainsi le thème non essentiel qui ne constitue pas l'objet d'observation dans les écrits scientifiques. Dans ce cas, il s'agit de dégager le thème que l'auteur ne traitera pas ici ou de l'anticipation d'objections du lecteur sur des thèmes non nécessaires ou non pertinents (32), en précisant ensuite le thème à prendre en compte (26, 33).

[32]C'est tout le propos des lignes qui suivent, dont on notera d'une part qu'elles n'engagent que leur auteur, d'autre part qu'elles reposent sur un ensemble impressionniste : **il n'est pas question ici** d'enquêtes statistiques, mais de ressenti, de témoignage subjectif - un peu de subjectivité ne fait jamais de mal. (Science de l'information)

[33]**La question essentielle n'est plus** la taille de l'État par rapport au marché, **mais plutôt** sa capacité de gestion et sa capacité de concertation avec le secteur privé. (Économie)

En résumé, la forme négative pour dégager le thème non pertinent n'est pas très présente, et peut être accompagnée de tournure de précision du thème et le marqueur d'opposition, *mais* comme les exemples 26 et 33.

Il est à noter que l'emploi des adjectifs d'IMPORTANTANCE dans la routine N° 1 permet de montrer que l'auteur parvient à discriminer les informations importantes pour le lecteur et donc maîtrise le sujet. Ainsi, l'auteur a un point de vue surplombant ici.

7.3. Synthèse

Pour conclure, l'écrit scientifique en français présente une abondance de routines productives qui permettent la mise en relief des faits, des informations, ou des entités scientifiques. Dans cette catégorie, nous avons retenu 3 sous-fonctions : la mise en évidence, le (co)constat et le focus. Ces trois fonctions concernent la hiérarchisation des informations pour le lecteur, et avec chacune une fonction spécifique. La mise en évidence comporte plus de routines et vise à montrer l'importance d'une information dans le texte. Cette information peut être un fait scientifique, un constat, une déduction, une information. Le (co)constat vise à mettre en avant un constat de l'auteur, en enrôlant le lecteur comme témoin de l'argumentation et

l'invitant à partager le constat. Le focus, un mécanisme plus simple, regroupe des routines des constructions très spécifiques qui mobilisent des noms abstraits capables d'envelopper un empan d'information comme *question, enjeu, problème*. Ces routines permettent de se focaliser sur un thème spécifique qui mérite d'être signalé à un moment précis de l'énonciation.

Les routines de topicalisation s'actualisent sous la forme de plusieurs structures syntaxiques. Elles apparaissent souvent en cooccurrence avec des marqueurs discursifs, en particulier avec des marqueurs de topicalisation (*quant à, concernant, sur ce point*) pour certaines routines. Ces marqueurs sont très récurrents et permettent aux routines de topicalisation de s'inscrire plus solidement au plan du contenu et argumentatif.

Bien sûr, les routines de topicalisation reflètent parfaitement la propriété interlocutive des écrits scientifiques : la hiérarchisation et la sélection des informations ou des thèmes spécifiquement importants pour le lecteur demandent à l'auteur de prendre en compte son destinataire. Un article de recherche comporte des informations abondantes, et les éléments essentiels doivent être mis en évidence, afin que le lecteur puisse suivre le fil de conducteur de la recherche.

Chapitre 8. Routines de guidage à travers les parties textuelles, l'actualisation énonciative et les disciplines

Dans ce chapitre, nous nous intéressons aux régularités des routines de guidage portant sur les positions textuelles, l'actualisation énonciative et les disciplines. Dans les tableaux de modélisations des routines dans les chapitres 5, 6 et 7, nous avons indiqué d'une façon succincte où se trouvent les routines dans le texte sans entrer dans les détails, car les routines ne se trouvent pas toujours aux mêmes endroits dans les articles de recherches. L'observation de leurs positions textuelles permet de mieux comprendre la pratique de l'écriture scientifique, ce qui peut être appliqué ultérieurement dans la didactique des routines. Par exemple, on pourra préciser que les routines d'annonce de l'objectif d'étude seront préférentiellement utilisées dans la partie introductive.

Au-delà des emplacements des routines, nous présentons également les formes d'actualisation dans la situation d'énonciation. D'abord, la capacité d'actualisation dans le processus énonciatif est l'un des critères définitoires fondamentaux des routines (*cf.* section 1.2.5), par rapport aux autres unités phraséologiques telles que les *lexical bundles* ou les *motifs* que nous avons observés. Il convient donc de mener une observation sur cet aspect, afin de découvrir les modalités d'inscription des routines dans l'énonciation des écrits scientifiques. Cet aspect a été évoqué rapidement dans les sous-sections des éléments périphériques et des paramètres énonciatifs des trois chapitres précédents, mais dans ce chapitre, nous voulons proposer une vision globale sur ces paramètres.

Enfin, nous procéderons à une analyse quantitative sur les routines de guidage en suivant deux objectifs : 1) la répartition des routines de guidage dans le corpus global ; 2) les particularités disciplinaires, à l'aide, en particulier, des analyses factorielles des correspondances, afin de tester quelques hypothèses sur les modes rhétoriques des disciplines.

8.1. Positions textuelles des routines de guidage

La position textuelle a été évoquée succinctement pour certaines routines dans les chapitres 5, 6 et 7, notamment pour les routines de structuration, sans proposer une étude systématique sur ce point. Les routines sont-elles spécifiquement plus fréquentes dans une partie textuelle spécifique que les autres ? Nous pouvons formuler l'hypothèse que l'emplacement des routines est étroitement lié à leurs fonctions. Par ailleurs, nous discuterons quelques propriétés syntaxiques et typographiques pour certaines routines.

Nous avons expliqué dans la partie méthodologie (*cf.* section 4.6.2.2) que l'extraction ne permet pas de fournir de façon systématique les répartitions des positions textuelles des routines¹⁰⁶, et nous avons observé ces données manuellement. Ainsi, une telle procédure ne permet pas une observation systématique des routines en fonction des parties textuelles. Néanmoins, nous pouvons proposer des observations en fonction des grandes catégories de routines. Nous verrons que les routines relevant des mêmes fonctions manifestent des propriétés partagées en ce qui concerne les lieux textuels. Les parties textuelles étudiées comportent principalement : le résumé, l'introduction, le corps principal, la conclusion, les notes. Dans le cas où le corps principal d'un article se divise en sous-sections avec des thématiques précises, nous ne faisons plus la différence entre ces sections, et ne l'indiquons que par le corps principal.

8.1.1. Emplacement des routines de structuration dans le texte

Les routines de structuration, qui concernent l'organisation discursive du texte, comportent quatre sous-fonctions : la planification de la structure, l'annonce de l'objectif d'étude global ou local, la navigation textuelle et la conclusion/déduction.

En fonction de leur fonctionnement rhétorique (*cf.* Chapitre 5), nous supposons que :

- les routines de planification et d'annonce de l'objectif se trouvent davantage dans les parties introductives comme le résumé et l'introduction ou au début des sous-sections pour annoncer le plan ou l'objectif partiel ;

- les routines de navigation apparaissent à travers tout le texte et sont plus présentes dans le corps principal des articles ;

- les routines de conclusion/déduction ont tendance à se situer dans la conclusion, ou dans la discussion si l'article possède une sous-section ainsi nommée.

Pour vérifier ces hypothèses, nous avons observé les routines de structuration en examinant leurs contextes gauches et droits jusqu'à ce qu'un titre de section apparaisse. De manière générale, les routines de structuration présentent certaines régularités qui sont très étroitement liées à leurs fonctions. En effet, l'emplacement peut parfois modifier leurs rôles fonctionnels.

D'abord, sans surprise, les routines de planification et d'annonce de l'objectif/l'objet d'étude global ou local se trouvent principalement dans l'introduction (exemple 1) ou le résumé

¹⁰⁶ Les informations sur les parties textuelles ne sont pas exploitables dans le Lexicoscope, contrairement à Scienquest.

(exemple 2) des articles de recherche. Quelques occurrences peuvent être identifiées au début d'une sous-section du corps principal du texte (exemple 3).

- [1] La structure de cet article est la suivante : **la section 2 est consacrée à la présentation de notre modèle théorique. Les données utilisées pour son estimation sont décrites dans la section 3. La section 4 est consacrée à la présentation et à la discussion des estimations économétriques. La section 5 conclut.** (Économie, routines de planification dans l'introduction)
- [2] **Cet article propose une analyse** géographique de l'événement territorial que constitue le retour du siège du gouvernement et du Parlement allemands à Berlin à partir d'une relecture des débats parlementaires du 20 juin 1991. (Anthropologie, routine d'annonce de l'objectif dans le résumé)
- [3] **Dans la section suivante, nous étudierons** le modèle avec des spécifications linéaires particulières des fonctions α . (Économie, routine de planification dans le début d'une sous-section du corps principal de l'article)

Le fait que les routines de planification et d'annonce de l'objectif soient très fréquentes dans l'introduction est lié aux propriétés rhétoriques de cette partie textuelle cruciale pour un article de recherche. Rappelons que Swales (1990, p. 141) souligne l'importance de l'annonce de l'objectif d'étude et de la structure d'organisation de l'article dans l'introduction, qui consiste en deux étapes inscrites dans le mouvement rhétorique « occuper la niche » de son modèle « créer un espace de recherche » (pour plus de détails, voir sections 5.3.1.3 et 5.3.2.3). Selon cet auteur, l'annonce de l'objectif d'étude occupe une place plus importante que la planification du discours. La première est obligatoire, tandis que la deuxième n'est pas indispensable. Certaines disciplines privilégient plus l'annonce du plan que d'autres, comme l'informatique (Swales, 1990, p. 161) ou l'économie dans notre corpus (on reviendra sur ce point dans la section dédiée aux particularités disciplinaires).

En parallèle, nous identifions quelques routines d'annonce de l'objectif/l'objet qui se positionnent dans la partie conclusive ou dans le corps principal avec un temps verbal au passé. Dans ce cas précis, la routine remplit une fonction secondaire de rappel de l'objectif du discours :

- [4] **Notre objectif était de tester** la validité de ce modèle en fonction d'un autre contenu ; nous pensons l'avoir atteint. (Psychologie, dans la conclusion)

En ce qui concerne les routines de navigation textuelle, nous avons remarqué qu'elles sont principalement réparties dans le corps principal du discours. Mais les routines N° 2 (*comme nous l'avons dit, on va le voir*), N° 4 (*je passe maintenant à, on y revient, nous revenons sur ce point plus tard*), N° 6 (*voir annexe X, voir la note X*) et N° 7 (*voir l'article de quelqu'un, se référer à l'ouvrage de quelqu'un*) présentent des singularités textuelles remarquables. Ces

routines apparaissent souvent dans deux emplacements spécifiques : en incise ou dans une note de bas de page ou de fin d'article. L'incise peut s'intégrer directement dans une proposition ou apparaître entre parenthèses. En fonction de ces modalités, nous résumons ces éléments dans le tableau *infra* (+ signifie oui, - signifie non, ++signifie un degré élevé) :

Routines	Incise	Entre parenthèses	Dans les notes
N° 2 (<i>comme nous l'avons dit, on va le voir</i>)	++	+	-
N° 4 (<i>on y revient, nous revenons sur ce point plus tard</i>)	+	+	-
N° 6 (<i>voir annexe X, voir la note X</i>)	+	++	++
N° 7 (<i>voir l'article de quelqu'un, se référer à l'ouvrage de quelqu'un</i>)	+	++	++

Tableau 8-1 : Positions textuelles particulières de certaines routines de navigation textuelle

La routine N° 2 apparaît majoritairement en incise dans le texte. Elle peut se placer au début, au milieu d'une proposition ou entre parenthèses, mais pas dans les notes. La routine N° 4 se trouve dans le corps du texte, parfois entre parenthèses, notamment pour l'actualisation lexicale *on y revient, on revient sur ce point*. La routine N° 6 peut se trouver dans le corps du texte (souvent entre parenthèses) ou dans les notes. La routine N° 7 peut apparaître dans tous les endroits indiqués, mais est beaucoup plus présente entre parenthèse et dans les notes.

Les notes occupent une place fondamentale dans les sciences humaines et sociales pour présenter des arguments et des informations supplémentaires qui soutiennent une thèse dans le texte principal des articles de recherche. Elles reflètent « le profil intellectuel propre des différentes communautés scientifiques, [...renvoyant], au-delà des sources précises de l'information scientifique, aux théories et aux écoles théoriques auxquelles elles sont ou espèrent être liées » (Grafton, 1998, p. 21). Dans la pratique des articles de recherches, les notes sont des commentaires hors du texte principal, servant à ajouter une référence, un commentaire sur un élément précis du texte principal. Elles ont souvent une fonction de preuve. Facultatifs, elles ne touchent pas le fond du discours, mais fournissent au lecteur des informations complémentaires sans alourdir le texte principal. Les notes peuvent introduire une citation, expliquer une source, ce qui est le cas des routines N° 6 et N° 7. Mettre les éléments de références dans les notes au lieu de les insérer dans le discours principal est une pratique très courante dans la communauté scientifique. Cela permet de hiérarchiser l'information. En parallèle, les références assemblées dans les notes permettent au lecteur de se reporter vers d'autres discours, s'il veut obtenir plus d'informations.

Contrairement aux notes qui ne sont pas dans le discours principal, l'incise s'intègre au processus d'énonciation. Ce phénomène syntaxique a été beaucoup étudié dans le discours

rapporté, notamment pour les citations comme *Mon frère, Marie l'a dit, est arrivé* (voir Bonami & Godard, 2008 ; Wilmet, 2010, p. 549–551). Dans nos routines, la routine N° 2 de navigation textuelle fonctionne généralement comme incise et possède deux constructions : l'une avec la conjonction *comme* (*comme nous l'avons vu/constaté... comme nous le verrons...*), l'autre sans *comme* (*on l'a vu, on va voir, que nous avons exposé plus haut*). Elles n'ont pas les mêmes statuts syntaxiques. La première n'est pas une incise à proprement dit, car elle correspond à une phrase subordonnée qui a un lien syntaxique avec la proposition principale et un lien logique avec le contexte textuel. Il s'agit d'une preuve déjà communiquée qui peut soutenir le raisonnement de la phrase suivante. Dans l'exemple 5, l'information *ceux-ci ne peuvent plus être purement et simplement assimilés à des formes ancestrales d'assistance ou de charité* a déjà partagé dans la première partie du texte.

[5] **Comme nous l'avons noté en première partie**, ceux-ci ne peuvent plus être purement et simplement assimilés à des formes ancestrales d'assistance ou de charité. (Sciences politiques)

La construction sans *comme* peut être considérée comme un type d'«incise progressive»¹⁰⁷ qui « ne dispose pas d'autonomie sémantique ou syntaxique » (Grossmann, 2018, p. 107). Les incises de ce type montrent « l'enrôlement du lecteur dans l'argumentation ou le rappel de points déjà traités » (*ibid.*) (6) ou le point à traiter dans le fils du discours (7).

[6] L'ambiguïté du statut de ces « fils illégitimes » résulte des deux logiques contradictoires **que nous avons exposées plus haut**. (Anthropologie)

[7] À cet égard, l'institutionnalisation de l'ethnologie a, **on va le voir**, deux conséquences très différentes : la professionnalisation des enquêteurs entraîne une modification des rapports entre théorie et pratique ; l'intégration de la discipline à l'université sous la coupe de la sociologie durkheimienne ou postdurkheimienne conduit à concevoir le fait ethnographique comme un fait d'ordre moral, non réductible au document ou à la pièce muséographique. (Anthropologie)

Contrairement aux incises non autonomes, les routines N° 6 (8) et N° 7 (9) s'inscrivent dans le discours en fonctionnant comme une phrase indépendante de type injonctif.

[8] **Pour une présentation détaillée de ce concept, voir la note de synthèse** parue dans la Revue française de pédagogie (Sarrazy, 1995). (Science de l'éducation)

[9] **Voir en annexe A**. (Économie)

¹⁰⁷ Ce terme vient de Cornulier (1978), « L'incise, la classe des verbes parenthétiques et le signe mimique ». Pour les syntacticiens, ce sont les vraies incises.

On peut considérer que les routines fermées par des parenthèses ont un statut spécifique comme des incises autonomes. La position entre parenthèses signifie que le contenu ne touche pas le contenu principal du discours, mais fournit une information complémentaire.

Les routines N° 6 (10) et N° 7 (11) ont tendance à apparaître dans les parenthèses :

[10] Le corpus actuel de la recherche comprend au final 180 articles tous antérieurs à janvier 2004 (**voir Annexes, graphique 1**). (Science de l'informatique)

[11] Ce soi qui ne va pas de soi, Paris, Descartes& Cie, 2002 (**voir notamment le chapitre** « Entre philosophie et sciences humaines » : 315-332) (Anthropologie).

Dans notre étude, nous regroupons tous les types d'incises abordés plus haut par le terme de « subordonnées incises » pour qualifier les routines qui se situent entre parenthèses (les exemples 10 et 11), ou entre deux virgules (7), ou au début (5) ou à la fin d'une phrase (6), ayant pour fonction d'introduire une information supplémentaire qui n'a pas de lien direct avec le contenu principal, et intégrant un mot subordonnant.

En résumé, les notes, les incises et les passages entre parenthèses sont des pratiques très courantes dans les écrits scientifiques. Le fait de placer les routines dans ces emplacements permet au scripteur de se concentrer sur le contenu du premier plan du texte pour ne pas dévier l'attention du lecteur, mais de veiller simultanément au besoin du lecteur d'avoir des explications supplémentaires, sur la navigation textuelle, sur les sources bibliographiques.

Enfin, contrairement à notre hypothèse, les routines de conclusion/déduction ne sont pas identifiées majoritairement dans la conclusion. Elles se trouvent souvent dans le corps principal ou dans la discussion où une conclusion discursive sur un paragraphe ou une déduction d'une chaîne argumentative est essentielle. Pour expliquer pourquoi les routines de conclusion/déduction sont présentes dans la discussion, nous avons fait une requête rapide pour la routine [*on/nous*] (*en*) [*conclure/déduire*] et deux marqueurs *pour conclure, en conclure* dans la plateforme Scienquest¹⁰⁸ qui propose des résultats répartis dans les parties textuelles des articles. Nous avons identifié 42 occurrences dans la partie du développement (le corps principal), mais aucune occurrence n'est présente dans la conclusion. En revanche, les marqueurs conclusifs *pour conclure, en conclusion*, sont bien identifiés à la fois dans le corps principal et dans la conclusion. Il semble que les auteurs préfèrent utiliser les routines de

¹⁰⁸ Nous avons présenté très rapidement cette plateforme dans la note de bas de page 60, en fournissant le site. Elle inclut 300 articles de notre corpus TermITH. Elle propose neuf parties textuelles : introduction, développement-le corps principal dans notre étude, conclusion, résumé, note de bas de page, remerciements, annexes, avant-propos et mots-clés)

conclusion/déduction dans le corps principal pour formuler des conclusions locales d'une partie textuelle ou d'une chaîne d'arguments, mais emploient en revanche sans préférence les marqueurs de conclusion dans le corps principal ou dans la conclusion.

Pour conclure, l'emplacement des routines de structuration manifeste des tendances remarquables selon leurs fonctions rhétoriques qui sont en rapport avec les propriétés rhétoriques des parties textuelles des articles de recherche. Les routines de planification et d'annonce de l'objectif/l'objet d'étude dans le texte sont plus fréquentes dans la partie introductive, alors que les routines de conclusion/déduction ont tendance à se placer dans le corps principal, en particulier, dans la discussion. Les routines de navigation textuelle manifestent une diversité de placement dans le texte. Elles sont présentes dans le corps du texte, ou dans les notes, souvent comme incises.

8.1.2. Positions textuelles des routines métalinguistiques

Les routines métalinguistiques visent principalement au décodage du code et du contenu scientifiques. Les fonctions retenues pour ces éléments sont la reformulation, la réparation, la précision terminologique et l'emprunt terminologique. Ces routines sont-elles étroitement liées à certaines parties textuelles spécifiques ?

Nous pouvons ainsi imaginer que les routines à fonction métalinguistique apparaissent plutôt dans le début de l'article et non en conclusion. Par ailleurs, on peut supposer que les routines de précision et d'emprunt terminologique font appel à des éléments typographiques spécifiques comme les guillemets ou l'italique.

Contrairement aux routines de structuration discursive qui se trouvent parfois spécifiquement dans l'introduction, le début d'une section textuelle ou dans les notes, les routines métalinguistiques dans notre corpus semblent apparaître sans préférence de positions textuelles, en particulier les routines de reformulation et de réparation.

Mais les routines de précision et d'emprunt terminologique apparaissent plutôt dans le début et le corps principal de l'article et non en conclusion. Un terme scientifique (nouveau ou déjà existant) renvoie à une notion, un concept, un fait ou un phénomène scientifique qui constituent le noyau du développement argumentatif du texte scientifique. Son introduction a ainsi tendance à être mise en place soit au début soit au milieu de l'article, afin d'inscrire le texte dans un cadre disciplinaire, comme la définition et l'emprunt terminologique. Il est plus rare que l'on introduise un terme scientifique dans la partie conclusion.

Nous relevons aussi des occurrences de routines métalinguistiques dans les notes, comme l'exemple 12 :

[12] [3] **Nous reprenons le terme** à Philippe Le Guern (2010), qui évoque le fait d'être fan. (Science de l'information, dans la note)

[13][109] Initialement, **le terme** de pèlerin **désigne** l'étranger, le voyageur, avant de s'appliquer à celui qui fréquente un sanctuaire, proche ou lointain. (Histoire, dans la note)

Ces informations ne sont peut-être pas essentielles, mais méritent tout de même une explication. Cela montre que l'auteur est capable de hiérarchiser les données pour le lecteur, comme nous l'avons observé plus haut.

En résumé, on peut conclure que les routines métalinguistiques apparaissent dans la plupart des parties textuelles, à condition qu'elles manifestent un besoin d'éclaircissement sur l'usage de la langue et des termes disciplinaires, sauf dans la conclusion.

Si les routines métalinguistiques ne présentent pas de régularité aussi remarquable que les routines de structuration, les routines de précision et d'emprunt terminologiques, présentent quelques dispositifs typographiques spécifiques dans le corpus. Ces routines font appel aux guillemets (« » ou "") pour encadrer les énoncés définitoires dans les routines de définitions (14), l'appellation attribuée dans les routines de dénomination (15), ou le terme utilisé dans les routines de désignation (16), d'emprunt terminologique (17).

[14] Or le territoire, forestier ou non, **peut être défini comme** un « espace à métrique topographique », par opposition aux réseaux qui relèveraient d'une « métrique topologique [16] ». (Géologie)

[15]**On appellera** ici « formalistes » les premières conceptions, et « antiformalistes » les deuxièmes. (Sociologie)

[16] Ainsi, **le terme** "état émotionnel négatif" **désigne** ici l'expérience émotionnelle caractérisée par une valence négative et un éveil plutôt élevé (anxieux, inquiet, stressé, troublé) en opposition avec un état émotionnel caractérisé par une valence positive et un éveil plutôt faible (calme, détendu, tranquille, confiant). (Psychologie)

[17] Ces données permettent en effet de distinguer les programmes selon leur degré de « mise en danger » des invités politiques, **pour reprendre les termes d'Erik Neveu** (Neveu, 2003). (Science de l'information)

8.1.3. Positions textuelles des routines de topicalisation

Rappelons que les routines de topicalisation consistent à mettre en évidence une information importante, à fournir des constats, et à focaliser sur un thème important. Selon leurs fonctions, il y a plus de chance que les routines de topicalisation se trouvent dans le corps

principal ou la conclusion, et qu'elles soient beaucoup moins présentes dans le résumé, l'introduction ou les notes.

En fait, nous n'avons relevé qu'une occurrence de routines de focus [*le(la, les, l')*] [*enjeu/question/problème/point*] [*principal/essentiel/important*] [*est de/repose dans*] et *il est question de* dans le résumé (18), et quelques occurrences de routines de mise en évidence dans les notes. Les autres se trouvent toutes dans le corps principal.

[18] **Le problème essentiel est** sans doute celui des prépositions souvent dites "vides" ou, du moins, dont certains emplois peuvent être dits "vides". (Linguistique, dans le résumé)

[19] [4] **Il faut** tout de même **signaler** les témoignages laconiques sur quelques combats entrevus par des explorateurs occidentaux, tel le Français Simon de La Loubère au xvii^e siècle ou le Britannique James Low au xviii^e, sans oublier le voyageur Henry G. Quaritch - Wales au xix^e siècle. (Anthropologie, dans la note)

8.1.4. Synthèse

En résumé, on a pu constater que les routines de structuration présentent plus de préférences au plan textuel que les autres fonctions de routines. D'après les observations, nous pouvons récapituler les positions des routines dans le tableau ci-dessous (+ signifie oui, - signifie non, ++ signifie un degré élevé) :

Catégories fonctionnelles	Sous-fonctions	Résumé et Introduction	Corps principal	Conclusion	Entre parenthèses	Dans les notes
Structuration	Planification	++	+(début d'une section)	-	-	-
	Annonce de l'objectif/l'objet d'étude	++	+	+	-	-
	Navigation textuelle	-	++	-	++	++
	Conclusion/déduction	-	++	+	-	-
Métalinguistique		+	++	-	-	+
Topicalisation		+	++	+	-	+

Tableau 8-2 : Récapitulation des positions textuelles des routines de guidage

Ce tableau présente les tendances principales du positionnement textuel des routines qui sont liées aux besoins rhétoriques spécifiques d'une section textuelle, comme la planification qui se situe toujours dans l'introduction.

Les routines ne présentent pas seulement des régularités sur les leurs emplacements dans le texte, mais aussi sur les marques d'actualisation énonciative.

8.2. L'actualisation énonciative des routines de guidage

Nous arrivons maintenant à l'observation sur les marques d'actualisation énonciative des routines de guidage. Rappelons que l'un des critères principaux dans la définition des routines adoptée dans la présente étude porte sur le plan énonciatif : les routines doivent être considérées fondamentalement dans le cadre énonciatif. D'après ce que nous avons observé dans les Chapitres 5, 6 et 7, l'inclusion des routines dans la situation d'énonciation se réalise de deux façons : de façon explicite avec la présence des déictiques, des expressions anaphoriques dans la routine, ou de façon implicite avec la réalisation des éléments périphériques comme les marqueurs ou les éléments anaphoriques ou cataphoriques dans la proposition subordonnée qui suit la routine. Les routines se distinguent l'une de l'autre en raison des fonctions différentes.

8.2.1. Les marques explicites d'actualisation énonciative des routines

Nous remarquons, en observant les routines de guidage d'une manière systématique, quelques manières explicites de l'actualisation sur le plan énonciatif qui correspondent à la présence des déictiques et du temps grammatical du verbe. Conçue comme un « passage de la langue au discours », l'actualisation se fait généralement par trois catégories typiques d'opérations dans le discours monologal (Bres, 2001, p. 253) : 1) l'actualisation temporelle, spatiale, personnelle via des anaphores ou des déictiques qui fonctionnent en tant que paramètres de référenciation ; 2) l'actualisation modale ; 3) l'actualisation sémiotique qui ne nous intéresse pas ici.

Dans nos données, nous relevons bien sûr les deux premières opérations à travers l'emploi :

1. Des pronoms personnels déictiques *je, nous, on* (20), des pronoms démonstratifs *ce, cet, ces* (21), des adjectifs possessifs *mon, notre* (22), et des anaphores *le* (23). Ces éléments sont très présents et se cumulent régulièrement dans les routines de structuration (*nous* et *dans cet article* dans l'exemple 20).

[20] **Nous proposons dans cet article d'analyser** les possibilités d'un développement durable à partir du comportement des consommateurs. (Économie)

[21] **Cet article propose une étude** de l'action patrimoniale développée actuellement à la Martinique (Antilles françaises). (Anthropologie)

[22] **Notre objectif est de montrer** que, dans un contexte de comparaison intergroupe menaçant pour l'identité sociale, les effets des comparaisons sociales

sur l'estime de soi sont médiatisés par l'identification à l'endogroupe.
(Psychologie)

[23]**Pour le dire encore autrement**, l'individu, porteur d'un sens subjectivement visé, est-il le niveau de réflexion adéquat de l'analyse sociologique ?
(Sociologie)

2. Des marqueurs temporels *auparavant, plus tard, maintenant*, ou spatiaux/locaux *en première partie, ci-dessous, plus haut, ici, dans l'introduction...*

[24]En effet, **comme nous l'avons rappelé ci-dessus**, il semble bien que l'influence de la variable immigration comme celle de la variable « langue parlée à la maison » soit « englobée » dans l'effet exercé par le statut socio-économique et culturel de la famille d'origine des élèves. (Sciences de l'éducation)

[25]C'est ce que **nous allons tenter de montrer dans les pages qui suivent**.
(Histoire)

3. Des verbes modaux *vouloir, pouvoir* (26, 27), des formes impératives (28, 29), des modalités déontiques (30, 31).

[26]**L'on peut** cependant **noter que** ces deux extraits trahissent une certaine ambiguïté quant à l'utilisation des termes « communication » et « information ».
(Sciences de l'information)

[27]En employant ce terme en référence à Max Weber [36], **nous voulons souligner que** son mode de légitimation consiste à activer un sentiment subjectif d'appartenance communautaire et à prétendre incarner le groupement en mettant en avant certains de ses attributs. (Sciences politiques)

[28][26] **Voir à ce sujet l'article de** Wajih Halwa intitulé « Capital punishment, Amman's messy urban sprawl ». (Histoire)

[29]**Notons également qu'**en conformité avec notre hypothèse une majorité (9 sur 11) de participants commençant la construction par une stratégie analytique n'effectuent pas ultérieurement de changement (Test binomial exact : $p = 0,065$; ns). (Psychologie)

[30]**Il faut noter que** le modèle peut aussi être enrichi de concepts et de relations non trouvés par Caméléon. (Linguistique)

[31]**Il est à noter que** les morts s'échelonnèrent entre le 5 mai et le 20 août, ce qui manifeste clairement que le coût lié à leur mort était assumé. (Sociologie)

4. Du temps grammatical des verbes (le présent, le futur, le passé, tous les exemples cités plus haut).

On peut résumer que ces voies d'actualisation des routines dans la situation d'énonciation scientifique sont plutôt manifestes. Les éléments actualisés sont inclus dans les routines, ce qui permet donc d'identifier plus directement dans le processus énonciatif. Ces éléments énonciatifs sont observés dans les routines de structuration, dans les routines

métalinguistiques (par exemple *cela*, le dans *cela veut dire que*, le terme « » désigne... ; *on* et *nous* dans les routines de précision terminologique *nous définit X comme Y*, *on appelle Y X*) et dans la plupart des routines de topicalisation (*nous notons que...*). Mais nous avons constaté encore certaines routines qui ne permettent pas d'observer directement la réalisation actualisée.

8.2.2. Le mode implicite d'actualisation énonciative des routines

Les modes implicites d'actualisation concernent davantage les routines de topicalisation, en particulier les deux configurations lexico-syntaxiques impersonnelles :

1. Routine N° 2 de mise en évidence <Une information/un constat s'avère 'essentiel(le)' méritant d'être 'mis(e) en évidence'> : *il est [important/fondamental/essentiel] de [noter/souligner/observer/constater...]* ;
2. Routines N° 2 de focus <Un élément constitue le 'thème' en question> : *il est question de.*

Pour ces routines à structure impersonnelle (sauf les routines à modalité déontique comme *il [faut/est à/importe] de [noter/constater/signaler/observer...]*, le pronom impersonnel, en jouant le rôle du sujet syntaxique, ne possède pas de sens référentiel et n'est donc pas capable d'indiquer l'inclusion des routines dans l'énonciation. Dans ce cas, la mission de la mise en énonciation est accomplie par le complément d'objet, c'est-à-dire, les syntagmes ou les phrases subordonnées, ou par les marqueurs discursifs, comme nous l'avons vu dans la section 7.2.1.2.

[32] À cet égard, **il est important de noter que** *les observations* concernant le vent ne répondaient pas à un désir de connaissance scientifique, mais à des considérations pratiques. (Histoire)

[33] Néanmoins, **il est important de souligner que** *cette validation* a été réalisée avec des lycéens âgés de 12 à 18 ans.

[34] *Le charaf* **dont il est question ici** est donc moins un statut social qu'une valeur symbolique ou sémiologique qui permet de légitimer une forme de pratique religieuse et rituelle. (Anthropologie)

[35] Enfin, **il est important de souligner** le poids que pèse, dans le processus adoptif, la germanité prise, ici, dans toutes ses expressions de sexe, d'âge, de biologique et de social. (Anthropologie)

Dans les extensions de ces routines, les moyens linguistiques énonciatifs sont les suivants : les indices spatio-temporels *ici* (34), les éléments anaphoriques (pronoms démonstratifs ou articles définitifs : *les observations*, *cette validation*, *Le charaf* dans les exemples 32, 33 et 34). Ces éléments linguistiques permettent de situer le discours scientifique dans le temps et l'espace par rapport au moment de l'énonciation.

En outre, les marqueurs discursifs peuvent également favoriser l'intégration des routines dans les lieux énonciatifs du discours scientifique. Prenons le marqueur *enfin* comme exemple d'illustration, qui peut réaliser 7 valeurs pragmatiques différentes (Bertrand & Chanet, 2005). Parmi celles-ci, nous distinguons les valeurs d'organisateur textuel, de synthèse, de correction et d'énumération comme les plus représentatives des écrits scientifiques. Dans notre exemple (35), *enfin* véhicule une valeur d'organisation et d'énumération, qui clôture la sous-partie du discours et indique le dernier point essentiel pour le lecteur. Avec l'emploi du marqueur discursif *enfin*, la proposition produite après *enfin* est reliée aux propositions antérieures.

Il en va de même pour le marqueur de topicalisation *à cet égard* (32) et le marqueur de concession *néanmoins* (33). Le premier précise le thème dont il est en question, et apparaît avec la routine *il est important de noter que*. Le deuxième marqueur *néanmoins* manifeste bien la caractéristique interlocutive des écrits scientifiques en anticipant l'interprétation du lecteur.

En résumé, l'opération d'actualisation énonciative des routines s'effectue soit directement par les pronoms personnels, les déictiques et les anaphores, les verbes modaux et les formes impératives déjà inclus et facilement identifiés dans les routines, soit de manière implicite à travers des extensions comme le complément ou les marqueurs discursifs. Parmi tous les éléments permettant de réaliser l'inscription dans la situation d'énonciation, les pronoms personnels dans les routines de guidage revêtent un intérêt particulier. Il est intéressant de procéder à une observation de cette catégorie en fonction des verbes associés comme Fløttum *et al.* (2006, p. 82–86). Les verbes du LST identifiés dans les routines sont en partie comparables à ceux qui sont étudiés dans le cadre des travaux du projet KIAP (*cf.* section 2.3.3), ce qui nous permettra donc d'étudier les rôles d'auteur dans les écrits scientifiques en nous inspirant de cette étude.

8.2.3. Les pronoms personnels et les rôles d'auteur des routines

Les pronoms auteurs sont des éléments énonciatifs méritant une observation plus approfondie. Dans notre corpus, 44,93 % d'occurrences (on reviendra dans la section 8.3.1.4) dont 15 sur 37 routines au total incluent un pronom personnel renvoyant à l'auteur. *nous* et *on*, y sont plus utilisés que *je*. L'emploi de ces pronoms reflète donc une forte particularité des rôles d'auteurs du discours scientifique, en particulier les articles de recherches. Notre objectif n'est pas de mener une étude exhaustive sur les pronoms personnels¹⁰⁹, mais d'identifier d'une

¹⁰⁹ Les pronoms personnels dans les écrits scientifiques font l'objet d'étude de nombreuses études. On peut citer l'ouvrage *L'écrit scientifique : du lexique au discours* (2013) et la revue *lidil* N° 41 *Énonciation et rhétorique dans l'écrit scientifique* en 2010.

manière générale des rôles qu'occupe l'auteur, en observant les classes sémantiques des verbes associés. En outre, nous mènerons aussi une observation rapide sur les trois pronoms personnels *je, on et nous*.

8.2.3.1 *Les rôles d'auteur scientifique à travers les pronoms personnels*

Rappelons que Fløttum *et al.* (2006, p. 82–86) proposent quatre rôles rhétoriques de l'auteur à travers une étude sur les pronoms à la première personne dans les articles de recherche en français, en anglais et en norvégien : a) l'énonciateur se charge de la tâche d'écriture ; b) le chercheur explique la recherche ; c) l'argumentateur qui développe le raisonnement ; d) l'évaluateur exprime l'évaluation et l'émotion. Ces rôles ont été identifiés à l'aide des verbes associés et des éléments métatextuels dans le contexte. L'étude montre que les rôles principaux joués par l'auteur scientifique concernent celui de chercheur (46 %), celui d'énonciateur (26 %), celui d'argumentateur (14 %), ou encore deux rôles à la fois comme énonciateur & chercheur (5 %) et énonciateur & argumentateur (2 %).

Les routines de guidage ont pour objectif principal d'accompagner le lecteur en rendant plus clair et facile à lire le discours. Nous supposons donc que c'est principalement le rôle d'énonciateur qui sera ici en jeu. Pour tester notre hypothèse, nous récapitulons les pronoms avec les verbes mis en œuvre dans les 15 routines :

Fonctions rhétoriques	Routines	Pronoms personnels	Verbes les plus fréquents associés à la routine
Structuration	N° 2 de routines de planification (<i>dans la section 2, nous présentons...</i>)	<i>je</i> (3), <i>on</i> (1), <i>nous</i> (37)	Verbes d'ANALYSE_INFO : <i>aborder, analyser, examiner, étudier, considérer, présenter, caractériser, décrire, illustrer, représenter,</i> Verbes de DIFFUSION : <i>proposer, exposer, discuter, préciser, formuler...</i>
	N° 5 de routines d'annonce de l'objectif d'étude (<i>nous proposons dans cet article d'analyser</i>)	<i>je</i> (4), <i>on</i> (6), <i>nous</i> (58)	Verbes de FINALITÉ : <i>chercher à, tenter, viser à, (s')attacher</i> Verbes de PROPOSITION : (<i>s'</i>) <i>proposer</i>
	N° 6 de routines d'annonce de l'objectif d'étude (<i>nous nous intéressons à</i>)	<i>je</i> (1), <i>on</i> (3), <i>nous</i> (79)	Verbes de THÈME : (<i>s'</i>) <i>intéresser, (se) focaliser, (s')interroger</i>
	N° 2 de routines de navigation textuelle (<i>comme nous allons le voir</i>)	<i>je</i> (11), <i>on</i> (200), <i>nous</i> (203)	Verbes de CONSTAT : <i>voir, noter,</i> Verbes de DIFFUSION : <i>dire, mentionner, indiquer, préciser, exposer</i>
	N° 4 de routines de navigation textuelle (<i>je reviendrai sur ce point</i>)	<i>je</i> (13), <i>on</i> (24), <i>nous</i> (47)	Verbes de PASSAGE : <i>passer, revenir</i>
	N° 8 de routines de navigation textuelle (<i>on peut se référer à quelqu'un</i>)	<i>je</i> (7), <i>on</i> (56), <i>nous</i> (4)	Verbes de RÉFÉRENCE : <i>se référer, citer, reporter, mentionner</i>
	N° 1 de routines de conclusion (<i>on peut conclure que</i>)	<i>je</i> (1), <i>on</i> (30), <i>nous</i> (14)	Verbes d'INFÉRENCE : <i>conclure, déduire</i>
Métalinguistique	N° 1 de routines de reformulation (<i>on peut également dire</i>)	<i>je</i> (0), <i>on</i> (23), <i>nous</i> (1)	Verbes de DIFFUSION : <i>parler, dire</i>
	N° 1 de routines de réparation (<i>on ne peut pas dire</i>)	<i>je</i> (1), <i>on</i> (11), <i>nous</i> (3)	Verbes de DIFFUSION : <i>parler, dire</i>
	N° 2 de routines de précision terminologique (<i>on définit X comme</i>)	<i>je</i> (0), <i>on</i> (14), <i>nous</i> (12)	Verbes de CONCEPTION : <i>définir, concevoir, qualifier</i>
	N° 3 de routines de précision terminologique (<i>on appelle X Y</i>)	<i>je</i> (24), <i>on</i> (61), <i>nous</i> (65)	Verbes de DÉNOMINATION : <i>nommer, appeler,</i>
	N° 5 de routines de précision terminologique (<i>j'utilise le terme X pour désigner Y</i>)	<i>je</i> (3), <i>on</i> (31), <i>nous</i> (12)	Verbes d'USAGE : <i>utiliser, employer</i>
	N° 1 de routines d'emprunt terminologique (<i>on reprend le terme de quelqu'un</i>)	<i>je</i> (5), <i>on</i> (5), <i>nous</i> (5)	Verbes d'USAGE : <i>reprendre, utiliser, employer, mobiliser</i>
Topicalisation	N° 1 de routines de mise en évidence (<i>nous pouvons mentionner</i>)	<i>je</i> (6), <i>on</i> (65), <i>nous</i> (18)	Verbes de DIFFUSION : <i>préciser, indiquer, mentionner, souligner</i> Verbes d'INSISTANCE : <i>insister</i>
	Routine de (co)constat (<i>nous pouvons noter que</i>)	<i>je</i> (8), <i>on</i> (262), <i>nous</i> (45), notons (15)	Verbes de CONSTAT : <i>noter, constater, observer, relever, remarquer, repérer, voir</i>
TOTAL		<i>je</i> (87), <i>on</i> (792), <i>nous</i> (603), notons (15)	

Tableau 8-3 : Récapitulation des pronoms personnels et des verbes dans les routines de guidage

En combinant les verbes entrant dans les routines avec les pronoms, ainsi que leurs fonctions rhétoriques dans les écrits scientifiques, nous avons observé que la plupart des

pronoms personnels dans les routines de guidage jouent principalement deux rôles : le rôle d'énonciateur et le rôle de chercheur, ou les deux à la fois.

L'identification des rôles d'auteur à l'aide des verbes n'est pas toutefois sans problème. En effet, nous remarquons que le classement des verbes de Fløttum *et al.* (2006, p. 82–86) peut engendrer des interprétations erronées des rôles d'auteur, car le sens d'un verbe peut varier selon le contexte. Par exemple, Fløttum *et al.* (2006, p. 84) classent le verbe *proposer* comme un verbe lié au rôle d'argumentateur. Néanmoins, *proposer* est qualifié comme verbe de COMMUNICATION dans LST qui correspond au rôle d'énonciateur. Nous pensons donc qu'il peut jouer les deux rôles à la fois. En outre, les éléments métatextuels (*dans la section 2, dans cet article, plus haut...*) et les fonctions rhétoriques de ces routines dans le discours scientifique peuvent aussi aider à mieux comprendre les rôles d'auteur.

Nous résumons nos remarques sur les rôles d'auteur des routines dans le tableau 8-4 (+ indique oui, - indique non, ++ indique plus représentatif dans le cas d'avoir les deux rôles en même temps) :

Routines	Énonciateur	Chercheur
N° 2 de routines de planification (<i>dans la section 2, nous présentons...</i>)	++	+
N° 5 de routines d'annonce de l'objectif d'étude (<i>nous proposons dans cet article d'analyser</i>)	++	+
N° 6 de routines d'annonce de l'objectif d'étude (<i>nous nous intéressons à</i>)	++	+
N° 2 de routines de navigation textuelle (<i>comme nous allons le voir</i>)	+	-
N° 4 de routines de navigation textuelle (<i>je reviendrai sur ce point</i>)	+	-
N° 8 de routines de navigation textuelle (<i>on peut se référer à quelqu'un</i>)	+	++
N° 1 de routines de conclusion (<i>on peut conclure que</i>)	+	++
N° 1 de routines de reformulation (<i>on peut également dire</i>)	+	-
N° 1 de routines de réparation (<i>on ne peut pas dire</i>)	+	-
N° 2 de routines de précision terminologique (<i>on définit X comme</i>)	-	+
N° 3 de routines de précision terminologique (<i>on appelle X Y</i>)	++	+
N° 5 de routines de précision terminologique (<i>j'utilise le terme X pour désigner Y</i>)	+	++
N° 1 de routines d'emprunt terminologique (<i>on reprend le terme de quelqu'un</i>)	+	++
N° 1 de routines de mise en évidence (<i>nous pouvons mentionner</i>)	+	-
Routine de (co)constat (<i>nous pouvons noter que</i>)	-	+

Tableau 8-4 : Récapitulation des rôles d'auteur dans les 15 routines de guidage avec pronom personnel

Dans l'ensemble, les routines avec les pronoms personnels permettent à l'auteur de jouer les rôles d'énonciateur et de chercheur, le premier rôle étant plus marqué : 13/15 routines, tandis que le rôle de chercheur occupe 10/15 routines. Ce résultat correspond à notre hypothèse.

Dans le rôle d'énonciateur, les routines de reformulation, de réparation, de mise en évidence n'utilisent que des verbes de COMMUNICATION, en clarifiant le code langagier ou en attirant l'attention sur l'information essentielle. L'auteur dans ces routines s'incarne comme étant l'énonciateur du discours chargé de vérifier l'encodage et de communiquer les idées.

La routine N° 5 de précision terminologique (*j'utilise le terme X pour désigner Y*) et la routine N° 1 d'emprunt terminologique renvoient clairement au rôle de chercheur. Il en va de même pour la routine de (co)constat qui relève nettement d'une activité de recherche.

La coexistence du rôle d'énonciateur et de celui de chercheur est aussi manifeste (plus de la moitié de routines : 8/15), comme dans la routine N° 3 de précision terminologique (*on appelle X Y*). Dans l'exemple 36, on voit bien que la dénomination d'une question n'est pas formulée par hasard, mais correspond à la réflexion de l'auteur après l'observation sur les caractéristiques de cette question. L'auteur montre donc une partie de son identité en tant que chercheur capable de choisir une dénomination pour désigner une question de recherche spécifique. En parallèle, l'acte de dénommer un phénomène par une expression (*question socialement vive*) est bien métalinguistique, permet aussi à l'auteur de jouer le rôle d'énonciateur.

[36] Nous appellerons "question socialement vive" une question qui possède les caractéristiques suivantes. (Sciences de l'éducation)

Les pronoms personnels en question (sauf *je*) peuvent renvoyer au lecteur ou aux autres en dehors de l'auteur. Dans ces cas précis, manifestent-ils des différences dans les routines ? Nous allons explorer cela dans la section suivante.

8.2.3.2 *L'observation des pronoms personnels et l' enrôlement du lecteur*

Les trois pronoms personnels présents dans les routines ont des renvois différents que nous résumons ici :

Pronoms personnels	Renvois dans les articles de recherche à
<i>JE</i> (87)	exclusivement l'auteur
<i>NOUS</i> (647) et <i>ON</i> ¹¹⁰ (740)	1. L'(es)auteur(s) 2. L'(es)auteur(s) + le(s) lecteur(s) 3. L'(es)auteur(s) + les collègues de la communauté discursive 2. L'(es)auteur(s) + tout le monde

Tableau 8-5 : Renvois des pronoms personnels dans les routines

Il est évident que le pronom *on* est le plus présent, tandis que *je* est le moins présent. *Je* est plus présent dans deux routines : la routine N° 5 de précision terminologique (*on appelle X Y*) (24/150) et la routine N° 1 d'emprunt terminologique (*on reprend le terme de quelqu'un*) (5/15). Dans ces deux routines, les pronoms *nous* et *on* désignent plutôt l'auteur, car l'acte de dénomination et d'emprunt doit être l'acte de l'auteur. De la même façon, les pronoms personnels *on* et *nous* dans la routine N° 2 de précision terminologique (*on définit X comme*) et

¹¹⁰ Le pronom personnel *on* possède encore deux valeurs liées à 1) le(s) lecteur(s) ; 2) les autres chercheurs ((Fløttum et al., 2006, p. 117), qui font l'objet d'étude dans les routines de guidage, parce que nous l'avons exclu au moment du nettoyage manuel des occurrences de routines (voir le Chapitre 4).

dans la routine N° 1 de mise en évidence (*nous pouvons mentionner*). Dans la routine de mise en évidence, les verbes *mentionner, signaler, remarquer, souligner* ont comme agent de l'action exclusivement l'auteur. Dans les routines de planification, d'annonce, de reformulation et de réparation, nous soulignons le même renvoi pour les pronoms personnels (*je, nous, on*).

Par contre, les *on* et *nous* dans les routines de navigation textuelle, de conclusion et de (co)constat sont plus complexes. Ils sont souvent inclusifs et enrôlent le lecteur dans l'énonciation pour l'inviter à enrôler le lecteur, notamment avec les verbes de CONSTAT (*noter, remarquer, constater, observer, relever, voir...*). En revanche, les pronoms personnels apparaissent avec les verbes de COMMUNICATION dans ces routines comme *présenter, souligner, mentionner, dire* ne renvoient qu'à l'auteur. C'est aussi la raison pour laquelle le pronom *je* n'apparaît qu'avec quatre verbes de COMMUNICATION dans la routine de navigation textuelle (*dire, signaler, mentionner, souligner*), et est très rare dans la routine de (co)constat.

8.2.4. Synthèse

Les routines de guidage dans les écrits scientifiques peuvent s'inscrire dans la situation de l'énonciation de plusieurs manières. Certaines actualisations sont très apparentes et faciles à détecter : c'est le cas des déictiques, des pronoms personnels, des anaphores, des temps verbaux et des modalités dans les routines. D'autres sont moins explicites et doivent être observées dans les propositions ou syntagmes qui suivent les routines. Il s'agit ici des marqueurs métatextuels, discursifs, de temporalité et de localisation qui apparaissent avec les routines, ainsi que les anaphores dans la partie d'extension des routines qui permettent d'intégrer les routines dans l'énonciation. Il faut constater que la plupart de routines s'actualisent de façon explicite, ce qui est correspond aussi aux caractéristiques énonciatives des routines que nous avons développées dans le premier Chapitre.

Les pronoms personnels constituent des éléments très intéressants pour étudier les traces de l'auteur dans les écrits scientifiques. Nous avons examiné les rôles d'auteur associés aux verbes, aux éléments métatextuels et aux fonctions rhétoriques des routines. Cet examen permet de montrer que l'approche de l'identification des rôles d'auteur à l'aide des verbes proposée par que Fløttum et *al.* (2006, pp. 82–86) ne peut pas être appliquée directement sans vérification dans le contexte (comme nous l'avons souligné dans le Chapitre 2). Notre observation, qui prend en compte aussi les fonctions rhétoriques des routines, permet de mettre en lumière que l'auteur des routines de guidage, comme ce que nous avons supposé, joue en priorité le rôle d'énonciateur du discours, mais aussi le rôle de chercheur qui est essentiel pour un auteur

scientifique. Enfin, nous notons que les pronoms personnels renvoient aux différents groupes de protagonistes dans les articles de recherche. Leurs distributions dans les routines présentent un lien étroit avec leurs fonctions. Les *on* et *nous* inclusifs sont plus fréquents dans les routines qui enrôlent le lecteur pour par exemple le (co)constat ou la navigation dans le texte, mais les pronoms *on* et *nous* peuvent renvoyer seulement à l'auteur, en particulier dans les routines dont les verbes désignent un acte de l'auteur comme la définition (*on définit X comme*), l'annonce de l'objectif (*nous nous intéressons dans cet article à*). *Je* est plus présent dans les routines qui dénotent l'acte de l'auteur, comme la dénomination (*on appelle X Y*).

8.3. Distribution des routines de guidage

Dans cette section, nous voulons mettre en lumière la répartition des routines ainsi que leur emploi à travers les dix disciplines de notre corpus : économie, linguistique, psychologie, sociologie, sciences politiques, science de l'éducation, science de l'information, géographie et histoire. Notre observation quantitative des routines permettra de mettre en évidence des éléments intéressants sur les pratiques d'écriture dans les disciplines.

Nous proposerons d'abord d'observer les résultats principaux sur la répartition des routines de guidage dans le corpus global, ainsi que les proportions des sous-fonctions pour chacune des trois catégories principales (structuration, métalinguistique et topicalisation), afin de rendre compte de l'emploi général de ces routines. Ensuite, nous procéderons à des observations sur les routines à travers les disciplines en relevant dans le détail les différences disciplinaires.

8.3.1. Répartition des routines dans le corpus global

Rappelons que notre typologie a été élaborée en nous basant sur les classements interactifs et métatextuels des modèles de Hyland et d'Ädel (*cf.* section 3.2.3). Le modèle de Hyland se focalise sur les marqueurs courts, tandis que celui d'Ädel porte sur des phrases qui sont plus proches de nos routines. Ces deux modèles proposent également des observations quantitatives qui peuvent servir de point de départ pour formuler nos hypothèses. Nous récapitulons d'abord les répartitions des fonctions issues de ces deux études, dans le tableau 8-6 :

Hyland (2005, p. 92)¹¹¹	Fréquences/1000 mots
Connexion logique (marqueurs de transition : addition, comparaison, conséquence)	12,8
Se référer aux actes, séquences ou étapes du discours (marqueurs d'encadrement : structuration, étiquetage des étapes discursives, anticipation, changement thématique)	5,0
Référence intratextuelle (marqueurs enphoriques)	4,0
Référence intertextuelle (marqueurs d'évidentialité)	6,9
Explication et clarification (marqueurs de gloses)	7,1
Ädel (2006, p. 79)¹¹²	Fréquences/100 000 mots
Sur le code : (fonctions métalinguistiques)	
Définition	18
Commentaire énonciatif	14
Sur la structure du texte : (fonctions métatextuelles)	
Introduction du thème	22
Focus sur le thème	11
Conclusion	5
Exemplification	4
Rappel	16
Ajout	1
Argumentation	1
Contextualisation	3

Tableau 8-6 : Fréquences des fonctions de Hyland (2005, p. 92) et d'Ädel (2006, p. 79) citées dans notre étude

De manière générale, on peut constater dans l'étude de Hyland que les marqueurs de structuration (les 4 premiers parmi les 5 fonctions) sont plus fréquents que les marqueurs de gloses : 28,7 contre 7,1. Dans l'étude d'Ädel, on observe le même résultat : les expressions des fonctions métatextuelles sont plus nombreuses que celles des fonctions métalinguistiques (63 contre 32). Quant aux marqueurs liés à la topicalisation, ils sont classés dans ces deux études comme éléments intégrants des fonctions structurantes : dans les marqueurs d'encadrement (5,0) et la fonction de focus sur le thème (11). Ce sont les moins nombreux. Rappelons de notre côté que notre définition de la topicalisation n'est pas totalement identique à celle de ces deux études, ce qui pourrait aboutir à des résultats différents.

À partir de ces observations effectuées dans les écrits scientifiques en anglais, et à partir des routines identifiées dans notre corpus (18 routines de structuration, 13 routines métalinguistiques, 6 routines de topicalisation), nous pouvons supposer que les écrits scientifiques en français ont tendance à employer plus de routines de structuration et moins de routines de topicalisation, en termes de fréquence d'occurrences.

¹¹¹ L'observation de ces marqueurs métadiscursifs est effectuée dans 28 articles de recherche en quatre disciplines : Microbiologie, Marketing, Astrophysique et Linguistique appliquée avec 160 000 mots au total.

¹¹² Nous n'avons retenu ici que les observations sur les corpus natifs d'environ 247 000 mots au total, qui comprennent 290 essais écrits par des étudiants universitaires américains et britanniques. L'étude d'Ädel utilise aussi un corpus des apprenants suédois d'anglais pour une étude contrastive de l'emploi des éléments métadiscursifs entre les natifs et non-natifs.

Nous proposons d'observer la répartition des trois fonctions principales des routines de guidage dans la figure 8-1. Les résultats valident partiellement notre hypothèse : les routines de structuration sont effectivement les plus nombreuses, mais les routines de topicalisation s'avèrent plus fréquentes que les routines métalinguistiques.

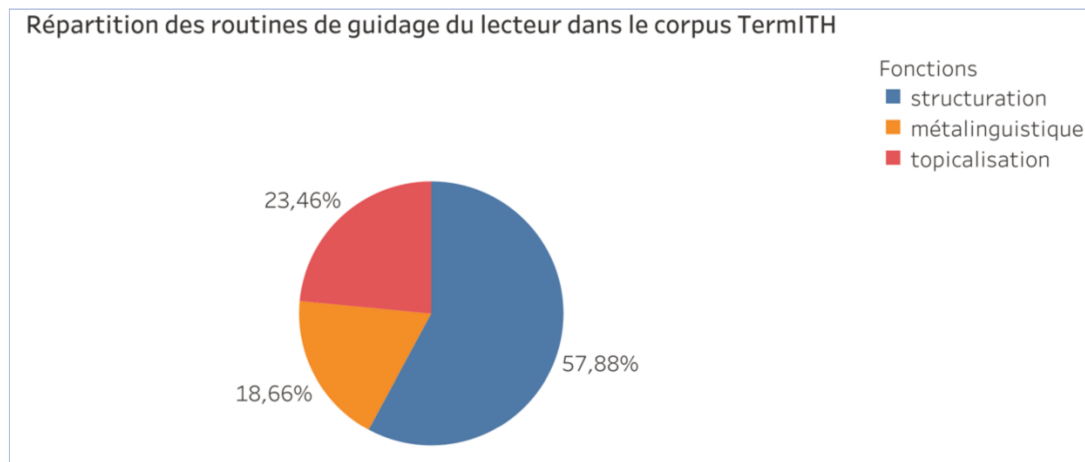


Figure 8-2 : Proportion des fréquences d'occurrence des trois fonctions des routines de guidage

Nous pouvons ainsi dire que les articles de recherches en sciences humaines et sociales révèlent un vif besoin de structurer discursivement le contenu propositionnel. D'une part, les routines de structuration remplissent des fonctions visant à organiser les parties textuelles, comme informer sur le plan du texte, annoncer au lecteur les objectifs ou objets principaux du texte, faire voyager le lecteur dans le texte pour identifier des informations plus précises, clôturer une chaîne argumentative ou une partie textuelle, voire le texte entier. D'autre part, les routines de topicalisation, en tant qu'ensemble autonome, visent principalement à la hiérarchisation des informations dans le discours afin de souligner les plus importantes pour le lecteur. On peut dire que la structuration porte sur le plan structurel, et la topicalisation se focalise sur le plan informationnel. Les deux fonctions peuvent donc toutes deux aider à organiser le discours, mais sur des plans différents.

Selon cette figure, il semble que les routines métalinguistiques soient les moins nombreuses, occupant une place de 18,66 % des occurrences. Cependant, ce chiffre proche de celui de la topicalisation (23,46 %), nous permet d'avancer que les routines veillant à l'emploi de la langue s'avèrent également essentielles pour les articles de recherche. Il faut par ailleurs tenir compte du fait que les écrits scientifiques en français emploient beaucoup de marqueurs courts à fonction métalinguistique au lieu d'expressions longues comme les routines : on utilise notamment les marqueurs de reformulation comme *en d'autres termes*, *c'est-à-dire*. À titre indicatif, Tran (2014) identifie 1390 occurrences des marqueurs de reformulation dans

291 articles de recherche en SHS, alors que nous n'avons observé que 622 occurrences de routines métalinguistiques dans 500 articles (nous reviendrons sur ce point dans la section 8.2.1.2.). Passons maintenant à l'observation des emplois des routines au sein de ces trois catégories principales.

8.3.1.1 *Routines de structuration : proportion et fréquences*

Parmi les routines de structuration, celles de navigation sont les plus nombreuses dans les articles de recherche, suivies par les routines d'annonce de l'objectif/l'objet (voir la figure 8-2). Les deux autres fonctions sont moins fréquentes. La fréquence de la fonction de navigation s'explique par le fait que ces propositions métadiscursives enchaînent le matériau propositionnel, assurant donc sur le plan textuel la cohésion du discours.

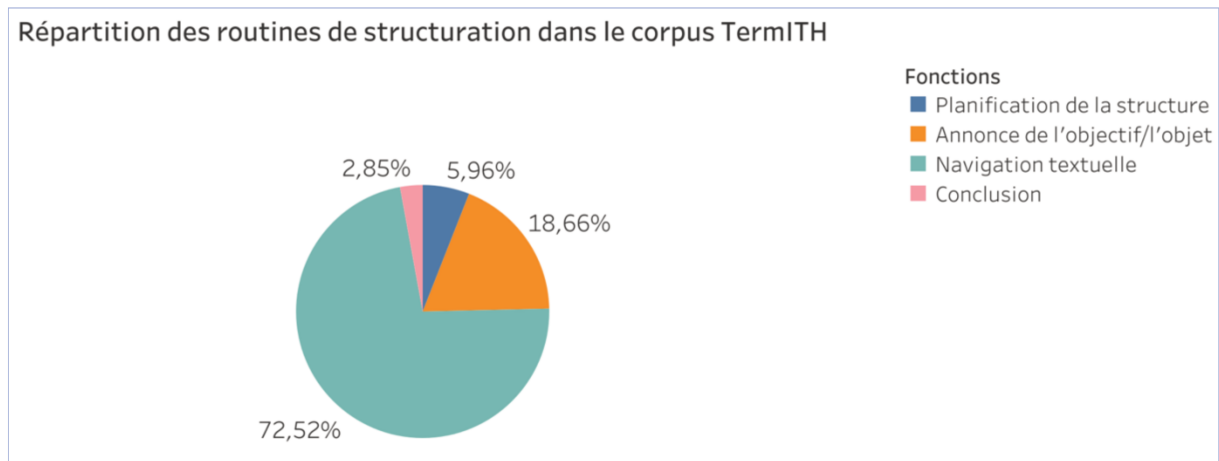


Figure 8-3 : Proportion des quatre sous-fonctions dans la structuration

Les routines de planification de la structure sont moins nombreuses. Deux raisons peuvent expliquer la faible présence des routines de planification : 1) Les articles de recherches ne contiennent pas forcément un plan clairement annoncé, comme le montre la figure 8-3. Seule l'économie semble y recourir de façon fréquente, tandis qu'à l'opposé on voit que l'histoire et l'anthropologie n'annoncent pas le plan avec les routines. 2) Le plan n'est annoncé qu'une fois dans un texte, alors qu'on peut avoir plusieurs renvois.

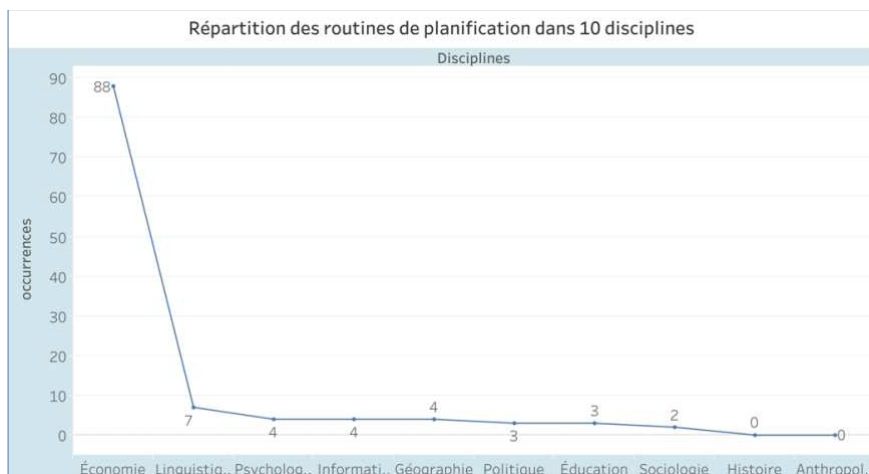


Figure 8-4 : Occurrences des routines de planification à travers les disciplines

En ce qui concerne les routines de conclusion, comme nous l'avons expliqué dans la section 5.3.4, ces marqueurs de conclusion (*pour conclure, en conclusion, pour finir en résumé, en un mot*) (Tran, 2014, p. 94) peuvent remplir aisément la fonction de conclusion. L'utilisation des propositions plus complexes, comme les routines, n'est donc pas nécessairement utile face aux marqueurs de conclusion. Cela explique le faible taux de routines de conclusion/déduction dans notre corpus.

De façon plus détaillée, la figure 8-4 présente les occurrences des routines de structuration par ordre décroissant pour chaque sous-classe. Certaines routines sont très fréquentes comme la routine d'annonce N° 1 (*L'objectif de cet article est d'analyser... notre objectif consiste à évaluer...*), de navigation textuelle N° 7 (*voir l'article de quelqu'un, se référer à l'ouvrage de quelqu'un*), de navigation textuelle N° 2 (*comme nous l'avons dit, on va le voir*) et de navigation textuelle N° 6 (*voir la note/annexe...*)

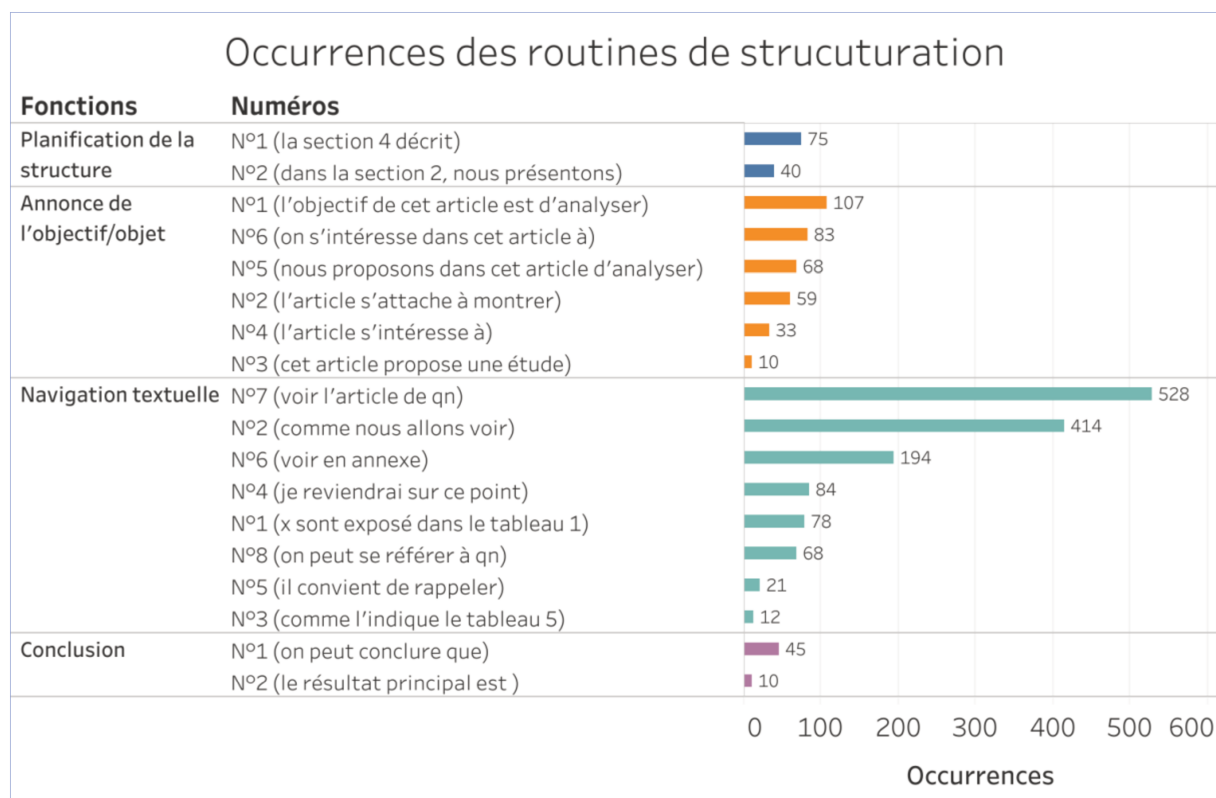


Figure 8-5 : Occurrences des routines de structuration

Il existe aussi trois routines qui ont une fréquence inférieure à 15. Ce sont les cas de la routine d'annonce N° 3 (*L'article propose une analyse... Cet article propose une étude...*), de navigation N° 3 (*comme l'indique le tableau*), de conclusion N° 2 (*Le premier résultat important est que... Le résultat principal de notre étude est de montrer...*). Les occurrences des routines peuvent refléter la préférence des configurations syntaxiques et sémantiques dans les articles de recherche pour une même fonction. Il semble que l'on ait tendance à utiliser plus souvent la structure N° 1 *la section 1 décrit* que la structure N° 2 avec le pronom auteur *dans la section 2, nous présentons* pour annoncer le plan. Parmi les routines d'annonce d'objectif d'étude, la routine N° 1 (*l'objectif de cet article est d'analyser*) est la plus présente, suivie par la routine N° 6 (*on s'intéresse dans cet article à*). Par contre, les routines qui prennent *l'article* comme sujet sont moins nombreuses (N° 2 et 4 et 3). On peut dire que les articles de recherche tendent à indiquer l'objectif soit directement avec le mot *objectif*, soit avec des propositions qui permettent de mettre en évidence la présence de l'auteur en tant qu'énonciateur du discours.

Pour la navigation textuelle, la routine N° 7 (navigation intertextuelle, *voir l'article de qn*), la routine N° 2 (navigation intertextuelle d'anticiper *on va le voir*, et de rappel *nous l'avons dit précédemment*) et la routine N° 6 (navigation intratextuelle, *voir en annexe*) l'emportent sur les autres routines de navigation, même sur toutes les routines de structuration. Ces trois

routines sont fondamentales dans les écrits scientifiques. Comme nous l'avons souligné dans le chapitre 2, les écrits scientifiques constituent un lieu particulièrement interdiscursif en se référant aux discours des pairs, afin de se positionner dans la discipline. L'abondance de la routine N° 7 met en évidence cette propriété dialogique des écrits scientifiques. En parallèle, les routines N° 2 et N° 6 permettent de créer des liens cohésifs entre les empan textuels, ce qui assure la cohésion d'un texte scientifique.

Nous voulons souligner aussi la prédominance des routines sans trace de l'auteur (11 routines dans 18 au total, avec un taux d'occurrence de 58,37 %). On va voir que cette préférence n'est pas toujours valide dans les autres routines.

8.3.1.2 *Routines métalinguistiques : proportions et fréquences*

Nous nous intéressons maintenant aux routines métalinguistiques dans le corpus. Selon la figure 8-5 présentée ci-dessous, les routines de précision terminologique représentent la majorité des routines retenues dans la catégorie « métalinguistique ». Il est logique d'expliquer ce point par un besoin essentiel du discours scientifique : l'auteur doit s'assurer de la bonne compréhension du terme disciplinaire employé, que ce terme existe déjà dans le domaine ou qu'il s'agisse d'une nouvelle dénomination forgée par l'auteur.

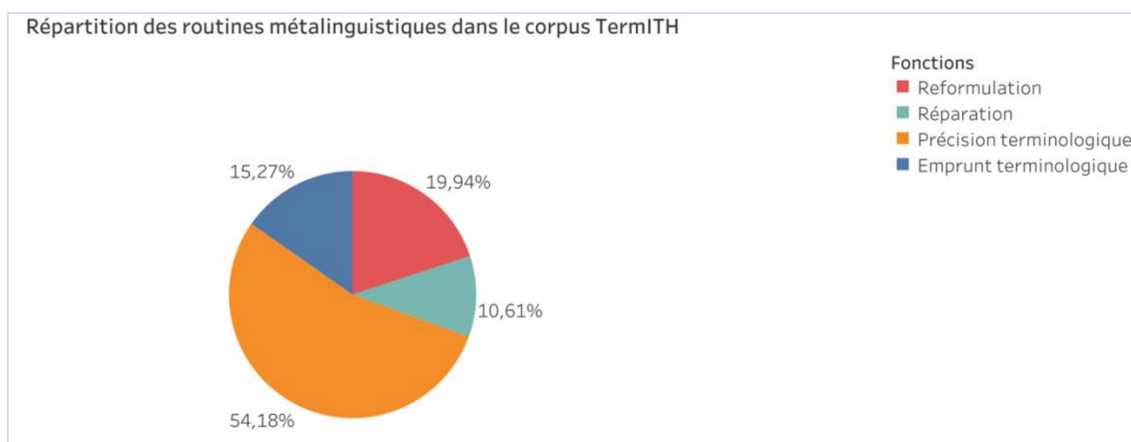


Figure 8-6 : Proportion des quatre sous-fonctions métalinguistiques dans le corpus

La reformulation se classe en seconde place, alors que la réparation et l'emprunt terminologique sont moins productifs. Néanmoins, il ne serait pas pertinent d'affirmer que le discours mobilise peu la reformulation pour éclairer les termes. Ici, il ne faut pas oublier que la reformulation se réalise plus couramment avec des marqueurs courts comme *c'est-à-dire*, *autrement dit*. Nous l'illustrons ici en comparant les occurrences des marqueurs de reformulation dans l'étude de Tran (2014) et celles des routines métalinguistiques dans notre corpus.

	Marqueurs de reformulation	Occurrences	Au total
Marqueurs de reformulation chez Tran (2014) dans 291 articles	<i>C'est-à-dire</i>	704	1390
	<i>Autrement dit</i>	197	
	<i>À savoir</i>	139	
	<i>En d'autres termes</i>	119	
	<i>En somme</i>	57	
	<i>En résumé</i>	21	
	<i>En un mot</i>	14	
	<i>En définitive</i>	54	
	<i>Au total</i>	49	
	<i>Somme toute</i>	13	
	<i>Au fond</i>	12	
<i>En fin de compte</i>	11		
Routines métalinguistiques dans notre corpus de 500 articles	Reformulation	124	622
	Réparation	64	
	Précision terminologique	339	
	Emprunt terminologique	95	

Tableau 8-7 : Comparaison entre les marqueurs de reformulation (Tran, 2014, p.298) et les routines métalinguistiques

En comparant ces données avec celles des routines métalinguistiques (622 occurrences au total, dont 124 occurrences des routines de reformulation), nous pouvons constater que, dans 291 articles, les marqueurs de reformulation ont déjà dépassé l'ensemble des occurrences des routines métalinguistiques repérées dans 500 articles. On peut en déduire que l'opération de reformulation se réalise davantage par des marqueurs courts : cela est probablement dû au fait que ces marqueurs disposent déjà d'une forte capacité à opérer une reformulation, avec la simplicité liée au figement en termes syntaxique et sémantique. Il est possible qu'il existe une complémentarité entre les marqueurs courts et les routines, car ces deux types de marqueurs présentent des différences rhétoriques dans la manière de « dire ». De nombreux marqueurs courts de reformulation intègrent une dimension de changement de point de vue ou introduisent une synthèse (*en somme, en résumé, en définitive, au total, somme toute, au fond et en fin de compte*). Les routines de reformulation et de réparation quant à elles utilisent spécifiquement les mots à fonction métalinguistique *dire, parler et signifier* avec des formes moins figées et soulignent de ce fait plus explicitement l'intention de reformulation.

Pour comprendre quelle routine est la plus utilisée dans notre corpus, nous proposons un examen plus détaillé par routines. Nous pouvons identifier dans la figure 8-6 trois routines très fréquentes :

- 1.a N° 3 de précision métalinguistique. Il s'agit d'une routine de dénomination <L' **auteur** 'attribue un nom - Y' à un concept/phénomène/fait scientifique X> (*on appelle X Y, on désigne X (par le terme) Y*) ;
- 1.b N°2 de reformulation < 'L' **énoncé source** est 'reformulé' > (*cela veut dire que*) ;
- 1.c N° 2 d'emprunt terminologique <Un ' **terme/définition** ' à **quelqu'un d'autre que**

l'auteur est **'emprunté(e)'**> (*pour reprendre la définition de quelqu'un*).

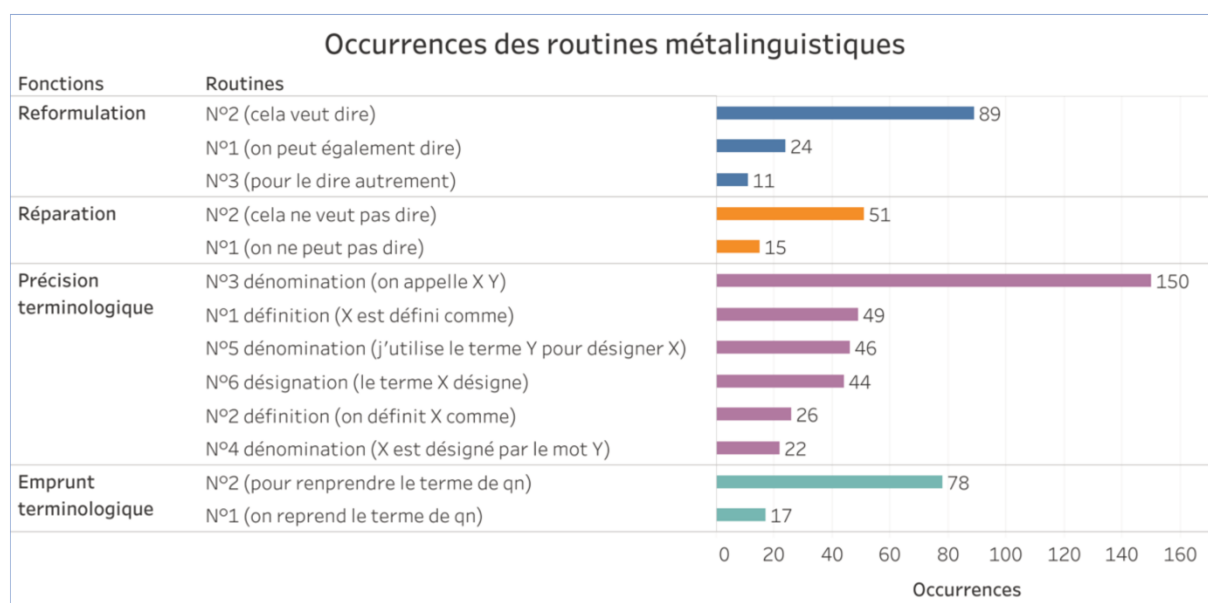


Figure 8-7 : Occurrences des routines métalinguistiques

Pour chaque sous-fonction, nous remarquons également une légère préférence pour les configurations lexico-syntaxiques sans recours aux pronoms personnels (7 routines parmi les 13, dont 54,98 % d'occurrences). On peut noter ici la routine de reformulation N° 2 (*[cela/ce qui] [veut dire/signifie] que*), de reformulation N° 3 (*pour (le) dire autrement*), de réparation N° 2 (*[cela/ce qui] [ne veut dire/signifie] pas que*), de précision terminologique N° 1 (*X est [défini/conçu] comme...*), N° 4 (*X est désigné par [le terme/le mot/l'expression]*) et N° 6 (*[le terme/le mot/l'expression] [désigne/signifie]...*) dans la même catégorie, enfin, d'emprunt terminologique N° 2 (*pour [reprendre/utiliser] [le mot/l'expression/la définition/le vocabulaire/la notion/le terme] de quelqu'un*). La présence de l'auteur dans les routines métalinguistiques a tendance à s'effacer derrière des stratégies visant à rendre plus clair l'emploi du code langagier.

8.3.1.3 Routines de topicalisation : proportion et fréquences

Nous avons retenu 3 sous-fonctions et 6 routines de topicalisation. Selon la figure 8-7, on peut observer que les routines de mise en évidence et de (co)constat sont beaucoup plus nombreuses que les routines de focus, en occupant respectivement 47,19 % (369 occurrences) et 42,20 % (330 occurrences) du total. Le taux faible des routines de focus peut être partiellement expliqué par le fait qu'une routine de structuration, plus précisément la routine N° 6 d'annonce d'objectif <L' **'auteur** **'se focalise sur'** un thème (dans le **'discours en cours'**)> peut remplir également une fonction de focus quand elle se réalise lexicalement plus

simplement comme *je me centre sur/nous nous intéressons à...* mais cette routine a été classée dans les routines de structuration.

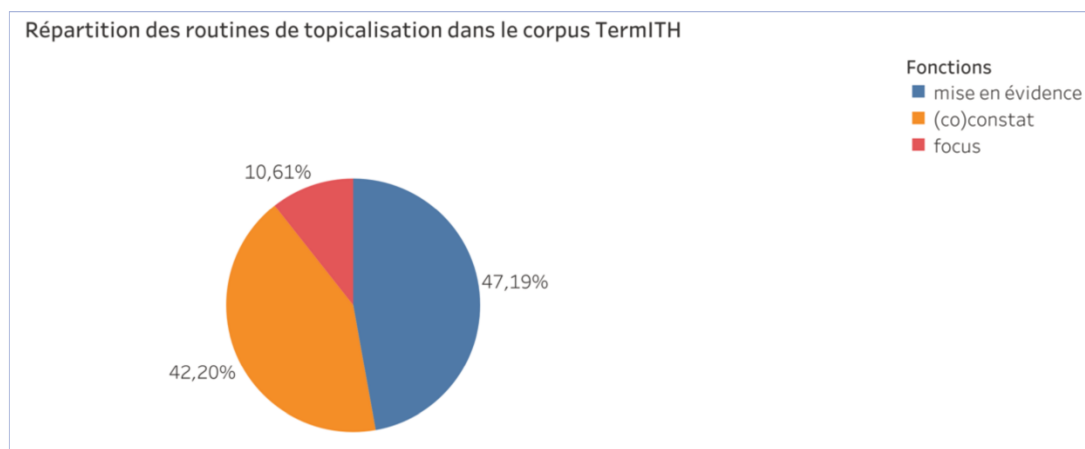


Figure 8-8 : Répartition des trois sous-fonctions dans topicalisation

Plus précisément, on distingue trois routines fréquentes dans la figure 8-8 : les routines N° 1 (< **L'’auteur’ ‘met en évidence’** une information pour son lecteur> [*on/nous/je*] [*souligner/insister*]) et N° 3 (< Une information/un constat **‘nécessite’** d’être **‘mis(e) en évidence’** > (il [*faut/importe de/convient de*] [*noter/constater/remarquer/relever/révéler*]) de la mise en évidence et la routine de (co)constat (< **L'’auteur’ ‘fait un constat’** (ou invite le lecteur à faire un constat) après une observation > [*on/nous/je*] (*vouloir/pouvoir*) [*noter/constater/remarquer/observer/voir...*]).

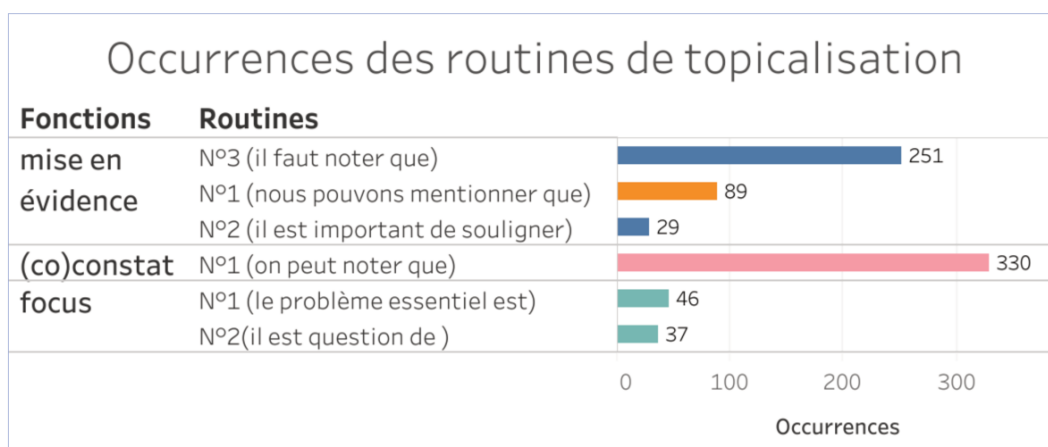


Figure 8-9 : Occurrences des routines de topicalisation

En outre, nous notons que les routines aux structures impersonnelles sont très usuelles dans notre corpus : les routines de mise en évidence N° 2 (*il est [important/fondamental/essentiel...] de [noter/constat/remarquer/observer/préciser...]*), N° 3 et N° 4 (*il est à [noter/observer/remarquer]*), la routine de focus N° 2 (*il est question de*). Les deux premières routines citées plus haut peuvent se transformer en structure <Pronom auteur +

Verbe> comme la routine N° 1 de mise en évidence et la routine de (co)constat. Par ailleurs, ces phrases impersonnelles permettent de rendre plus évident l'élément à topicaliser, comme l'indiquent Abeillé & Godard (2021, p. 1970) :

La construction impersonnelle peut s'accompagner d'une distinction entre fond et focus. Dans certains contextes, en effet, le référent introduit par la construction peut être présenté comme le focus, le reste de la phrase constituant le fond.

Ainsi, dans le cas des routines impersonnelles, la présence de l'auteur est cachée, mais l'intention de signaler une information essentielle au lecteur est mise en scène et focalisée. Cela permet aussi de rendre l'écrit scientifique plus objectif.

Nous rappelons ici que les routines construites avec le pronom auteur sont moins présentes dans les routines de structuration et les routines métalinguistiques, mais encore moins présentes dans les routines de topicalisation. Bien que le nombre des routines personnelles soit plus faible (2 routines parmi 6 au total), leur nombre d'occurrences apparaît plus élevé et occupe une proportion de 53,58 %. La forte préférence des routines avec le pronom auteur s'explique par l'abondance des routines de (co)constat. Formuler un constat renvoie plutôt à un acte de l'auteur qui joue un rôle de chercheur.

8.3.1.4 Synthèse

Pour récapituler, nous reprenons toutes les données des routines de guidage dans le tableau 8-8, et proposons une observation plus complète des routines de guidage portant la trace de l'auteur (l'emploi du pronom personnel renvoyant à l'auteur).

Catégories principales	Sous-fonctions	Occurrences	Taux des routines avec le pronom auteur
Structuration	Planification de la structure	115	35,65 %
	Annonce de l'objectif	360	41,94 %
	Navigation textuelle	1399	40,46 %
	Conclusion	55	81,82 %
	TOTAL	1929	41,63 %
Métalinguistiques	Reformulation	124	19,35 %
	Réparation	66	22,73 %
	Précision terminologique	337	65,88 %
	Emprunt terminologique	95	15,79 %
	TOTAL	622	44,37 %
Topicalisation	Mise en évidence	369	24,12 %
	(co)Constat	330	100 %
	Focus	83	0%
	TOTAL	782	53,58 %
TOTAL		3333	44,94 %

Tableau 8-8 : Distribution des fonctions des routines de guidage

Nous remarquons dans chaque catégorie principale, il existe une sous-fonction qui mobilise plus de constructions avec les pronoms personnels : la conclusion (81,82 %), la précision terminologique (65,88 %), le (co)constat (100 %). Il semble que l'auteur des écrits scientifiques manifeste plus de subjectivité quand il formule des conclusions ou des constats, et élabore sa terminologie du domaine. La conclusion et le constat renvoient principalement aux résultats d'une démarche argumentative de la recherche. C'est le moment où l'auteur doit mettre en valeur son rôle de chercheur et mettre en lumière les résultats obtenus après son observation. La précision terminologique regroupe plus de routines avec la présence de l'auteur exprimée par un pronom personnel. L'auteur s'engage plus dans les actes définitoires comme la dénomination, la définition.

Les occurrences des pronoms auteur sont néanmoins un peu moins présentes que les routines sans pronoms auteur, ce qui correspond à une caractéristique des écrits scientifiques marquée par l'effacement énonciatif. Les écrits scientifiques possèdent nombreux « tours objectivants » (Grossmann, 2010, p. 419) : l'emploi des passifs (*X est défini comme... X est décrit dans le tableau...*), l'emploi des sujets métonymiques comme *ce travail, cette étude* (Tutin, 2010, p. 36). Dans notre corpus, nous observons cela dans de nombreuses routines comme *l'objectif de cet article est de..., cet article propose une étude...* En résumé, on constate que les routines de guidage sont plus fréquentes dans la fonction métatextuelle et dans la fonction de topicalisation.

8.3.2. Routines de guidage à travers les disciplines

Bien que ces disciplines appartiennent au même champ des sciences humaines et sociales, elles présentent également des différences non négligeables (Tutin, 2010, p. 177). Dans cette section, nous mènerons quelques observations sur la comparaison des disciplines à travers les routines de guidage. L'objectif est d'esquisser les différences et les similitudes les plus saillantes entre les 10 disciplines à partir des résultats de l'analyse. Étant donné qu'il s'agit de caractériser les disciplines à travers l'utilisation d'un certain nombre de routines, qui constituent autant de facteurs de rapprochement ou de distinction, nous proposons la méthode d'exploration fondée sur les analyses factorielles des correspondances (AFC). Les fréquences absolues sont utilisées pour les analyses factorielles des correspondances (AFC) et les fréquences relatives des routines dans les disciplinaires sont utilisées pour vérifier les résultats obtenus.

8.3.2.1 Hypothèses sur les disciplines

Avant d'entrer dans les résultats d'analyses, nous mentionnons les hypothèses à partir de certaines études qui permettent de comprendre quelques grandes caractéristiques des disciplines.

Plusieurs études s'intéressent à la comparaison entre les disciplines. Tutin (2010, p. 214–215) a comparé les pronoms auteurs et les verbes de positionnement (*considérer, penser, estimer, contester, opposer*) dans trois disciplines issues du corpus Scientext : la psychologie expérimentale, les sciences de l'éducation et la linguistique. Les résultats montrent que la psychologie, une discipline dont l'étude se fonde largement sur l'expérimentation, préfère mettre en évidence les hypothèses et les résultats (*nous concluons*), l'auteur des articles en sciences de l'éducation exprime davantage des intentions (*nous voulons*) et des opinions (*nous pensons*). La linguistique utilise plus de verbes tant pour évoquer la démonstration et les résultats, que pour marquer les intentions et les opinions. Une autre étude sur les mêmes disciplines en examinant les verbes de constat (*constater, noter, remarquer*) montre que la linguistique utilise plus de verbes de constat, suivie par les sciences de l'éducation, et que la psychologie utilise moins de verbes constatifs.

Fløttum *et al.* (2006, p. 57) ont aussi observé les différences entre trois disciplines : l'économie, la linguistique et la médecine. Nous ne nous intéressons ici qu'aux différences entre l'économie et la linguistique, qui relèvent des disciplines des sciences humaines et sociales. Les résultats montrent que l'économie mobilise plus de marqueurs métatextuels et de pronoms de première personne, alors que les pronoms personnels indéfinis, la négation, les connecteurs d'opposition (*but*) et références bibliographiques sont plus productifs dans la linguistique. Grossmann *et al.* (2009, p. 191), dans une autre étude portant sur la filiation scientifique en économie et en linguistique, montrent que l'économie emprunte plutôt des objets scientifiques construits complexes comme *méthode, modèle, théorie, approche*. En revanche, la linguistique emprunte des objets scientifiques construits simples, tels que *concept, notion, idée, définition* que nous pouvons aussi trouver dans les routines de précision et d'emprunt terminologique.

Selon ces études, on peut imaginer d'abord que l'économie est spécifique. Cela correspond aux caractéristiques de cette discipline : certaines approches quantitatives sont plus proches des sciences « dures », et elle a donc besoin d'une organisation discursive et d'une explication plus explicites que les autres disciplines humaines, comme on l'a vu dans la section 2.2.1. On pourrait ainsi supposer que l'économie emploie plus de routines de

structuration textuelle que la linguistique, peut être aussi plus que les autres disciplines. En revanche, les routines d'emprunt terminologique et de précision terminologique pourraient être moins fréquentes en économie qu'en linguistique.

Ensuite, nous supposons que la psychologie se distingue parce qu'elle utiliserait plus de routines de conclusion pour présenter les résultats. Les routines de conclusion et les routines liées à la terminologie (précision et emprunt terminologique) seraient aussi très nombreuses dans la linguistique.

L'histoire est, quant à elle, une science qui s'appuie largement sur les documents, et donc utilise plus de citations (Bondi, 2007, p. 84). Cela permet de supposer que l'histoire emploie plus de routines de navigation textuelle (notamment, les renvois intertextuels) et plus de routines d'emprunt terminologique que les autres disciplines.

La géographie est la science de l'espace. Elle se caractérise par la description des régularités territoriales en recourant aux tableaux, aux figures, autrement dits, à des éléments péritextuels (Ounoughi & Duvillard, 2016). À l'instar de l'histoire, elle se réfère aussi aux sources externes (ibid.). Ces caractéristiques nous amènent à présumer que les routines de navigation textuelle et d'emprunt terminologique seraient également très fréquentes dans cette discipline.

Auría (2008, p. 143) indique que les sciences de l'information emploient plus d'adjectifs évaluatifs (*important, essential, interesting, etc.*) dans l'introduction que la linguistique appliquée. Ces adjectifs évaluatifs forment des expressions comme *it is thus critical to develop, a problem of central interest, the major focus of this study is to examine* (ibid. p. 135–136). On peut ainsi s'attendre à ce que les routines de topicalisation soient productives dans cette discipline.

La sociologie, les sciences politiques et l'anthropologie concernent toutes la société, les groupes sociaux et les cultures, et on peut supposer qu'elles sont proches l'une de l'autre.

Pour tester nos hypothèses, nous avons élaboré un tableau récapitulatif en indiquant les fréquences relatives des routines dans les disciplines. En parallèle, nous avons effectué les analyses factorielles des correspondances (AFC) en utilisant les fréquences absolues des routines dans les disciplines afin de montrer la distance entre les disciplines sur les principaux axes, ainsi que les associations entre disciplines et routines.

8.3.2.2 Distances entre les disciplines

Nous observons d'abord les distances entre les disciplines et relevons les particularités essentielles des dimensions rhétoriques dans les disciplines. Deux figures AFC et un tableau sont proposés dans cet examen. La figure 8-9 propose une vue globale des disciplines s'appuyant sur les fréquences absolues des sous-fonctions des routines. Le premier examen par AFC montre que l'économie occupe une position marginale, tellement distinctive que le premier axe se résume à la distinction économie vs les autres disciplines, qui s'explique par une surreprésentation des routines de planification et, dans une moindre mesure, de conclusion. Pour mieux comprendre les distances entre les autres 9 disciplines (figure 8-10), nous sommes obligée d'exclure l'économie dans les observations suivantes à l'aide des AFC. Ainsi, la figure 8-10 illustre de façon plus précise la distance entre les 9 disciplines régies par les sous-fonctions. Pour entrer davantage dans les détails, nous proposons aussi la figure 8-11 en 3D de la figure 8-10 en 2 D. Le tableau 8-9 recense les fréquences relatives pour 100 000 mots (fréquence relative=fréquence absolue/taille de sous-corpus* 100 000) des routines par sous-fonctions et par disciplines.

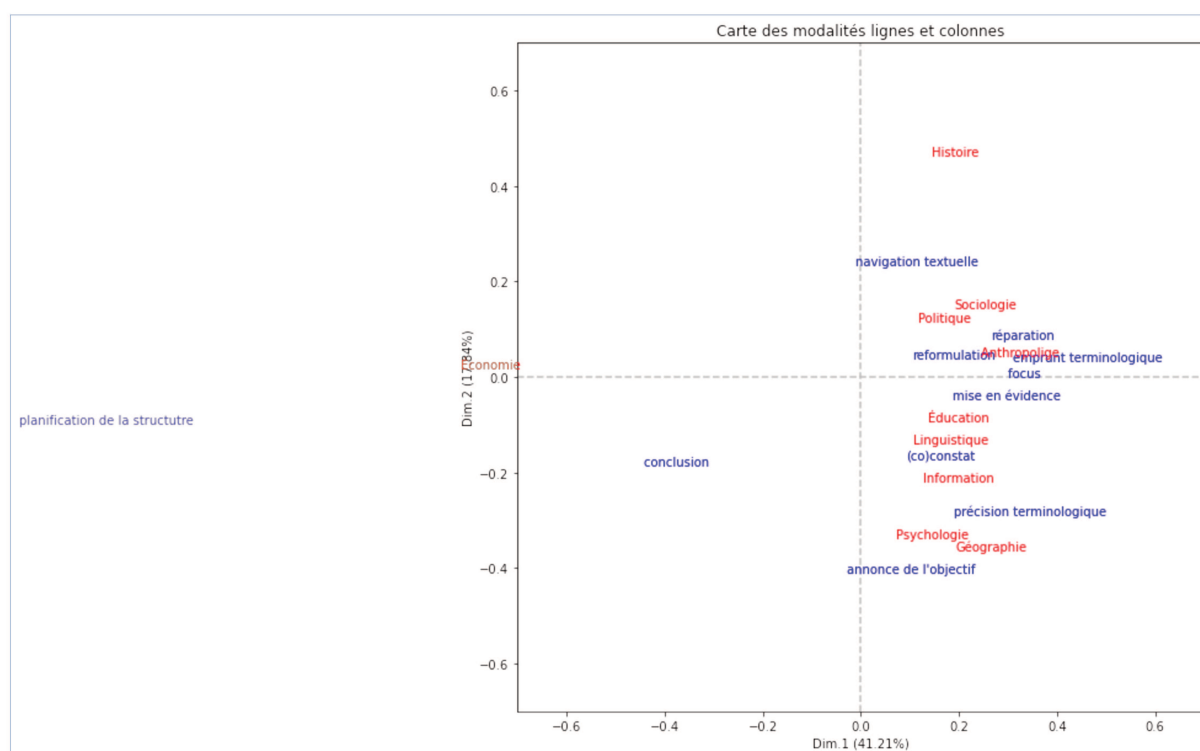


Figure 8-9 : Vue globale sur les disciplines en 2D

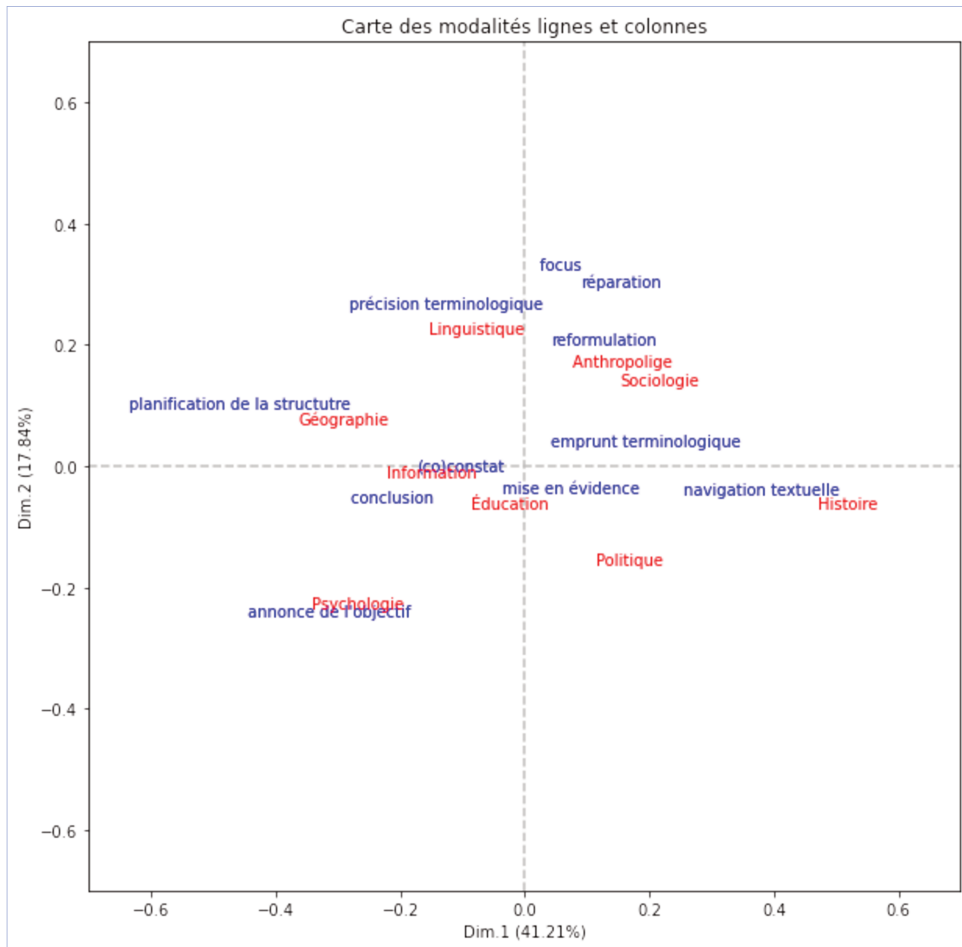


Figure 8-10 : Vue globale sur les disciplines en 2D sans économie

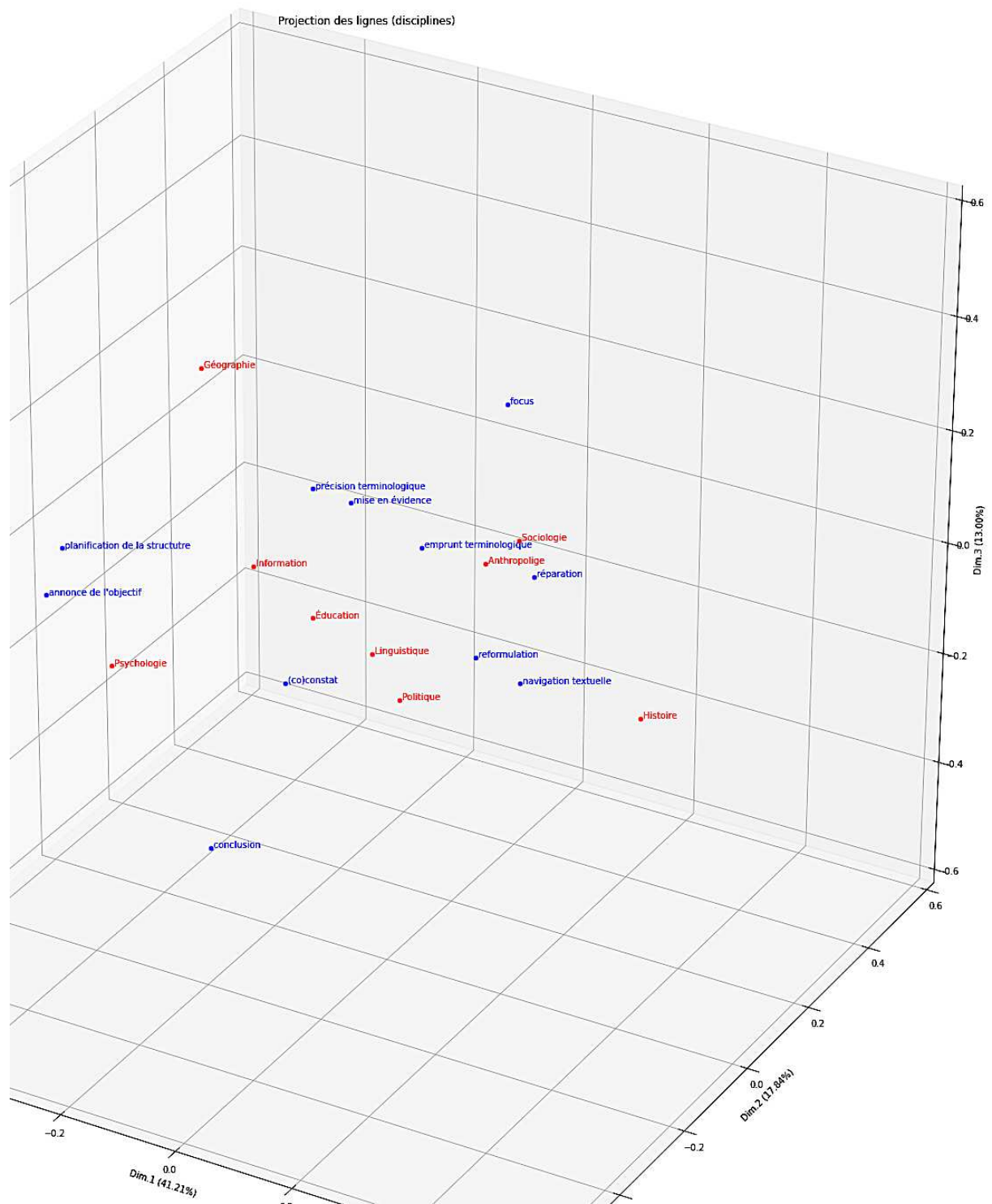


Figure 8-11 : Distance entre les disciplines sans économie, en retenant les 3 principaux axes factoriels, expliquant environ 72 % de l'inertie

Fréquences relatives (=fréquence absolue/taille de sous-corpus* 100 000) pour 100 000 mots

Catégories principales		Sous-fonctions	Éco	Lin	Psy	Soc	Poli	Édu	Info	Géo	Anthro	His	Moyenne
Structuration	planification de la structure		17,8	1,6	1,0	0,4	0,5	0,7	1,0	1,0	-	-	2,4
	annonce de l'objectif		10,9	7,0	14,4	3,7	7,3	12,5	10,8	7,2	4,0	1,6	7,94
	navigation textuelle		44,3	28,2	21,1	27,9	32,8	30,9	22,6	10,7	25,1	33,0	27,66
	conclusion		3,2	2,3	2,2	1,1	0,9	0,5	1,0	-	-	0,4	1,16
	<i>Total</i>		76,2	39,1	38,7	34,2	41,5	44,6	35,4	18,9	29,1	35,0	39,16
Métalinguistiques	reformulation		2,8	4,9	1,9	3,0	2,2	3,4	2,5	0,7	2,4	1,8	2,37
	réparation		0,8	2,8	0,5	1,3	0,9	2,2	0,5	1,2	1,6	1,3	1,31
	précision terminologique		5,9	13,9	6,9	6,5	4,0	7,7	10,0	8,2	8,1	2,6	6,69
	emprunt terminologique		0,8	2,3	1,0	2,0	3,1	1,4	3	2,0	1,8	1,7	1,91
	<i>Total</i>		10,3	23,9	10,3	12,8	10,2	14,7	16	12,1	13,9	7,4	12,28
Topicalisation	mise en évidence		6,5	6,3	9,3	9,1	8,2	8,2	6,8	9,5	6,5	5,9	7,63
	(co)constat		7,5	11,7	10,1	4,1	7,8	8,2	5,8	4,2	7,3	3,4	7,01
	focus		1,0	1,9	1,2	3,7	0,9	2,4	1,5	1,2	2,6	0,8	1,72
	<i>Total</i>		15	19,9	20,6	16,9	16,9	19	14,1	14,9	16,4	10,2	16,36
TOTAL		101,6	83,1	69,6	62,7	68,8	78	65,4	46,2	59,5	52,6	67,8	

Tableau 8-9 : Fréquences relatives pour 100 000 mots des routines dans les disciplines

À l'aide des figures 8-9, 8-10 et 8-11, on peut relever rapidement que l'économie est très éloignée des autres disciplines, ce qui montre qu'elle est bien spécifique. Cette spécificité est marquée, comme nous l'avons vu, par le suremploi des routines de planification (17,8) et de conclusion (3,2), comme le confirme le tableau 8-9. Selon ce tableau, on peut aussi relever les tendances plus générales :

- 1 Trois disciplines (l'économie, la linguistique et les sciences de l'information) emploient plus de routines que les autres ;
- 2 Trois disciplines (la géographie, l'histoire, et l'anthropologie) ont tendance à sous-employer les routines de guidage.

Comme l'économie est très spécifique, on se focalise d'abord sur l'emploi des routines en économie pour tester les hypothèses faites dans la section précédente sur cette discipline. Tout d'abord, selon le tableau 8-9, on observe très clairement l'abondance des routines de guidage dans l'économie (101,6) par rapport aux autres disciplines, ce qui confirme la

singularité de cette discipline (voir figure 8-9). Au sein des routines de guidage identifiées en économie, la dimension de la structuration est très prédominante, ce qui correspond à notre hypothèse sur l'abondance des routines de structuration dans cette discipline. Une autre hypothèse concernant le sous-emploi des routines à fonction de précision et d'emprunt terminologiques en économie, par rapport à la linguistique, est aussi validée par le tableau 8-9 : l'économie utilise respectivement 5,9 et 0,8 occurrences pour 100 000 mots (inférieurs aux fréquences moyennes 6,69 et 1,91), tandis que la linguistique en emploie 13,9 et 2,3 occurrences, ce qui est supérieur à la moyenne. Cela peut aussi vouloir dire que la terminologie en économie est plus normalisée, et qu'on n'a donc pas besoin de la définir. Contrairement à l'économie, la linguistique utilise plus de routines métalinguistiques et de topicalisation que les autres disciplines, en particulier les routines de précision terminologique et de (co)constat. Cela correspond à l'hypothèse de Grossmann *et al.* (2009, p. 191) : les linguistes ne préfèrent pas reprendre l'idée, modèle ou théorie des autres (l'emprunt terminologique), mais ont tendance à présenter leur propre point de vue, dans notre cas, à travers la précision terminologique et le (co)constat.

Le fait que l'économie est très distincte des autres disciplines gêne la comparaison entre autres disciplines avec l'AFC. Sur la figure 8-10, on voit bien que les autres disciplines se différencient essentiellement selon l'axe vertical, qui ne représente que 17 % de l'inertie et ne permet pas une interprétation à lui seul. C'est pourquoi nous avons proposé un affichage des 3 axes principaux sans l'économie à l'aide de l'AFC basée sur les fréquences absolues des sous-fonctions (voir annexe 3). Les résultats qui présentent les distances entre les disciplines à travers l'examen des dimensions rhétoriques sont illustrés dans la figure 8-11.

Les 9 disciplines se trouvent principalement dans deux groupes : la psychologie et la géographie semblent plus proches, tandis que les disciplines restantes (les sciences de l'information, les sciences de l'éducation, la linguistique, les sciences politiques, la sociologie, l'anthropologie et l'histoire) se trouvent sur un autre axe. La psychologie et la géographie se distinguent des autres sur les 2 premiers axes, et la géographie se distingue également selon le 3e. Il semble que la dimension de l'annonce de l'objectif soit plus présente en psychologie que les autres dimensions rhétoriques. En outre, la psychologie emploie plus de routines de topicalisation, notamment le (co)constat (20,6/100 000 dans le tableau 8-10). Ces remarques rejoignent le constat fait par Tutin (2010) : la psychologie a tendance à utiliser plus de verbes pour la démonstration des résultats et des hypothèses. Nous considérons que l'importance des résultats et des hypothèses peut expliquer, dans une certaine mesure, les routines de la fonction

d'annonce de l'objectif. Les résultats servent à répondre à ces questions, et les hypothèses permettent de formuler des propositions sur une question de recherche. Les deux sont étroitement liés à l'objectif de recherche. En combinant l'étude de Tutin (2010) et ces résultats, on peut donc supposer que les psychologues ont tendance à présenter plus explicitement les objectifs de l'étude, les hypothèses et les résultats.

La sociologie et l'anthropologie sont assez proches et partagent la fonction d'emprunt terminologique, ce qui correspond à nos attentes. Mais les sciences politiques, également présumées proches de ces deux disciplines, semblent plutôt éloignées. Par ailleurs, on remarque une distance importante entre l'histoire et la géographie dans la figure 8-11. On supposait ces deux disciplines proches en matière de l'utilisation des routines de navigation textuelle et d'emprunt terminologique. Quand on vérifie les fréquences relatives dans le tableau 8-6 pour ces deux disciplines, il apparaît en effet que l'histoire et la géographie possèdent plus de routines de navigation textuelle (10,7 et 33,0) que les autres routines, ce qui correspond à notre hypothèse. Bien que l'utilisation des routines de navigation entre ces deux disciplines soit proche, cela ne permet pas d'expliquer pourquoi elles sont éloignées globalement. Nous proposons donc une autre analyse factorielle des correspondances plus précises faisant apparaître la corrélation entre les routines précises et les disciplines. Les résultats sont présentés dans la figure 8-12. La discussion sera présentée dans la section suivante en recourant aux fréquences relatives des routines dans les 9 disciplines (sauf l'économie).

8.3.2.3 *Corrélation entre les disciplines et les routines*

Nous nous intéressons maintenant aux corrélations entre les disciplines et les routines. Ces relations corrélatives permettent de mieux comprendre les différences et les similitudes entre les disciplines, en particulier sur l'emploi d'une routine ou d'une partie de routines qui marque une discipline vis-à-vis des autres. Deux types de données sont utilisées ici : une AFC des routines à travers les 9 disciplines (fréquences absolues) (figure 8-12) et un tableau des fréquences relatives par million de mots¹¹³ des routines dans les disciplines (tableau 8-10).

Nous avons déjà souligné que la psychologie est une discipline marquée par la dimension d'annonce de l'objectif. La figure 8-12 précise que cette dimension se traduit spécifiquement par la surreprésentation d'une routine d'annonce de l'objectif *l'objectif de cet*

¹¹³ Pour mettre en évidence la différence entre les fréquences relatives des 37 routines de guidage dans les 10 disciplines, nous avons élargi la proportion de 100,000 à 1,000,000.

article est d'analyser... avec une fréquence relative de 86 occurrences par million de mots, alors que la deuxième place est occupée par l'économie qui a 32 occurrences par million de mots.

En ce qui concerne l'histoire, elle se caractérise par le suremploi de la routine de navigation intertextuelle *voir l'article de quelqu'un* (233/million), qui occupe la première place dans toutes les disciplines. Cette routine est aussi la deuxième plus fréquente pour la géographie (42/million), après la routine de mise en évidence *il faut noter que* (57/million). Par ailleurs, la navigation intratextuelle, notamment avec la routine *comme nous allons le voir* (42/million) est aussi très présente en géographie. Ces résultats confirment bien les caractéristiques de ces disciplines, qui s'appuient largement sur des documents et des éléments péritextuels. Malgré cette similitude, l'histoire n'est nullement proche de la géographie. La distance entre ces deux disciplines est causée principalement par l'emploi des autres routines. L'histoire montre une grande attirance pour la routine *voir l'article de quelqu'un* ((N° 7 des routines de navigation textuelle), ce qui correspond à nos attentes : la dimension interdiscursive est très marquée en histoire, où les bibliographies sont donc très importantes.

Fréquences relatives (=fréquence absolue/taille de sous-corpus* 100 000) pour 100 000 mots

Sous-fonctions	Routines	Éco	Lin	Psy	Socio	Poli	Édu	Info	Géo	Anthro	His
planification	<i>la section 4 décrit</i>	115	2	5	2	5	5	8	10	-	-
	<i>dans la section 2, nous présentons</i>	63	14	5	0	0	2	3	-	-	-
annonce de l'objectif	<i>L'objectif de cet article est d'analyser</i>	32	23	86	6	9	36	23	25	8	3
	<i>L'article s'attache à montrer</i>	6	0	10	13	15	26	23	20	10	5
	<i>cet article propose une étude</i>	4	0	5	2	0	0	0	2	6	1
	<i>L'article s'intéresse à</i>	4	0	2	6	18	12	10	7	4	4
	<i>nous proposons dans cet article d'analyser</i>	26	26	19	4	13	12	30	7	8	1
	<i>nous nous intéressons à</i>	24	16	22	2	15	29	13	10	2	1
navigation textuelle	<i>les résultats sont exposés dans</i>	81	12	41	4	7	17	5	0	2	0
	<i>comme nous allons le voir</i>	105	124	57	68	115	110	93	42	103	44
	<i>comme l'indique le tableau</i>	0	2	12	0	0	7	0	5	0	1
	<i>je reviendrai sur ce point</i>	26	52	5	13	11	19	10	2	12	19
	<i>il convient de rappeler</i>	40	5	5	0	15	0	3	2	4	4
	<i>voir en annexe</i>	103	19	41	57	20	79	48	5	20	16
	<i>voir l'article de quelqu'un</i>	95	52	45	129	131	72	60	42	95	233
conclusion	<i>on peut conclure que</i>	20	23	12	11	7	5	10	0	2	4
	<i>le résultat principal est de</i>	12	0	10	0	0	0	0	0	0	0
reformulation	<i>on peut également dire</i>	2	12	2	4	2	10	3	0	8	6
	<i>cela veut dire</i>	26	38	17	22	13	24	18	7	10	12
réparation	<i>pour le dire autrement</i>	0	0	0	4	7	0	5	0	6	0
	<i>on ne peut pas dire</i>	2	7	2	2	0	7	5	0	2	4
précision terminologique	<i>cela ne veut pas dire</i>	6	21	2	11	9	14	5	12	14	9
	<i>X peut être défini comme</i>	6	9	24	4	5	12	15	25	8	3
	<i>on définit X comme</i>	14	14	7	2	0	10	8	2	2	0
	<i>on appelle X Y</i>	10	77	29	35	22	38	43	20	36	13
	<i>X est désigné par le mot Y</i>	4	12	5	11	0	2	5	2	2	3
	<i>J'utilise le terme X pour désigner</i>	22	7	2	7	9	7	15	17	10	4
emprunt terminologique	<i>le terme X renvoie à</i>	2	19	2	6	4	7	15	15	22	4
	<i>on reprend le terme de quelqu'un</i>	6	7	2	2	5	0	8	0	2	3
mise en évidence	<i>pour reprendre le terme de quelqu'un</i>	2	16	7	18	26	14	25	20	16	14
	<i>nous pouvons mentionner</i>	6	23	19	24	18	24	20	25	14	12
	<i>il est important de souligner</i>	10	0	12	7	4	7	3	10	4	4
c(co)constat	<i>il faut noter que</i>	49	40	62	59	38	43	60	57	47	43
	<i>nous pouvons noter que</i>	75	117	101	41	78	82	58	42	73	34
focus	<i>le problème essentiel est</i>	10	16	7	20	5	7	8	10	6	5
	<i>il est question de</i>	0	2	5	17	4	17	8	2	20	3

Tableau 8-10 : Fréquences relatives par 1 000 000 mots des routines dans les disciplines

Par contre, la géographie, est caractérisée plutôt par la routine *l'article s'attache à montrer* (N° 2 des routines d'annonce de l'objectif d'étude) et de la routine *la section 4 décrit* (N° 1 des routines de planification de la structure), récapitulées dans le tableau 8-11.

Fréquences relatives par million de mots		
	<i>l'article s'attache à montrer</i>	<i>la section 4 décrit</i>
Histoire	20	0
Géographie	5	10

Tableau 8-11 : Fréquences relatives par million de mots de deux routines dans l'histoire et la géographie

La linguistique éprouve une grande attraction pour une routine de reformulation (N° 2 des routines de reformulation) *cela veut dire/cela signifie que...*, aussi pour une routine de dénomination (*Y est désigné par le mot X*). En effet, la linguistique mobilise beaucoup de routines métalinguistiques dans l'ensemble des routines de guidage (23,9 occurrences pour 100 000 mots, alors que la moyenne est de 12,28 pour 100 000, cf. tableau 8-9).

Si l'anthropologie et la sociologie montrent une préférence pour la dimension d'emprunt terminologique dans la figure 8-11, les routines fréquentes de cette dimension rhétorique dans les deux disciplines ne sont pourtant pas identiques : dans la figure 8-12, on voit que la sociologie est associée à la routine *on reprend le terme de quelqu'un*, tandis que l'anthropologie est plus marquée par la routine *pour reprendre le terme de quelqu'un*.

Pour les points situés près de l'origine, nous ne pouvons pas interpréter davantage les relations entre routines et disciplines, car la distance entre les points rouges et les points bleus n'est pas directement interprétable, étant donné que dans un tel *biplot*, ce sont les directions prises dans l'espace qui peuvent être comparées.

Pour conclure, les analyses quantitatives permettent de dégager des régularités assez remarquables sur les emplois des routines de guidage dans notre corpus, comme les routines de structuration qui sont les plus nombreuses, et le fait que l'économie se détache des autres disciplines. Néanmoins, il faut admettre que nos observations sur les disciplines ne prétendent pas être très précises, notre corpus étant sans doute trop modeste pour pouvoir tirer des conclusions définitives et de portée générale. Nous pouvons dessiner de grandes tendances, qui mériteraient d'être précisées par des enquêtes plus approfondies dans chaque discipline.

8.3.3. Synthèse

Dans cette section, nous avons testé nos hypothèses sur la répartition des routines de guidage, ainsi que les particularités des disciplines à l'aide des observations quantitatives. Nous

constatons que la plupart des hypothèses établies au préalable ont été validées, comme la présence importante des routines de structuration dans l'ensemble du corpus par rapport aux autres routines. La dimension de la structuration textuelle est fondamentale pour la garantie de la lisibilité du texte scientifique. Par contre, les routines de topicalisation sont plus nombreuses que les routines métalinguistiques, ce qui contredit notre hypothèse initiale. Cela est peut-être lié au fait que les marqueurs sur lesquels se basent nos hypothèses sont moins adaptés à cette fonction. Malgré le fait que les routines métalinguistiques soient les moins fréquentes, elles se réalisent à travers de nombreuses routines de précision terminologiques. Ces routines sont essentielles pour formuler des définitions, attribuer un nom à un nouveau concept et expliquer le lien entre signe linguistique et chose concernée.

Concernant les disciplines, l'économie se détache par le suremploi des routines de structuration et le sous-emploi des autres dimensions. La psychologie est également assez spécifique, marquée par l'emploi des routines de topicalisation et d'annonce de l'objectif, en particulier la routine *l'objectif de cet article est d'analyser...* L'histoire et la géographie ont besoin de faire des liens avec les sources documentaires, se traduisant par l'emploi des routines de navigation textuelle. En résumé, on voit que l'équilibre des dimensions rhétoriques, ainsi que certaines routines précises, peuvent mettre en lumière des caractéristiques disciplinaires intéressantes. Ce que nous avons observé dans cette section ne présente que les tendances globales des disciplines. Nous pensons que l'entrée par les routines est pertinente pour mieux comprendre les modes argumentatifs et les critères de scientificité des disciplines. Mais une étude plus approfondie devra être menée ultérieurement, afin de dégager plus précisément ces spécificités.

8.4. Synthèse

Dans ce chapitre, nous nous sommes intéressée à trois aspects des routines de guidage du lecteur : les positions textuelles, les modes d'actualisation énonciative, et enfin les répartitions des routines et les similitudes et différences entre les disciplines.

L'emplacement textuel des routines respecte des pratiques rhétoriques très courantes de du discours scientifique, en particulier les articles de recherche ayant pour objectif la publication. Les routines de planification et de l'annonce de l'objectif/objet sont utilisées davantage dans le début du texte ou dans le début d'une section textuelle. Les routines de navigation peuvent se trouver dans les incises, les notes ou entre parenthèses. Les routines de précision et d'emprunt terminologique intègrent régulièrement des éléments typographiques spécifiques comme des guillemets. En réalité, conformément à nos attentes, la plupart des

routines ne sont pas fixées à une section spécifique du texte, et peuvent être utilisées selon le besoin réel de l'énonciation.

Sur le plan énonciatif, nous avons vu que les routines de guidage manifestaient un ensemble de moyens linguistiques pour s'actualiser explicitement. Ces marques énonciatives explicites permettent d'identifier les dimensions dialogiques en jeu. La plupart des routines appartiennent à la dimension interlocutive, mais elles peuvent aussi s'inscrire dans le dialogisme interdiscursif (les routines qui renvoient aux discours d'autrui) ou dans le dialogisme intralocutif (comme les routines de reformulation).

En ce qui concerne les analyses quantitatives, on peut relever que les routines de structuration sont les plus nombreuses suivies par les routines de topicalisation, puis les routines métalinguistiques. Pour la fonction de structuration, notre étude confirme les observations effectuées en anglais sur les marqueurs plus courts (Hyland, 2005, Ädel, 2006). En revanche, les routines métalinguistiques apparaissent moins fréquentes que les routines de topicalisation, contrairement aux travaux de Hyland et Ädel. Cela est probablement lié à l'abondance des marqueurs courts à fonction métalinguistique. Par contre, les routines de topicalisation ont un rôle très spécifique et n'ont pas de véritable équivalent sous forme de marqueurs courts.

L'observation quantitative sur les occurrences des routines avec les pronoms auteurs montre que les routines sans pronoms personnels sont les plus utilisées dans le corpus, ce qui confirme la tendance à l'effacement énonciatif souvent observé dans le discours scientifique. Quant aux routines de guidage construites avec les pronoms personnels, on constate, en suivant la typologie des rôles d'auteur de Fløttum *et al.* (2006), que l'auteur scientifique assure en premier lieu un rôle d'énonciateur et moins fréquemment, un rôle de chercheur, mais pas un rôle d'argumentateur ou d'évaluateur.

Par ailleurs, des modes d'écriture disciplinaires peuvent être mis en évidence par les routines de guidage. Nos résultats rejoignent quelques observations déjà effectuées (Fløttum *et al.*, 2006 ; Tutin, 2010, Grossmann *et al.* 2009). Nous avons d'abord remarqué une tendance principale en ce qui concerne l'emploi des routines : 1) l'économie, la linguistique et les sciences de l'information suremployent les routines ; 2) la géographie, l'histoire, et l'anthropologie ont tendance à sous-employer ces éléments.

Parmi les dix disciplines examinées, certaines disciplines sont spécifiques :

- 1) L'économie se trouve éloignée des neuf autres disciplines. Elle se caractérise par la grande productivité des routines de structuration textuelle et l'utilisation plus limitée

des routines de topicalisation et des routines métalinguistiques. Elle paraît plus proche des sciences « dures » qui ont une architecture textuelle très structurée. Les termes et les définitions semblent plus normalisés en économie, car elle utilise très peu de routines de précision terminologique.

- 2) La psychologie possède plus de routines d'annonces de l'objectif (*l'objectif de cet article est d'analyser*) pour parler de la problématique, et les routines de (co)constat (*on peut constater que*) pour présenter les résultats.
- 3) La linguistique semble utiliser davantage de routines métalinguistiques et de topicalisation que les autres disciplines, notamment les routines servant à préciser une terminologie et à formuler un constat. Formuler son propre point de vue sur la terminologie et sur l'observation des résultats semble très valorisé dans cette discipline.
- 4) L'histoire est remarquable pour l'utilisation abondante des routines de navigation intertextuelle (*voir l'article de quelqu'un*) ce qui affirme la place primordiale de la dimension interdiscursive dans cette discipline, basée largement sur l'étude des documents.

En résumé, les routines présentent bien un intérêt pour étudier les modes rhétoriques des disciplines.

Conclusion de la partie III :

Au terme de cette partie, nous avons identifié 12 fonctions rhétoriques réparties en trois plans sur lesquels se déploient les routines de guidage du lecteur. Sur le plan structurel, les fonctions inventoriées montrent l'essence de l'organisation du discours scientifique : il nécessite un plan clair, une annonce de l'objectif/objet précis, un guidage pour la navigation entre les différentes parties textuelles. Sur le plan métalinguistique, une attention particulière doit être portée à la production des énoncés pour garantir la clarté des termes et des expressions utilisées et la bonne interprétation du message. Sur le plan de la topicalisation, il est à souligner que certaines idées dans le discours méritent une attention particulière par rapport aux autres. Cela permet de hiérarchiser les informations pour le lecteur.

Nous constatons, par le biais des observations sur les caractéristiques linguistiques sur les plans syntaxiques, sémantiques et fonctionnels, que les expressions récurrentes nommées « routines de guidage du lecteur » partagent des champs sémantiques et lexicaux dans chaque sous-catégorie. C'est aussi grâce à ces régularités que nous les avons classées dans une même catégorie fonctionnelle. Outre les régularités sémantiques et lexicales, il faut remarquer qu'une même routine est aussi susceptible d'avoir plusieurs variations syntaxiques. En parallèle, les routines intègrent couramment avec les marqueurs discursifs, créant ainsi un lien plus fort dans la dynamique discursive.

Nous avons principalement intégré les routines de guidage dans la dimension interlocutive du dialogisme, étant donné que la visée de ces routines concerne avant tout le lecteur du discours scientifique. Cependant, les routines de guidage s'investissent de façon complexe dans les dimensions dialogiques. Les trois dimensions, à savoir, interdiscursive, interlocutive et intralocutive, sont toutes concernées par ces routines. Nous pouvons rappeler par exemple que les routines de navigation intertextuelle présentent à la fois les valeurs interdiscursive et interlocutive, et les routines d'emprunt terminologique appartiennent aux dimensions intralocutive et interdiscursive.

Les observations des positions textuelles et des répartitions des routines ainsi que leurs situations d'emploi à travers les disciplines, permettent d'une part de comprendre les contraintes rhétoriques de certaines parties textuelles des articles de recherche ; d'autre part, de compléter des études déjà effectuées sur la comparaison des disciplines. Notre étude montre par ailleurs que l'étude des pratiques d'écritures des disciplines à partir des routines semble être une piste de recherche intéressante.

Conclusion générale

Dans cette étude, nous nous sommes focalisée sur le classement fonctionnel et la modélisation des routines de guidage du lecteur dans les écrits scientifiques en français. Notre travail est à la fois théorique et méthodologique. Sur le plan théorique, nous avons circonscrit les routines de guidage du lecteur en lien avec d'autres notions proches et rivales dans la phraséologie et proposé une modélisation aux routines, en synthétisant leurs principales caractéristiques. Sur le plan méthodologique, en recourant aux méthodes issues de la linguistique de corpus et du traitement automatique des langues, notre travail a permis de montrer que la méthode de requête à l'aide des classes de mots et des ALR est valable, ce qui nous a permis d'identifier et de classer de façon onomasiologique les routines accompagnant la lecture dans le corpus TermITH-transdisciplinaire qui comporte 500 articles de recherche en Sciences humaines et sociales avec 5 millions de mots.

En guise de conclusion, nous retracerons d'abord les résultats principaux de notre étude et puis signalerons les limites et ouvrirons enfin les perspectives pour des travaux futurs.

Résultats principaux

Les routines de guidage du lecteur constituent un objet d'étude inscrit dans l'extension du domaine de la phraséologie. À travers les observations sur les notions connexes et les routines de guidage identifiées, ainsi que l'association avec l'analyse du discours et la linguistique de corpus, nous avons formulé une définition plus claire de ce phénomène dans les écrits scientifiques. Plutôt que d'être des phrases ou des propositions isolées, les routines sont des groupes d'expressions récurrentes qui se construisent autour d'un prédicat, partageant des traits communs sur les plans lexical, syntaxique et sémantique. Elles s'inscrivent dans la situation de l'énonciation des écrits scientifiques de façon explicite ou implicite, et remplissent une série de fonctions métadiscursives qui permettent de mieux organiser le discours et mieux orienter l'interprétation du contenu scientifique. Cette définition des routines de guidage peut s'appliquer aux autres discours argumentatifs, par exemple le discours scientifique oral, ou le discours politique juridique.

Outre de la délimitation des routines de guidage, nous avons procédé à un classement des routines selon des critères lexico-syntaxiques, sémantique et fonctionnel en les organisant de façon onomasiologique, en nous inspirant des fonctions du métadiscours de Hyland (2005)

et Ädel (2006, 2010). Notre typologie comporte 37 routines réparties dans 11 sous-fonctions, regroupées en 3 grandes catégories fonctionnelles :

- La structuration textuelle comporte : 2 routines de planification de la structure du discours, 6 routines d'annonce de l'objectif d'étude, 8 routines de navigation textuelle et 2 routines de conclusion ;
- Les fonctions métalinguistiques comportent : 3 routines de reformulation, 2 routines de réparation, 6 routines de précision terminologique et 2 routines d'emprunt terminologique ;
- La topicalisation comporte : 3 routines de mise en évidence, 1 routine de (co) constat et 2 routines de focus.

En termes de fonctions rhétoriques, la structuration textuelle concerne l'organisation du contenu propositionnel dans le but de proposer une bonne architecture tout au long du texte permettant au lecteur de mieux suivre le discours. Les routines dans cette catégorie sont très abondantes dans notre corpus, représentant 57,88 %. Parmi celles-ci, les routines de navigation textuelles pour envoyer le lecteur dans un autre lieu textuel ou dans un autre discours d'autrui sont les plus fréquentes.

Les fonctions métalinguistiques servent à rendre le contenu propositionnel plus clair. En clarifiant les énoncés précédents par les procédés reformulatifs et réparatifs, l'auteur anticipe une compréhension difficile ou biaisée du lecteur de certains éléments du discours. La précision et l'emprunt terminologique consistent un aspect rhétorique fondamental du discours scientifique. L'emploi des termes disciplinaires peut renvoyer à plusieurs actes comme la définition, la dénomination, la désignation et l'emprunt d'un terme ou une définition aux pairs.

La topicalisation permet de hiérarchiser les informations pour le lecteur en relevant celles qui sont essentielles à retenir. Cette catégorie occupe 23,46 % des routines de guidage, suivie par les routines métalinguistiques (18,66 %). Les routines de mise en évidence et de (co) constat sont les plus fréquentes, ce qui permet de constater que les écrits scientifiques en SHS mettent l'accent sur l'observation des faits scientifiques.

Notre étude ne vise pas seulement le classement des routines de guidage du lecteur selon leurs fonctions rhétoriques. Un autre apport important de cette thèse est de proposer une modélisation linguistique qui permet de décrire systématiquement les routines en prenant en compte tous les aspects linguistiques des routines, tels que les caractéristiques lexicosyntaxiques, sémantiques, fonctionnelles, énonciatives et dialogiques, ainsi que les éléments

périphériques, positions textuelles. Dans cette modélisation, nous avons accordé comme Pecman (2004) des étiquettes conceptuelles aux classes sémantiques des paradigmes lexicaux des routines. Nous avons aussi formulé des descriptions sémantiques à chaque routine et proposé des formules syntaxiques avec des classes sémantiques du LST afin de mieux représenter les constructions lexico-syntaxiques des routines de guidage. La modélisation et l'observation des routines ont permis de relever des résultats intéressants.

Sur le plan lexico-syntaxique, les routines possèdent des constructions lexico-syntaxiques complexes, qui manifestent différents degrés de variabilité. Il existe des routines peu variables comme les routines de reformulation (*cela veut dire/signifie que..., pour le dire autrement*), mais aussi des routines plus variables tel que les routines d'annonce de l'objectif (*Le (°) [objectif/but] de ce (t) [article/papier/travail] [est de/consiste à] [analyser/examiner/décrire/présenter...]*).

Sur les plans sémantique et fonctionnel, les routines classées sous une fonction identique ont tendance à partager des paradigmes sémantiques similaires du Lexique Scientifique Transdisciplinaire. Les fonctions identifiées sont de nature métadiscursive et correspondent aux besoins rhétoriques des écrits scientifiques.

Sur les plans énonciatif et dialogique, nous avons observé les modes d'inscription des routines dans l'énonciation et les lieux dialogiques des écrits scientifiques à travers les routines de guidage. Les résultats montrent que :

- Les routines de guidage du lecteur présentent plusieurs modes d'inscription spécifique dans la situation d'énonciation : la majorité d'entre elles emploient des marques explicites comme les pronoms personnels de premières personnes et démonstratifs, les adverbes déictiques temporels ou spatiaux, les modalités et le temps des verbes ;
- Bien que les écrits scientifiques soient monologiques et aient une visée objectivante, ils manifestent une forte dimension dialogique interlocutive et interdiscursive à travers les routines. Ces routines visent, de prime abord, à rendre le texte plus clair pour le lecteur, et s'inscrivent donc largement dans la dimension interlocutive. Mais elles reflètent aussi la dimension interdiscursive qui est fondamentale dans les écrits scientifiques, en particulier à travers les routines de renvoi intertextuel et les routines d'emprunt terminologique.
- De façon attendue, l'auteur joue principalement le rôle d'énonciateur, mais aussi de façon secondaire celui de chercheur quand il emploie les routines de guidage, deux

rôles qui sont souvent étroitement mêlés. En revanche, les rôles d'argumentateur et d'évaluateur sont peu présents.

Dans les modélisations des routines, nous avons aussi observé rapidement les éléments périphériques qui apparaissent souvent à côté des routines. La plupart des éléments sont des marqueurs à fonction discursive comme la déduction, l'addition, le réexamen, etc. Ces éléments permettent d'inscrire les routines dans le processus argumentatif et discursif.

L'emplacement textuel des routines a aussi été indiqué dans les tableaux de modélisation des routines. Dans l'ensemble, il existe peu de routine qui sont liées à des parties textuelles spécifiques. Les routines de planification et d'annonce de l'objectif d'étude se trouvent souvent dans le début du texte (le résumé et l'introduction). Les routines de navigation textuelles peuvent se trouver en note ou entre parenthèses, permettant donc de hiérarchiser les informations tout en gardant la cohésion textuelle. Les routines métalinguistiques, même si elles ne se situent pas dans des lieux textuels spécifiques, ont recours à des éléments typographiques spécifiques (les guillemets, l'italique), ce qui reflète l'aspect conventionnel de la pratique typographique des écrits scientifiques.

L'idée de la modélisation, avec les étiquettes conceptuelles et les classes sémantiques dans la formule syntaxique, est de fournir une description plus abstraite en indiquant l'essentiel d'une routine. Nous pensons que cette modélisation peut être valable pour d'autres expressions complexes ayant des variants comme des routines.

À la fin de notre analyse, nous avons mené quelques observations des propriétés rhétoriques propres à chaque discipline, en nous basant sur les routines de guidage. Les résultats ont pu compléter les multiples études effectuées sur quelques disciplines (Fløttum *et al.*, 2006 ; Tutin, 2010, Grossmann *et al.* 2009) et permettre de mettre en lumière quelques pratiques d'écriture spécifique dans certaines disciplines. Par exemple, nos résultats montrent que l'économie est une discipline spécifique par rapport aux autres disciplines en sciences humaines et sociales. Proche des sciences « dures », elle emploie plus de routines de structuration pour organiser de façon explicite le contenu propositionnel. En revanche, les économistes ont tendance à utiliser moins de routines métalinguistiques, notamment pour les fonctions de précision et d'emprunt terminologique, peut-être parce que les termes sont plus stables dans cette discipline et n'ont pas besoin d'être précisés. Ces observations sur les différences et similitudes des modes rhétoriques des disciplines restent certes préliminaires, mais nous pensons que l'entrée par les routines pour étudier les disciplines pourrait inspirer les chercheurs travaillant dans le domaine de l'analyse du discours. Bien sûr, cet objectif peut aussi être réalisé

avec d'autres types de routines, comme celles qui renvoient à la filiation scientifique, à l'expression des attitudes ou aux sentiments.

Bien que notre étude fournisse des apports intéressants, il faut admettre qu'il existe encore des limites, mais aussi des perspectives.

Limites et perspectives

Nous sommes consciente des limites de la présente étude. Tout d'abord, nous rappelons la taille limitée de notre corpus et le genre peu varié. Le corpus ne comporte que 500 articles de recherche et n'inclut pas d'autres sous-genres des écrits scientifiques de la sphère universitaire : les mémoires, les thèses ou les HDR, etc. Par ailleurs, le Lexicoscope ne peut pas proposer une exploitation sur les sections textuelles des articles. Ces deux insuffisances ont donc abouti à deux limites principales : premièrement, nous n'avons identifié qu'un sous-ensemble de routines de guidage. Il existe des expressions bien intéressantes, mais qui ne répondent pas aux critères de fréquences ou de transdisciplinarité (au moins 5 fois répartie dans 3 disciplines au minimum). Nous pensons aussi que certaines routines peuvent avoir plus de variations lexicales si le genre discursif est étendu. Par exemple, les routines de planification de la structure dans notre corpus comportent souvent des noms de SECTION (*partie, section, introduction, conclusion*), mais pas des mots comme *chapitre, méthodologie, discussion*, qui se trouvent souvent dans les mémoires et les thèses. Deuxièmement, nous pensons qu'il serait bienvenu de mener une observation quantitative sur les positions textuelles des routines de guidage en indiquant très clairement leurs emplacements exacts. Cela permettrait de mieux comprendre la carte rhétorique des écrits scientifiques.

Une autre limite réside dans les analyses des routines. Nous avons bien indiqué la présence des marqueurs discursifs qui co-apparaissent avec ces routines de guidage. Quelques remarques ont été émises, mais de façon rapide. Ces éléments, bien que périphériques, devraient être davantage pris en considération dans le futur pour mieux comprendre la rhétorique et la discursivité des écrits scientifiques.

Cependant, notre étude peut ouvrir des perspectives. En linguistique, il sera intéressant et utile d'élargir les études des autres fonctions de routines, comme les routines d'attitude qui sont fondamentales dans les écrits scientifiques.

Nous pensons aussi aux applications en didactique pour le Français Langue Étrangère (FLE) et pour le Français sur Objectifs Universitaires (FOU). Cavalla (2015, p. 111), en

étudiant des erreurs de collocations dans les thèses rédigées en français par des doctorants non francophones natifs, a relevé que la phraséologie scientifique transdisciplinaire constitue bien l'une des difficultés rencontrées dans le cadre d'étude universitaire par ces étudiants non natifs de français. Ces éléments phraséologiques, selon Tutin (2013, p. 1), grâce à leur caractère préconstruit, peuvent favoriser non seulement le renforcement de l'appartenance des étudiants non natifs au discours scientifique, mais aussi faciliter leur écriture. Par conséquent, nous pouvons constater un vrai besoin d'élaborer des ressources de la phraséologie transdisciplinaire scientifique pour les apprenants de français. Les routines, comme d'autres ressources phraséologiques scientifiques, peuvent être introduites dans les cours de langues ou des ateliers de rédaction à l'université.

En outre, nous pouvons penser à des applications dans le domaine du Traitement Automatique des Langues. Le laboratoire Lidilem a déjà élaboré une base du LST qui recense des mots simples, des locutions et des collocations. Nous pouvons donc envisager un projet d'intégration des routines dans cette base afin de fournir aux apprenants de français un outil d'aide à la rédaction des écrits scientifiques plus performant. Ces moyens linguistiques permettraient aux étudiants non natifs de mieux maîtriser la rédaction des écrits universitaires.

Bibliographie

- Abeillé, A., & Godard, D. (2021). *La Grande Grammaire du français* (Actes Sud). Arles (France).
- Adam, J.-M. (2006). Intertextualité et interdiscours : Filiations et contextualisation de concepts hétérogènes. *Revue Tranel (Travaux neuchâtelois de linguistique)*, vol. 44, p. 3–26.
- Adam, J.-M. (2005). *La linguistique textuelle : Introduction à l'analyse textuelle des discours*. Paris : Armand Colin.
- Adam, J.-M. (2020). *La linguistique textuelle : Introduction à l'analyse textuelle des discours —4e éd.* Paris : Armand Colin.
- Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. John Benjamins Publishing.
- Ädel, A. (2010). Just to give you kind of a map of where we are going: A taxonomy of metadiscourse in spoken and written academic English. *Nordic Journal of English Studies*, 9(2), 69–97.
- Ädel, A., & Mauranen, A. (2010). Metadiscourse: Diverse and divided perspectives. *Nordic Journal of English Studies*, 9(2), 1–11.
- Aguilar, M. (2008). *Metadiscourse in Academic Speech: A Relevance-theoretic Approach*. Peter Lang.
- Aijmer, K. (2014). *Conversational Routines in English: Convention and Creativity*. Routledge.
- Amossy, R. (2005). De l'apport d'une distinction : Dialogisme vs polyphonie dans l'analyse argumentative. In *Dialogisme et polyphonie*. De Boeck Supérieur, p. 63–73).
- Amossy, R. (2015). *La présentation de soi : Éthos et identité verbale*. Presses Universitaires de France.
- Auría, C. P.-L. (2008). Stance and Academic Promotionalism: A Cross-Disciplinary Comparison in the Soft Sciences. *Atlantis: Revista de La Asociación Española de Estudios Anglo-Norteamericanos*, 30(1), p. 129-145.
- Authier-Revuz, J. (1995). *Ces mots qui ne vont pas de soi : Boucles réflexives et non-coïncidences du dire*. Paris : Larousse.
- Bakhtine, M. M. (1978). *Esthétique et théorie du roman* (D. Olivier Trad. & M. Aucouturier Préf., Trad.). Paris : Gallimard.

- Bakhtine, M. M. (1984). *Esthétique de la création verbale* (A. Aucouturier Trad. & T. Todorov Préf., Trans.). Paris : Gallimard.
- Bakhtine Mikhaïl Mikhaïlovitch. (1994). *The dialogic imagination : Four essays* (Holquist Michael & Emerson Caryl, Éd.s.). Austin: University of Texas press.
- Barbet, C. (2012). Devoir et pouvoir, des marqueurs modaux ou évidentiels ? *Langue française*, n° 173 (1), 49–63.
- Bandinelli, M. (2017). Segments phraséologiques et séquences textuelles. Questions méthodologiques et approches topologiques. *Corpus*, (17).
- Benveniste, É. (1970). L'appareil formel de l'énonciation. *Langages*, 5 (17), p. 12–18.
- Benveniste, É. (1974). *Problèmes de linguistique générale 2*. Paris : Gallimard.
- Berthoud, A.-C. (1982). *Activité métalinguistique et acquisition d'une langue seconde : Étude des verbes déictiques allemands*. Berne, Francfort-am-Main : Peter Lang.
- Bertrand, R., & Chanut, C. (2005). Fonctions pragmatiques et prosodie de enfin en français spontané. *Revue de Sémantique et Pragmatique*, (17), p. 41–68.
- Biber, D. (2006). *University Language: A Corpus-based Study of Spoken and Written Registers*. John Benjamins Publishing.
- Biber, D. (2009). A corpus-driven approach to formulaic language in english: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3), p. 275–311.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), p. 263–286.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, 25(3), p. 371–405.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics : Investigating Language Structure and Use*. Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English* (Pearson Education Limited). Harlow.
- Boch, F., & Grossmann, F. (2002). *Se référer au discours d'autrui : Quelques éléments de comparaison entre experts et néophytes*. *Enjeux*, 54, p. 41–51.

- Boch, F., Rinck, F., & Grossmann, F. (2009). *Le cadrage théorique dans l'article scientifique : Un lieu propice à la circulation des discours*. J.M. Lôpez Munoz, S. Marnette et L. Rosier. Actes du colloque international Ci-Dit La circulation des discours, Nota Bene, p. 23–42.
- Bonami, O., & Godard, D. (2008). Syntaxe des incises de citation. *Congrès Mondial de Linguistique Française*. Paris, France.
- Bondi, M. (2007). Authority and expert voices in the discourse of history. In K. Fløttum (Ed.), *Language and discipline perspectives on academic discourse*, p. 66–88. UK: Cambridge Scholars Publishing.
- Boutet, J., Gauthier, F., & Saint-Pierre, M. (1983). Savoir dire sur la phrase. *Archives de Psychologie*, (51), p. 205–228.
- Bres, J. (2001). Analyse du discours et dialogisme. *Diacrítica*, 16, p. 249–263.
- Bres, J. (2005). Savoir de quoi on parle : Dialogue, dialogal, dialogique ; dialogisme, polyphonie... In *Dialogisme et polyphonie*, p. 47–61. De Boeck Supérieur.
- Bres, J. (2017). Dialogisme, éléments pour l'analyse. *Recherches en didactique des langues et des cultures. Les cahiers de l'Acedle*, 14 (14–2).
- Bres, J., & Dufour, F. (2014). Du fonctionnement dialogique monologal des marqueurs dialogaux oui, non, si. *Arena Romanistica — Journal of Romance Studies*, (17), p. 58–69.
- Bres, J., & Nowakowska, A. (2008). "J'exagère ?..." Du dialogisme interlocutif. In M.-B. M. H. et C. N. M. Birkelund (Ed.), *L'énonciation dans tous ses états*, p. 1–27. Peter Lang.
- Bres, J., Nowakowska, A., & Sarale, J.-M. (2016). Anticipative interlocutive dialogism: Sequential patterns and linguistic markers in French. *Journal of Pragmatics*, 96, p. 80–95.
- Brunet, E. (2014). *La lexicométrie française : Naissance, évolution et perspectives*. 45 (1–2).
- Carter-Thomas, S. (2009). The French c'est-cleft: Function and frequency. In *La linguistique systémique fonctionnelle et la langue française*, p. 127–157. Paris : L'Harmattan.
- Carter-Thomas, S., & Jacques, M.-P. (2022). Mobilisation de routines discursives dans le discours scientifique oral : perspectives coisées. In M. Luodonpää-Manni, F. Grossmann & A. Tutin (Éd.), *Les routines discursives dans le discours scientifique oral et écrit*, p. 151–177. Grenoble : UGA Éditions.
- Cavalla, C. (2015). Collocations transdisciplinaires dans les écrits de doctorants FLS/FLE. *Linx*.

- Revue des linguistes de l'université Paris X Nanterre*, (72), p. 111-130.
- Cavalla, C. (2008). Les collocations dans les écrits universitaires : Un lexique spécifique pour les apprenants étrangers. In O. B. et I. Schaffner (Ed.), *Apprendre une langue de spécialité: Enjeux culturels et linguistiques*, p. 93–104. Editions Ecole Polytechnique.
- Cavalla, C. (2015). Collocations transdisciplinaires dans les écrits de doctorants FLS/FLE. *Linx. Revue des linguistes de l'université Paris X Nanterre*, (72), p. 111–130.
- Cavalla, C., & Tutin, A. (2008). Etude des collocations évaluatives dans les écrits scientifiques. *Europhras2008*. Présenté à Helsinki, Finland. Helsinki, Finland.
- Charaudeau, P., & Maingueneau, D. (2002). *Dictionnaire d'analyse du discours*. Paris : Éditions Seuil.
- Chevalier, G. (2019). Analyse des routines sémantico-rhétoriques à des fins didactiques. *Linguistica Atlantica*, 37(2).
- Cislaru, G., & Olive, T. (2017). Segments répétés, jets textuels et autres routines. Quel niveau de pré-construction ? *Corpus*, (17).
- Class, B., & Schneider, D. K. (2015). Manuel de recherche en technologie éducative. *Université de Genève, Suisse. publiée—EduTech Wiki(en-ligne https://edutechwiki.unige.ch/fr/Manuel_de_recherche_en_technologie_%C3%A9ducative)*.
- Colletta, J.-M. & Karcher, G. (2013). Linguistique textuelle et didactique de l'écrit. Les cours du CNED sur le site <https://www.cned.fr/>.
- Combettes, B. (1999). *Les marqueurs de topicalisation dans l'"Histoire d'un voyage en terre de Brésil" de Jean de Léry*, p. 25–40.
- Combettes, B. (2003). Les marqueurs de topicalisation en français : Un cas de grammaticalisation. In *Ordre et distinction dans la langue et le discours*, p. 149–167. Honoré Champion.
- Combettes, B. (2006). Du niveau textuel au niveau énonciatif dans la grammaticalisation : Le rôle du contexte, Abstract. *Langue française*, (149), p. 48–60.
- Combettes, B. (2007). Grammaticalisation des marqueurs de topicalisation en français : Les expressions du type pour ce qui regarde. *Langue française*, (156), p. 93–107.
- Combettes, B., & Prevost, S. (2001). Évolution des marqueurs de topicalisation. *Cahiers de praxématique*, (37), p. 103–124.

- Coulmas, F. (2011). *Conversational Routine : Explorations in Standardized Communication Situations and Prepatterned Speech*. Walter de Gruyter.
- Courtine, J.-J. (1981). Quelques problèmes théoriques et méthodologiques en analyse du discours, à propos du discours communiste adressé aux chrétiens. *Langages*, 15(62), p. 9–128.
- Crismore, A. (1983). *Metadiscourse: What it is and how it is Used in School and Non-school Social Science Texts*. University of Illinois at Urbana-Champaign.
- Crismore, A. (1984). The Rhetoric of Social Studies Textbooks: Metadiscourse. *Journal of Curriculum Studies*, 16(3), p. 279–296.
- Crismore, A. (1989). *Talking with readers: Metadiscourse as rhetorical act*. New York: Peter Lang.
- Crismore, A., Markkanen, R., & Steffensen, M. S. (1993). Metadiscourse in Persuasive Writing: A Study of Texts Written by American and Finnish University Students. *Written Communication*, 10(1), p. 39–71.
- Culioli, A. (1968). La formalisation en linguistique—Cahiers pour l'Analyse (An electronic edition). *Cahiers pour l'Analyse*, (9), p. 106–117.
- Culioli, A. (1999). *Pour une linguistique de l'énonciation : Formalisation et opérations de repérage*. Editions OPHRYS.
- Dahl, T. (2004). Textual metadiscourse in research articles: A marker of national culture or of academic discipline? *Journal of Pragmatics*, 36(10), p. 1807–1825.
- Davis, M., & Morley, J. (2018). Facilitating learning about academic phraseology: Teaching activities for student writers. *Journal of Learning Development in Higher Education*.
- Desclés, J.-P. (2012). Interactions entre les valeurs de pouvoir, vouloir, devoir. In *Interactions entre les valeurs de pouvoir, vouloir, devoir*, p. 49-66). Max Niemeyer Verlag.
- Diwersy, S., Gonon, L., Goossens, V., Kraif, O., Novakova, I., Sorba, J., & Vidotto, I. (2021). La phraséologie du roman contemporain dans les corpus et les applications de la PhraseoBase. *Corpus*, (22).
- Ducrot, O. (1984). *Le dire et le dit*. Paris : Les éditions de Minuit.
- Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL*

- Quarterly*, 42(3), p. 375–396.
- Eshkol-Taravella, I., & Grabar, N. (2018). Reformulations : De l'étude outillée dans les corpus disponibles vers leur détection automatique. *Langages*, N° 212(4), p. 5–16.
- Florez, M. (2013). La citation positionnée dans l'écrit scientifique. In *L'écrit scientifique : Du lexique au discours*, p. 67–84. Rennes : Presse universitaire de Rennes.
- Fløttum, K. (2003a). “Je” et le verbe. *F. TRIBUNE*, 14, p. 7–14.
- Fløttum, K. (2003b). Personal English, indefinite French and plural Norwegian scientific authors? Pronominal author manifestation in research articles. *Norsk Lingvistisk Tidsskrift*, 21, p. 21–55.
- Fløttum, K., Dahl, T., & Kinn, T. (2006). *Academic Voices: Across languages and disciplines*. John Benjamins Publishing.
- Flowerdew, J. (2003). Signalling nouns in discourse. *English for Specific Purposes*, 22(4), p. 329–346.
- Gombert, J.-É. (1990). *Le développement métalinguistique*. Paris : Presses universitaires de France.
- Grafton, A. T. (1998). *Les origines tragiques de l'érudition : Une histoire de la note en bas de page, traduit de l'anglais (américain) par Pierre-Antoine Fabre*. Paris : Éditions du Seuil.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications*, p. 145–160. Oxford: Oxford University Press.
- Granger, S., & Meunier, F. (2008). *Phraseology: An interdisciplinary perspective*. John Benjamins Publishing.
- Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective*, p. 27–49. John Benjamins.
- Granger, S., & Paquot, M. (2010). *The Louvain EAP Dictionary (LEAD)*, p. 321–326. Leeuwarden, The Netherlands.
- Gray, B., & Biber, D. (2015). Lexical frames in academic prose and conversation. In S. Hoffmann, B. Fischer-starcke, & A. Sand (Éds.), *Current Issues in Phraseology*, p. 09–134). Amsterdam: John Benjamins Publishing Company.
- Gries, S. Th. (2008). Phraseology and linguistic theory A brief survey. In S. Granger & F.

- Meunier (Éds.), *Phraseology: An interdisciplinary perspective*, p. 1–26. John Benjamins.
- Grobet, A. (2002). *L'identification des topiques dans les dialogues*. De Boeck Supérieur.
- Grossmann, F. (2010). L'Auteur scientifique. *Revue d'anthropologie des connaissances, Vol 4, n° 3(3)*, p. 410–426.
- Grossmann, F. (2011). Renvoyer aux sources du savoir : Voir et cf. Dans le texte scientifique. In A. Jaubert, J. M. López, Muñoz, S. Marnette, L. Rosier, & C. Stolz (Éds.), *Citations II: Citer pour quoi faire*, p. 207–222.
- Grossmann, F. (2014 a). De quelques routines phraséologiques liées aux verbes parenthétiques dans les genres scientifiques. *Congrès Mondial de Linguistique Française, 8*, p. 759–770. Berlin, Germany.
- Grossmann, F. (2014 b). Les verbes de constat dans l'écrit scientifique. In A. Tutin & F. Grossmann (Eds.), *L'écrit scientifique : Du lexique au discours. Autour de Scientext*, p. 85–100. Rennes : Presses Universitaires de Rennes.
- Grossmann, F. (2018). Verbes de constat et autres verbes “ parenthétiques ” Quel statut dans l'écrit scientifique ? *Arena Romanistica, (14)*, p. 106–120.
- Grossmann, F., & Tutin, A. (2010). Les marqueurs verbaux de constat : Un lieu de dialogisme dans l'écrit scientifique. *Actes Du Colloque Dialogisme : Langue, Discours*.
- Grossmann, F., & Tutin, A. (2013). Introduction. In A. Tutin & F. Grossmann (Éds.), *L'Écrit scientifique : Du lexique au discours. Autour de Scientext*, p. 9–23. Rennes : Presse universitaire de Rennes.
- Grossmann, F., Tutin, A., & Silva, P. P. G. D. (2009). Filiation et transfert d'objets scientifiques dans les écrits de recherche. *Pratiques. Linguistique, littérature, didactique, (143–144)*, p. 187–202.
- Guespin, L. (1971). *Le discours politique*. Paris : Didier Larousse.
- Guiraud, P. (1959). *Problèmes et méthodes de la statistique linguistique*. D. Reidel.
- Halliday, M. A. K. (1973). *Explorations in the Functions of Language*. Edward Arnold (Publishers) Ltd.
- Halliday, M. A. K. (1985). *An Introduction to functional grammar*. Londres: E. Arnold.
- Halliday, M. A. K., & Martin, J. R. (2003). *Writing Science: Literacy And Discursive Power*. Routledge.

- Halliday, M., & Matthiessen, C. M. I. M. (2014). *An Introduction to Functional Grammar*. Routledge.
- Harris, Z. S. (1959). The Transformational Model of Language Structure. *Anthropological Linguistics*, 1(1), p. 27–29. JSTOR.
- Hatier, S. (2016). *Identification et analyse linguistique du lexique scientifique transdisciplinaire. Approche outillée sur un corpus d'articles de recherche en SHS* (Thèse de doctorat). Université Grenoble Alpes. Grenoble.
- Hatier, S., Augustyn, M., Yan, R., Tran, T. T. H., Tutin, A., & Jacques, M.-P. (2016). French cross-disciplinary scientific lexicon: Extraction and linguistic analysis. In T. Margalitadze & G. Meladze (Éds.), *Proceedings of the XVII EURALEX International congress Lexicography & Linguistic diversity*, p. 355–365. Tbilisi: Ivane Javakhishvili Tbilisi State University.
- Hausmann, F. J., & Blumenthal, P. (2006). Présentation : Collocations, corpus, dictionnaires. *Langue française*, (150), p. 3–13.
- Howarth, P. (1998). Phraseology and Second Language Proficiency. *Applied Linguistics*, 19(1), p. 24–44.
- Hyland, K. (2001). Bringing in the Reader: Addressee Features in Academic Articles. *Written Communication*, 18(4), p. 549–574.
- Hyland, K. (2005). *Metadiscourse: Exploring Interaction in Writing*. Bloomsbury Publishing.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), p. 4–21.
- Hyland, K. (2009). *Academic Discourse: English In A Global Context*. A&C Black.
- Hyland, K. (2012). Bundles in Academic Discourse. *Annual Review of Applied Linguistics*, 32, p. 150-169.
- Hyland, K. (2017). Metadiscourse: What is it and where is it going? *Journal of Pragmatics*, 113, p. 16–29.
- Hyland, K., & Bondi, M. (2006). *Academic Discourse Across Disciplines*. Peter Lang.
- Intaraprawat, P., & Steffensen, M. S. (1996). The Use of Metadiscourse in Good and Poor ESL Essays. *Journal of Second Language Writing*, 4(3), p. 253–272.

- Ibekwe-Sanjuan, F., Chen, C., & Roberto, P. (2008). Identifying Strategic Information from Scientific Articles through Sentence Classification. *6th International Conference on Language Resources and Evaluation Conference (LREC-08)*, p. 1518–1522. Marrakesh, Morocco: ELDA.
- Intaraprawat, P., & Steffensen, M. S. (1996). The Use of Metadiscourse in Good and Poor ESL Essays. *Journal of Second Language Writing*, 4(3), p. 253–272.
- Jacques, M.-P. (2005). Structure matérielle et contenu sémantique du texte écrit. *Corela. Cognition, représentation, langage*, 3(2).
- Jacques, M.-P. (2011). “Nous appelons X cet Y” : X est-il un terme émergent? *9th International Conference on Terminology and Artificial Intelligence (TIA)*, p. 31–37, Paris, France, 8 au 10 novembre 2011.
- Jacques, M.-P. (2014). Structure textuelle de l'article scientifique. Les intertitres et la construction rhétorique en sciences humaines et sociales. In A. Tutin & F. Grossmann (Éds.), *L'écrit scientifique : Du lexique au discours. Autour de Scientext*, p. 199–215. Rennes : Presses Universitaires de Rennes.
- Jacques, M.-P. (2017). La structuration textuelle en discours scientifique : Comparaison oral/écrit. *CHIMERA: Romance Corpora and Linguistic Studies*, 4(1), p. 89–116.
- Jakobson, R. (1963). Essais de Linguistique Générale. *Les Etudes Philosophiques*, 18(4), p. 465–465.
- Jacques, M.-P., & Tutin, A. (Éds.). (2018). *Lexique transversal et formules discursives des sciences humaines*. Londre : ISTE Editions.
- Jacquey, E., Tutin, A., Kister, L., Jacques, M.-P., Hatier, S., & Ollinger, S. (2013). Filtrage terminologique par le lexique transdisciplinaire scientifique : Une expérimentation en sciences humaines. *TIA 2013*, p. 121–128.
- Jakobson, R. (1963). Essais de Linguistique Générale. *Les Etudes Philosophiques*, 18(4), p. 465–465.
- Jakobson, R. (1985). Metalanguage as a linguistic problem. *Selected Writings VII*, p. 113–121.
- Jaworski, A., Coupland, N., & Galasinski, D. (2012). *Metalanguage: Social and Ideological Perspectives*. Berlin : Walter de Gruyter.

- Ji, Y., & Tutin, A. (2019). Les routines métalinguistiques dans les écrits scientifiques en français. *Studii de lingvistică*, 9(2), p. 177–200.
- Ji, Y., & Tutin, A. (2020). Les routines de topicalisation dans les écrits scientifiques en français. In M. Mangeot & A. Tutin (Éds.), *Lexique(s) et genre(s) textuel(s) : Approches sur corpus-Actes de la conférence LTT 2018—11es journées du réseau Lexicologie, Terminologie, Traduction à Grenoble*, p. 227–242. Grenoble, France.
- Kerbrat-Orecchioni, C. (2005). *Le discours en interaction*. Paris : Armand Colin.
- Kerbrat-Orecchioni, C. (2009). *L'énonciation : De la subjectivité dans le langage*. Paris : Armand Colin.
- Kleiber, G. (1984). Dénomination et relations dénominatives. *Langages*, 19(76), p. 77–94.
- Kleiber, G. (2001). Remarques sur la dénomination. *Cahiers de praxématique*, (36), p. 21–41.
- Kraif, O. (2016). Le lexicoscope : Un outil d'extraction des séquences phraséologiques basé sur des corpus arborés. *Cahiers de Lexicologie*, 1(108), p. 91–106.
- Kraif, O., Novakova, I., & Sorba, J. (2016). Constructions lexico-syntaxiques spécifiques dans le roman policier et la science-fiction. *Lidil. Revue de linguistique et de didactique des langues*, (53), p. 143–159.
- Kraif, O., & Tutin, A. (2017). Des motifs séquentiels aux motifs hiérarchiques : L'apport des arbres lexico-syntaxiques récurrents pour le repérage des routines discursives. *Corpus*, (17).
- Langhans, B. (2005). Les marqueurs formés sur terme. In A. Steuckardt, A. Niklas-Salminen, F. Douay, D. Bouverot, & Collectif (Eds.), *Les marqueurs de glose*, p. 37–50. Aix-en-Provence : Presses de L'Université de Provence.
- Kraif, O., & Tutin, A. (2020). Collocations et linguistique de corpus : L'intuition des linguistes et les critères quantitatifs convergent-ils ? *Le Français Moderne - Revue de linguistique Française*, (n° 1), p. 84–101.
- Lane, P. (2008). Les frontières des textes et des discours : Pour une approche linguistique et textuelle du paratexte. *Congrès Mondial de Linguistique Française 2008*, p. 1379-1387. Paris : EDP Sciences.

- Langhans, B. (2005). Les marqueurs formés sur terme. In A. Steuckardt, A. Niklas-Salminen, F. Douay, D. Bouverot, & Collectif (Éds.), *Les marqueurs de glose*, p. 37–50. Aix-en-Provence : Presses de L'Université de Provence.
- Lautamatti, L. (1978). Observations on the Development of the Topic in Simplified Discourse. *AFinLAn vuosikirja*, p. 71–104.
- Lefebvre, M. (2006). Les écrits scientifiques en action : Pluralité des écritures et enjeux mobilisés. *Sciences de la société*, (67), p. 3–15.
- Legallois, D. (2009). À propos de quelques n-grammes significatifs d'un corpus poétique du XIXe siècle. *L'information grammaticale*, 121(1), p. 46–52.
- Legallois, D., Charnois, T., & Poibeau, T. (2016). Repérer les clichés dans les romans sentimentaux grâce à la méthode des « motifs ». *Lidil. Revue de linguistique et de didactique des langues*, (53), p. 95–117.
- Legallois, D., & Grea, P. (2006). L'objectif de cet article est de... Construction spécificationnelle et grammaire phraséologique. *Cahiers de praxématique*, (46), p. 161–186.
- Legallois, D., & Tutin, A. (2013). Présentation : Vers une extension du domaine de la phraséologie. *Langages*, (189), p. 3–25.
- Lewis, M. (1993). *The Lexical approach: The state of ELT and a way forward*. Hove, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord : Language teaching.
- Lewis, M. (1996). Pedagogical implications of the lexical approach. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (1st ed.), p. 255–270. Cambridge University Press.
- Lewis, M. (1997). *Implementing the Lexical Approach : Putting Theory Into Practice*. Language Teaching Publications.
- Loffler-Laurian, A. M. (1983). Typologie des discours scientifiques : Deux approches. *Études de Linguistique Appliquée*, 51, p. 8–20.
- Longrée, D., Luong, X., & Mellet, S. (2008). Les motifs : Un outil pour la caractérisation topologique des textes. *JADT*. Lyon, France. p. 733–744.
- Longrée, D., & Mellet, S. (2013). Le motif : Une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours, From language to speech : the 'motif',

- a new item to widen the scope of phraseology? *Langages*, (189), p. 65–79.
- Longrée, D., & Mellet, S. (2015). Les variantes des motifs chez les prosateurs latins : Entre récurrence générique et spécificité d’auteur, des formes révélatrices et caractérisantes. *Les Études Classiques*, 82(1–2), p. 65–87.
- Luodonpää-Manni, M., Grossmann, F., & Tutin, A. (Éds.). (2022). *Les routines discursives dans le discours scientifique oral et écrit*. Grenoble : UGA Éditions.
- Maingueneau, D. (1991). *L’analyse du discours : Introduction aux lectures de l’archive*. Paris : Hachette supérieur.
- Maingueneau, D. (2016). *Analyser les textes de communication*. Paris : Armand Colin.
- Maingueneau, D. (2021). *Discours et analyse du discours* (2ème édition). France : Armand Colin.
- Malherbe, M., & Pousseur, J.-M. (1985). *Francis Bacon, science et méthode : Actes du colloque de Nantes*. Paris : Vrin.
- Marchello-Nizia, C. (2007). Le principe de surprise annoncée. Grammaticalisation et pragmatization de cependant. *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics*, (1).
- Mayaffre, D. (2008). De l’occurrence à l’isotopie. *Syntaxe et Semantique*, N° 9(1), p. 53–72.
- Mel’Čuk, I. (2012). Phraseology in the language, in the dictionary, and in the computer. *Yearbook of Phraseology*, 3(1), p. 31–56.
- Mellet, S., & Longrée, D. (2009). Syntactical Motifs and Textual Structures: Considerations based on the Study of a Latin historical Corpus. *Belgian Journal of Linguistics*, 23(1), p. 161–174.
- Mellet, S., & Longrée, D. (2012). Légitimité d’une unité textométrique : Le motif. *ADT 2012, 11èmes Journées internationales d’analyse statistique des données textuelles*, Liège, Belgique, p. 715–728.
- Meyer, B. J. F. (1975). Identification of the Structure of prose and Its Implications for the Study of Reading and Memory. *Journal of Reading Behavior*, 7(1), p. 7–47.
- Muller, C. (1969). La statistique lexicale. *Langue française*, 2(1), p. 30–43.
- Muller, C. (1979). *Langue française et linguistique quantitative*. Genève : Slatkine.

- Nair, P. K. R., & Nair, V. D. (2014). Organization of a Research Paper : The IMRAD Format. In P. K. R. Nair & V. D. Nair (Éds.), *Scientific Writing and Communication in Agriculture and Natural Resources*, p. 13–25. Cham: Springer International Publishing.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford : Oxford University Press.
- Née, E., Sitri, F., & Veniard, M. (2014). Pour une approche des routines discursives dans les écrits professionnels. *SHS Web of Conferences*, 8, p. 2113–2124.
- Née, É., Sitri, F., & Veniard, M. (2016). Les routines, une catégorie pour l’analyse de discours : Le cas des rapports éducatifs. *Lidil. Revue de linguistique et de didactique des langues*, (53), p. 71–93.
- Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, 11(3), p. 283–304.
- Novakova, I., & Sorba, J. (2017). À la recherche de motifs autour de meurtre, crime, assassinat dans le roman policier contemporain. *Synergies Pologne* (14), p. 29–42.
- Ounoughi, S., & Duvillard, S. (2016). Pour une évolution vers l’interdisciplinarité : Analyse du discours et géographie. *23rd annual conference on Critical Geographies : Situated Solidarities*. Présenté à Grenoble, France.
- Pecman, M. (2004). *Phraséologie contrastive anglais-français : Analyse et traitement en vue de l’aide à la rédaction scientifique* (Thèse de doctorat). Université de Nice-Sophia Antipolis. Nice.
- Péry-Woodley, M.-P. (2000). *Une pragmatique à fleur de texte : Approche en corpus de l’organisation textuelle* (Thèse de doctorat). Université Toulouse le Mirail - Toulouse II. Toulouse.
- Pontille, D. (2007). Matérialité des écrits scientifiques et travail de frontières : Le cas du format IMRAD. In *Sciences et frontières*, p. 229–253. Louvain-la-Neuve : E.M.E. Éditions.
- Poudat, C. (2006). *Étude contrastive de l’article scientifique de revue linguistique dans une perspective d’analyse des genres* (Thèse de doctorat). **Université d’Orléans. Orléans.**
- Prévoist, S. (2003). Détachement et topicalisation : Des niveaux d’analyse différents. *Les cahiers de praxématique*, (40), p. 97–126.
- Quiniou, S., Cellier, P., Charnois, T., & Legallois, D. (2012). Fouille de données pour la

- stylistique : Cas des motifs séquentiels émergents. *11es Journées Internationales d'Analyse Statistique Des Données Textuelles (JADT'12)*, p. 821–833. Liège, Belgium.
- Rahman, M. (2004). Aiding the Reader: The use of Metalinguistic Devices in Scientific Discourse. *Nottingham Linguistic Circular*, 18, p. 29–48.
- Rastier, F. (2005). Enjeux épistémologiques de la linguistique de corpus. *La linguistique de corpus*, p. 31–45.
- Rastier, F., & Pincemin, B. (1999). Des genres à l'intertexte. *Cahiers de praxématique*, (33), p. 83–111.
- Rebeyrolle, J., & Tanguy, L. (2000). Repérage automatique de structures linguistiques en corpus : Le cas des énoncés définitoires. *Cahiers de Grammaire, Sémantique et corpus*(25), p. 153–174.
- Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Éds.), *English Corpus Linguistics : Studies in Honour of Jan Svartvik*. Longman.
- Rey-Debove, J. (1997). *Le métalangage : Étude linguistique du discours sur le langage*. Paris, France : Armand. Colin.
- Riegel, M. (1990). La définition, acte du langage ordinaire. De la forme aux interprétations. *La Définition*, p. 97–110.
- Rinck, F. (2006). *L'article de recherche en Sciences du langage et en Lettres : Figure de l'auteur et identité disciplinaire du genre*. (Thèse de doctorat). Université Stendhal (Grenoble).
- Rinck, F. (2010). L'analyse linguistique des enjeux de connaissance dans le discours scientifique, Abstract : Linguistic analysis of knowledge issues in scientific discourse. A literature review. *Revue d'anthropologie des connaissances*, 4, n° 3(3), p. 427–450.
- Rinck, F., & Tutin, A. (2007). Annoter la polyphonie dans les textes : Le cas des passages entre guillemets. *Corpus*, (6), p. 79–100.
- Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Construction*, 3(1), p. 95–119.
- Rossari, C. (1993). *Les Operations de Reformulation : Analyse Du Processus Et Des Marques Dans Une Perspective Contrastive Francais - Italien*. Berne ; New York: Peter Lang

- Gmbh, Internationaler Verlag Der Wissenschaften.
- Salem, A. (1986). Segments répétés et analyse statistique des données textuelles. *Histoire & Mesure*, 1(2), p. 5–28.
- Salem, A. (1987). *Pratique des segments répétés : Essai de statistique textuelle* (Laboratoire Lexicométrie et textes politiques, Éd.). Paris, France : Klincksieck.
- Sandor, A. (2007). Modeling metadiscourse conveying the author's rhetorical strategy in biomedical research abstracts. *Revue Française de Linguistique Appliquée*, Vol. XII(2), p. 97-108.
- Schmid, H.-J. (2000). *English Abstract Nouns as Conceptual Shells: From Corpus to Cognition*. Walter de Gruyter.
- Siepmann, D. (2005). *Discourse Markers Across Languages. A Contrastive Study of Second-Level Discourse Markers in Native and Non-Native Text with Implications for General and Pedagogic Lexicography* (1ère). London: Routledge.
- Siepmann, D. (2007). Les marqueurs de discours polylexicaux en français scientifique. *Revue française de linguistique appliquée*, XII(2), p. 123–136.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), p. 487–512.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford : Oxford University Press.
- Sinclair, J., Jones, S., & Daley, R. (2004). *English Collocation Studies: The OSTI Report*. (R. Krishnamurthy, Éd.). Bloomsbury Publishing.
- Sitri, F., & Tutin, A. (2016). Présentation. *Lidil. Revue de linguistique et de didactique des langues, Phraséologie et genres de discours : Patrons, motifs, routines*(53), p. 5–18.
- Sitri, F., & Veniard, M. (2017). Routines discursives, variation et normes de genre. *Langage et société*, N° 159(1), p. 99–114.
- Skubis, I. (2016). Analyse du texte scientifique-les métatextèmes. *Episteme*, p. 83–96.
- Steuckardt, A. (2009). Décrire la reformulation : Le paramètre rhétorique. *Cahiers de praxématique*, (52), p. 159–172.
- Steuckardt, A. (2018). Les marqueurs de reformulation formés sur dire : Exploration outillée. *Langages*, N° 212(4), p. 17–34.

- Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
- Tesnière, L. (1965). *Eléments de syntaxe structurale*. Paris, France: Klincksieck.
- Thomas, S., & Hawes, T. P. (1994). Reporting verbs in medical journal articles. *English for Specific Purposes*, 13(2), p. 129-148.
- Todorov, T. (1981). *Mikhaïl Bakhtine. Le principe dialogique. Suivi de : Ecrits du Cercle de Bakhtine*. Le Seuil.
- Tran, T. T. H. (2014). *Description de la phraséologie transdisciplinaire des écrits scientifiques et réflexions didactiques pour l'enseignement à des étudiants non-natifs : Application aux marqueurs discursifs* (Thèse de doctorat). Université de Grenoble. Grenoble.
- Tutin, A. (2007). Collocations du lexique transdisciplinaire des écrits scientifiques : Annotation et extraction des propriétés linguistiques dans la perspective d'une application didactique. *Cahiers de l'Institut de Linguistique de Louvain*, 31(2-4), p. 247-262.
- Tutin, A. (2010). *Sens et combinatoire lexicale : De la langue au discours. Dossier en vue de l'obtention de l'habilitation à diriger des recherches (HDR)*. Université Stendhal, Grenoble III., Grenoble, France.
- Tutin, A. (2014). La phraséologie transdisciplinaire des écrits scientifiques : Des collocations aux routines sémantico-rhétoriques. In A. Tutin & F. Grossmann (Eds.), *L'écrit scientifique : Du lexique au discours. Autour de Scientext*, p. 27-44. Rennes : Presses Universitaires de Rennes.
- Tutin, A. (2016, décembre). *Les collocations : Des associations binaires aux routines discursives*. Présenté à Université Paris-Sorbonne. Université Paris-Sorbonne.
- Tutin, A. (2017). La mise en scène de la surprise dans les écrits scientifiques de sciences humaines. *TRANEL. Travaux Neuchâtelois de Linguistique*, (65), p. 19-35.
- Tutin, A. (2018 a). Les expressions polylexicales transdisciplinaires dans les articles de recherche en sciences humaines : Retour d'expérience (Chapitre 3). In Jacques. M. P. & Tutin. A. (Éd.), *Lexique transversal et formules discursives des sciences humaines*, p. 73-89. Londres : ISTE.

- Tutin, A. (2018 b). Les routines phraséologiques du discours scientifique. *Phrasèmes en discours*. Présenté à Padova, Italy. Padova, Italy : G. Henriot.
- Tutin, A., & Grossmann, F. (2013). *L'Écrit scientifique : Du lexique au discours. Autour de Scientext*. Presses universitaires de Rennes.
- Tutin, A., & Ji, Y. (2022). Comment modéliser les routines sémantico-rhétoriques dans le discours scientifique ? In M. Luodonpää-Manni, F. Grossmann & A. Tutin (Éd.), *Les routines discursives dans le discours scientifique oral et écrit*, p. 37–64. Grenoble : UGA Éditions.
- Tutin, A., & Kraif, O. (2016). Routines sémantico-rhétoriques dans l'écrit scientifique de sciences humaines : L'apport des arbres lexico-syntaxiques récurrents. *Lidil. Revue de linguistique et de didactique des langues*, (53), p. 119–141.
- Vande Kopple, W. J. (1985). Some Exploratory Discourse on Metadiscourse. *College Composition and Communication*, 36(1), p. 82–93.
- Vande Kopple, & J, W. (1997). *Refining and Applying Views of Metadiscourse*. ERIC Clearinghouse.
- Vuillaume, M. (2018). Ici, un déictique pas tout à fait comme les autres. *Langue française*, 197(1), p. 5–34.
- Wahl, A., & Gries, S. Th. (2018). Multi-word Expressions: A Novel Computational Approach to Their Bottom-Up Statistical Extraction. In P. Cantos-Gómez & M. Almela-Sánchez (Éds.), *Lexical Collocation Analysis: Advances and Applications*, p. 85–109. Cham: Springer International Publishing.
- Williams, J. M. (1981). *Style: Ten lessons in clarity & grace*. Glenview, Ill: Scott, Foresman.
- Wilmet, M. (2010). *Grammaire critique du français*. De Boeck Supérieur.
- Yan, R. (2017). *Étude des constructions verbales scientifiques dans une perspective didactique : Utilisation des corpus pour le diagnostic des besoins langagiers en FLE à l'aide des techniques de TAL*. (Thèse de doctorat). Université Grenoble Alpes. Grenoble.
- Zufferey, S. (2020). *Introduction à la linguistique de corpus*. Londres : ISTE.

Annexe 1

Les classes sémantiques du LST utilisées comme pivot de recherche dans notre étude (65 classes du LST + une classe créée par nous). Les mots soulignés en rouge sont ajoutés par nous.

Verbes				
1	Analyse_info Analyse des informations : description des données, analyse des relations entre les variables, comparaison des résultats attendus, interprétation des écarts, analyse statistique, résultats et conclusions.	Action scientifique qui vise à prendre connaissance de quelque chose au moyen d'observations scientifiques	#Constat	constater, identifier, noter, observer, prendre acte, prendre connaissance, relever_1, repérer_1, repérer_2, voir, connaître, savoir, remarquer
2	Analyse_info	Action scientifique qui vise à mettre en évidence, à démontrer la vérité, la réalité, la valeur de quelque chose.	#Démonstration	démontrer_1, mettre au jour, mettre en évidence, mettre en exergue, mettre en lumière, mettre en relief, mettre en valeur, établir_2, montrer_2
3	Analyse_info	Action scientifique qui vise à étudier en détail, faire l'analyse de quelque chose pour en avoir une connaissance approfondie, juger de sa valeur.	#Examen	aborder, analyser, approfondir, considérer_2, contrôler, critiquer, étudier, examiner, explorer, faire le point, passer en revue, prendre en compte, prendre en considération, soumettre, tenir compte, tester, traiter_1, traiter_2, vérifier_1
4	Analyse_info	Donner un sens précis, parmi d'autres possibles, à un acte, à un fait, dont l'explication n'apparaît pas de manière évidente.	#Interprétation	éclairer_1, expliciter_1, expliquer_1, interpréter, justifier_1, raisonner
5	Analyse_info	Faire sortir d'un fait ou d'un événement la conséquence logique qu'elle contient implicitement.	#Analyse_inférence	conclure_2, déduire
6	Analyse_info	Action scientifique qui vise à délimiter des propriétés de qqc.	#Détermination	définir_1, déterminer_1
7	Analyse_info	Reconsidérer quelque chose d'abstrait en y apportant des informations supplémentaires, des éclaircissements et des modifications.	#Réexamen	mettre en perspective, nuancer, préciser_1, redéfinir, replacer, relativiser
8	Communication Classe regroupant les verbes de communication et de transfert des idées (selon le type de message, le moyen et la manière de transmission, faisant ou non partie d'une interaction).	Action d'attacher et de donner une importance particulière à quelque chose.	#Insistance	insister, mettre l'accent, souligner
9	Communication	Acte communicatif par lequel on rend manifeste par toutes sortes de signes (langage écrit, oral, etc.), de façon volontaire ou non, ce que l'on pense.	#Diffusion	citer_1, citer_2, diffuser, exposer, exprimer_1, garantir_2, indiquer_1, mentionner, préciser_2, reporter_2, signaler_1 dire, parler
10	Communication	Acte communicatif par lequel on énonce avec la concision et la netteté d'une formule.	#Formulation	émettre, énoncer, évoquer_2, faire allusion, formuler, invoquer, résumer

11	État Classe des verbes renvoyant à un état au sens général : une manière d'être, une situation ou une condition statique.	État de se montrer dans sa réalité, de se trouver dans une certaine situation.	Existence	agir (il s')_2, apparaître (il)_2, avérer (il s')_2, constituer_2, décliner (se)_2, demeurer_1, demeurer (il)_2, exister (il)_2, faire défaut, présenter (se)_3 rester
12	État	État d'avoir un lien avec quelque chose, avoir pour objet quelque chose.	Rapport	affecter, avoir trait, concerner, désigner_2, faire écho, intéresser_1, porter_3, référer (se)_2, renvoyer_2, viser_2 rappporter (se)
13	État	État d'avoir la même valeur qualitative, d'avoir pour équivalence ou de constituer le trait distinctif ou dominant de quelque chose tout en le rendant reconnaissable	Représentation	caractériser_2, composer_2, consister_1, équivaloir, faire figure, faire office, figurer_2, représenter_2, résider, signifier, symboliser, tenir lieu, entendre
14	Modalité Expression de l'attitude du locuteur par rapport au contenu propositionnel de son énoncé.		#Obligation	imposer_1 devoir
15	Modalité	Évaluation de la nécessité.	Nécessité	convenir_2, convenir (il)_3, exiger, importer (il), nécessiter, falloir
16	Processus humain Classe regroupant les verbes renvoyant aux processus humains généraux, qui ne relèvent ni du cognitif ni du domaine scientifique.	Faire usage de quelque chose, appliquer un procédé, une technique, faire agir un objet, une méthode afin d'obtenir un effet qui permette d'aboutir à une fin déterminée.	#Usage	accéder, adapter_3, appliquer, employer_2, emprunter, exploiter_1, mettre en œuvre, mobiliser_1, pratiquer, recourir, tirer profit, utiliser, prendre, reprendre, servir
17	processus nonhumain Classe regroupant les verbes renvoyant aux processus non humains qui ne sont pas statiques, qui expriment une action.	Rendre manifeste, laisser apparaître clairement la vérité d'un fait, les caractéristiques d'une entité abstraite.	Révélation	affirmer (s')_2, décrire_2, démontrer_2, dessiner_2, éclairer_2, expliciter_2, exprimer_2, illustrer_2, indiquer_2, marquer, montrer_1, refléter_1, révéler_1, signaler_2, spécifier_2, témoigner, véhiculer
18	processus nonhumain Classe regroupant les verbes renvoyant aux processus non humains qui ne sont pas statiques, qui expriment une action.	Faire usage de quelque chose, appliquer un procédé, une technique, faire agir un objet, une méthode afin d'obtenir un effet qui permette d'aboutir à une fin déterminée.	Usage	exploiter_2, mettre en jeu, mobiliser_2
19	Ancrage construction Verbes servant à la construction du modèle d'analyse : problématique, hypothèses, conceptualisation, définition, réflexion, références.	Avoir en vue et mettre en œuvre son action dans un but déterminé.	#Finalité	attacher (s')_2, consacrer (se)_1, efforcer (s'), employer (s')_1, prendre soin, proposer (se)_2, tendre_1, tenter, viser_1, destiner, chercher (à)
20	Ancrage construction	Se décider, se déterminer entre deux ou plusieurs choses ; opter pour un parti, pour une solution.	#Choix	adopter, choisir_1, choisir_2, conserver_2, prendre en charge, prendre parti, privilégier, retenir, sélectionner, préférer
21	Ancrage construction	Rassembler ses forces intellectuelles ou affectives et les faire porter sur un certain objet, à l'exclusion de tout autre ; y appliquer avec force (sa pensée, son attention, etc.).	#Thème	avoir affaire, centrer, concentrer_2, consulter, focaliser (se), intéresser (s')_2, interroger (s')_1, questionner

22	Ancrage_construction	Processus par lequel on forme ou crée dans son esprit une idée, un projet, une œuvre.	#Conception	concevoir_1, constituer_1, définir_2, développer_2, formaliser, structurer_1
23	Ancrage_construction	Action scientifique qui vise à attribuer un nom à quelque chose. N[sujet] +V+ N[COD]	#Dénomination	désigner_1, nommer, qualifier, appeler, dénommer, dénoter
24	Ancrage_construction	Donner un sens précis, parmi d'autres possibles, à un acte, à un fait, dont l'explication n'apparaît pas de manière évidente.	#Interprétation	éclairer_1, expliciter_1, expliquer_1, interpréter, justifier_1, raisonner
25	Ancrage_construction	Action scientifique qui vise à se servir de quelque chose comme appui, repère ; faire référence à.	#Référence	faire appel, référer (se)_1, reporter (se)_1
26	Observation Actions qui ont trait à la sélection des unités d'observation, à la collecte des données, au recueil des informations, et à l'expérimentation.	Considération générale sans jugement ou orientation particulière.	#Opinion_neutre	admettre_1, affirmer_1, assumer, concevoir_2, considérer_1, considérer_3, convenir_1, envisager, estimer_2, juger_1, prendre conscience, prétendre, reconnaître, représenter (se)_3, supposer_2, trouver, penser, percevoir, croire, essayer,
27	Temporalité Classe de verbes caractérisant un processus par rapport au temps : durée, fréquence, chronologie (localisation sur l'axe temporel), début et fin.	Verbe marquant le début d'un processus (sujet humain)	#Début	Amorcer, Commencer, lancer
28	Temporalité	Verbe marquant la fin d'un processus (sujet humain)	#fin	aboutir_2, conclure_1 finir
29	Passage	Verbe indiquant les étapes énonciatives ou logiques		Passer, suivre, revenir, retourner, rappeler, poursuivre, revoir
Noms				
30	État_qualité État ou qualité d'une entité abstraite ou concrète. avoir un <u>N</u> /être de <u>N</u>	Degré d'importance. <i>conférer un <u>N</u></i>	Importance_positif	importance, intensité, poids, portée, priorité
31		Évaluation de l'obligation la nécessité. <i>Répondre à un <u>N</u></i>	Nécessité	exigence, nécessité
32	Processus_cognitif Processus cognitif permettant de structurer/analyser les observables et/ou objets construits. <i>une personne fait/effectue/propose un <u>N</u></i>	Phase dans l'analyse, d'identification d'entités/de phénomènes. <i>un <u>N</u> montre que/un <u>N</u> visuel</i>	Constat	bilan, connaissance_2, constat, découverte, état_2, identification, localisation, observation, perception, prise de conscience, reconnaissance
33	Processus_cognitif	Processus de mise en évidence de qqc. <i>Un phénomène est éclairé par un <u>N</u> de SN</i>	Démonstration	démonstration, explication_1, mise au jour, mise en évidence, mise en valeur
34	Processus_cognitif	Processus de présentation d'un élément. <i>proposer un <u>N</u> détaillé</i>	Description	description, illustration_2, mise en scène, présentation
35	Processus_cognitif	Observation attentive dans un but d'explication. <i>fournir/conduire DétPoss <u>N</u> (détaillé/approfondi/systématique)</i>	Examen	analyse, considération_1, étude, étude de cas, examen, expertise, mise en perspective, prise en compte, prise en considération

36	Communication_support Acte de communication et de transfert des idées (selon le moyen de transmission). <i>Le N est consacré à</i>	Document écrit. <i>Lire, publier un N</i>	Document	article, compte rendu, document, documentation, essai_1, littérature, note_1, ouvrage, publication, synthèse, texte, thèse 1, volume 1,
37	Communication_support	Représentation graphique éclairant le texte. <i>Le N {numéral} montre/cf. N, sur N, dans N</i>	Graphique	figure_1, illustration_1, image_1, tableau
38	Communication_support	Sous-partie d'un document écrit. <i>Dans le N de l'ouvrage/cf. N</i>	Section	annexe, bibliographie, chapitre, conclusion_1, développement_1, extrait, introduction_1, note_2, section 1
39	Communication_expression : Acte de communication et de transfert des idées (selon le type de message). <i>Par/dans ce N, l'auteur précise</i>	Élément d'un énoncé. <i>Dans l'article, l'auteur utilise/énonce un N</i>	Énoncé	affirmation, définition, discours, expression, formulation, formule, mot, proposition, terme, vocabulaire, reformulation, phrase, citation, Notion (un peu de changement)
40	Communication_expression :		Mention	référence_1
41	Objet_scientifique Observables et objets construits de l'activité scientifique. <i>étudier/collecter/s'appuyer sur un N</i>	Représentation d'une entité par une autre. <i>X est considéré comme un N (de/pour) Y</i>	Représentation	équivalent, figure_2, genre, image_2, indication, indice_1, instance, marque, modèle_2, point de repère, référence_2, représentation, signe, symbole, traduction, type
42	Objet_scientifique	Thème, sujet abordé <i>L'article, l'auteur aborde/soulève un N</i>	Thème	centre d'intérêt, considération_2, enjeu, paradoxe, point_1, problématique, problème_2, projet_2, question, sujet, thématique, thème
43	Objet_scientifique	Élément servant de base à une explication, un jugement. <i>Plusieurs N expliquent ceci</i>	Explicatif_simple	concept, contenu, critère, exemple, facteur, indicateur, notion, paramètre, précision_2, sens_2, signification, variable.
44	Processus_cognitif Processus cognitif permettant de structurer/analyser les observables et/ou objets construits. <i>une personne fait/effectue/propose un N</i>	But déterminé d'une action. <i>Nous visons tel N</i>	Finalité	but, cible, finalité, objectif, objet_2, projet_1
45	Processus_cognitif	Processus de délimitation des propriétés de qqc. <i>un N repose sur SN</i>	Détermination	définition_2, détermination
46	Quantité Renvoie à la notion de quantité, de mesure. <i>Calculer un N/un N s'élève à</i>	Unité de mesure, de calcul. <i>ajouter un N au total</i>	Unité	chiffre, nombre, point_2
Adjectifs				
47	Importance	Dénote le degré d'importance, de notoriété : dont les effets sont grands, qui exerce une grande influence ; qui a de la valeur.	Importance_positif	central_2, critique_3, crucial, de base, de fond, de grande ampleur, de premier plan, de taille, déterminant, dominant, essentiel, fondamental, fondateur, important_1, intrinsèque, majeur, notable, prépondérant, primordial, principal,

				remarquable_2, révélateur, significatif, substantiel
48	Modalité		Certitude	certain_1, évident, supposé, susceptible sûr
49	Modalité		Nécessité	incontournable, indispensable, nécessaire
50	Temporalité	Caractérise une entité par rapport au temps : durée, fréquence, chronologie (localisation sur l'axe temporel), vitesse	Chronologie	actuel, annuel, antérieur, contemporain_1, contemporain_2, d'actualité, en construction, en cours, en débat, en discussion, en vigueur, final, futur, immédiat, proche_2, récent, simultané, temporel, ultérieur,
51	Temporalité_espace	Classe créée pour pallier une certaine discontinuité entre la catégorie « temps » et « espace ». Principalement des adjectifs discursifs/métatextuels.	Énumération	initial, préalable, précédent, premier, second, successif, suivant dernier, trois, quatre, cinq, deuxième, troisième, quatrième
52	Temporalité_espace		Temporalité_espace autre	absent, existant, extrême_1, présent
Adverbes				
53	Modalité	Introduit une garantie du locuteur sur un fait apparaissant comme certain, évident.	Certitude	a fortiori, à l'évidence, bel et bien, bien entendu, bien sûr, de toute évidence, évidemment, sans ambiguïté, sans aucun doute, véritablement_1
54	Modalité	Indique un élément nécessaire.	Nécessité	forcément, nécessairement
55	Discursif	Signale un changement de perspective ou non sur le thème développé.	Addition	aussi_3, d'ailleurs, de plus, de surcroît, du reste, également_3, en outre, en plus, entre autres, par ailleurs, voire
56	Discursif	Marque le dernier élément d'une série.	Conclusion	donc_1, en conclusion, enfin, in fine, pour conclure
57	Discursif	Énonce successivement les parties d'un tout. Le segment textuel est donc considéré comme l'un des termes d'une énumération.	Énumération	comme suit, d'abord_2, d'emblée, dans un premier temps, dans un second temps, deuxièmement, en premier_1, en premier lieu, ensuite_1, premièrement, tout d'abord
58	Discursif	Permet de reprendre les points principaux.	Récapitulation	bref, en bref, en résumé, en somme, en un mot
59	Discursif	Met en relief deux aspects (complémentaires ou opposés) d'un élément.	Structuration	à son tour, d'autre part, d'une part, dans le cas présent, en parallèle
60	Discursif	Signale un changement de perspective ou non sur le thème développé.	Clarification	autrement dit, concrètement, en d'autres termes, plus exactement, plus précisément, précisément_2, stricto sensu
61	Discursif	Permet de reconsidérer quelque chose en y apportant des informations supplémentaires, des éclaircissements et des modifications.	Réexamen	à terme_2, à vrai dire, après tout, au demeurant, au fond, au total_2, dans la pratique, dans les faits, de facto, de fait, effectivement, en définitive, en effet, en fait, en fin de compte, en réalité, finalement, somme toute

62	Espace	Exprime la propriété d'une entité par rapport à sa situation dans l'espace.	Localisation	à l'extérieur, à l'intérieur, ailleurs, au-delà, en bas, en haut, in situ, localement, nulle part, sur le terrain, ici
63	Espace	Renvoie à une unité textuelle.	Navigation_ textuelle	ci-dessous, ci-dessus, de là, ibid, infra, plus avant, précédemment 2
64	Temporalité	Caractérise une entité par rapport au temps : durée, fréquence, chronologie (localisation sur l'axe temporel).	Chronologie	à ce jour, à ce stade, à cette époque, à cette occasion, à court terme, à l'avance, à l'heure actuelle, à la fois, à long terme, à moyen terme, à nouveau, à présent, à terme_1, à tout moment, à un moment donné, actuellement, au cours du temps, au début, au départ, au même moment, au préalable, au quotidien, auparavant, d'abord_1, d'ores et déjà, dans la suite, dans le même temps, dans le passé, dans le présent, de nos jours, depuis longtemps, dès lors_1, désormais, du même coup, en amont, en même temps, en premier_2, en priorité, en temps réel, ensuite_2, finalement_2, initialement, jusqu'à maintenant, jusqu'à présent, jusqu'alors, jusque-là, jusqu'ici, par la suite, par le passé, peu après, plus que jamais, pour l'heure, pour l'instant, préalablement, précédemment, récemment, successivement, tantôt, tardivement, tout de suite, tout juste, déjà, puis
65	Manière	Permet d'insister sur l'exactitude d'un élément.	Précision	à la lettre, dans le détail, de plus près, en détail, en profondeur, essentiellement, exclusivement, plus en détail, précisément_1, spécifiquement, uniquement

Annexe 2

Les fréquences absolues des routines de guidage du lecteur à travers les disciplines.

	Sous-fonctions	Routines	Économie	Linguistique	Psychologie	Sociologie	Politique	Éducation	Information	Géographie	Anthropologie	Histoire
STRUCTURATION TEXTUELLE	Planification	N° 1 (la section 4 décrit)	57	1	2	2	3	2	3	4	0	0
		N° 2 (dans la section 2, nous présentons)	31	6	2	0	0	1	1	0	0	0
	Annonce de l'objectif/l'objet d'étude	N° 1 (l'objectif de cet article est d'analyser)	16	10	36	3	5	15	9	10	4	2
		N° 2 (l'article s'attache à montrer)	3	0	4	7	8	11	9	8	5	4
		N° 3 (cet article propose une étude)	2	0	2	1	0	0	0	1	3	1
		N° 4 (l'article s'intéresse à)	2	0	1	3	10	5	4	3	2	3
		N° 5 (nous proposons dans cet article d'analyser)	13	11	8	2	7	5	12	3	4	1
		N° 6 (nous nous intéressons à)	12	7	9	1	8	12	5	4	1	1
	Navigation textuelle	N° 1 (les résultats sont exposés dans)	40	5	17	2	4	7	2	0	1	0
		N° 2 (comme nous allons le voir)	52	53	24	37	63	46	37	17	51	34
		N° 3 (comme l'indique le tableau)	0	1	5	0	0	3	0	2	0	1
		N° 4 (je reviendrai sur ce point)	13	22	2	7	6	8	4	1	6	15
		N° 5 (il convient de rappeler)	2	2	2	0	8	0	1	1	2	3
		N° 6 (voir en annexe)	51	8	17	31	11	33	19	2	10	12
		N° 7 (voir l'article de quelqu'un)	47	22	19	70	72	30	24	17	47	180
N° 8 (on peut se référer à quelqu'un)		14	7	2	4	16	2	3	3	7	10	

	Conclusion	N° 1 (on peut conclure que)	10	10	5	6	4	2	4	0	1	3
		N° 2 (le résultat principal est de)	6	0	4	0	0	0	0	0	0	0
MÉTALINGUISTIQUE	Reformulation	N° 1 (on peut également dire)	1	5	1	2	1	4	1	0	4	5
		N° 2 (cela veut dire)	13	16	7	12	7	10	7	3	5	9
		N° 3 (pour le dire autrement)	0	0	0	2	4	0	2	0	3	0
	Réparation	N° 1 (on ne peut pas dire)	1	3	1	1	0	3	2	0	1	3
		N° 2 (cela ne veut pas dire)	3	9	1	6	5	6	2	5	7	7
	Précision terminologique	N° 1 (X peut être défini comme)	3	4	10	2	3	5	6	10	4	2
		N° 2 (on définit X comme)	7	6	3	1	0	4	3	1	1	0
		N° 3 (on appelle X Y)	5	33	12	19	12	16	17	8	18	10
		N° 4 (X est désigné par le mot Y)	2	5	2	6	0	1	2	1	1	2
		N° 5 (j'utilise le terme X pour désigner)	11	3	1	4	5	3	6	7	5	3
		N° 6 (le terme X renvoie à)	1	8	1	3	2	3	6	6	11	3
	Emprunt terminologique	N° 1 (on reprend le terme de quelqu'un)	3	3	1	1	3	0	3	0	1	2
N° 2 (pour reprendre le terme de quelqu'un)		1	7	3	10	14	6	10	8	8	11	
TOPICALISATION	Mise en évidence	N° 1 (nous pouvons mentionner)	3	10	8	13	10	10	8	10	7	9
		N° 2 (il est important de souligner)	5	0	5	4	2	3	1	4	2	3
		N° 3 (il faut noter que)	24	17	26	32	21	18	24	23	23	33
	(co)constat	N° 1 (nous pouvons noter que)	37	50	42	22	43	34	23	17	36	26
	Focus	N° 1 (le problème essentiel est)	5	7	3	11	3	3	3	4	3	4
		N° 2 (il est question de)	0	1	2	9	2	7	3	1	10	2

Résumé

Cette thèse s'intéresse à la question linguistique des routines de guidage du lecteur dans les écrits scientifiques en français. Ces routines de guidage du lecteur relèvent du phraséologisme scientifique transdisciplinaire, exercent des fonctions rhétoriques spécifiques de nature métadiscursive dans le discours scientifique. Leur présence ne vise pas à la transmission des contenus informatifs, mais à aider à faciliter la compréhension des contenus transmis lors de la communication. En nous basant sur des modèles métadiscursifs et à l'aide des techniques TAL, nous élaborons une typologie à l'accès onomasiologique et une modélisation linguistique pour les routines de guidage. Cette typologie servira à comprendre les stratégies rhétoriques généralement mobilisées dans les écrits scientifiques pour accompagner la lecture et fournira une façon efficace d'inventorier les routines de guidage. La modélisation met en place une description détaillée pour les routines de structuration. Nous montrons enfin que les routines de guidage permettent de comprendre l'aspect rhétorique des écrits scientifiques et la différence entre les disciplines.

Mots-clés : routines de guidage du lecteur, écrits scientifiques, fonction rhétorique, modélisation linguistique

Abstract

This thesis focuses on the reader guidance routines in scientific writing in French and their linguistic analysis. These reader guidance routines are part of the transdisciplinary scientific phraseology and exercise specific rhetorical functions of a metadiscursive nature in scientific discourse. Their presence is not aimed at the transmission of propositional contents, but at facilitating their understanding. Based on various metadiscursive models and by using NLP techniques, we develop a functional typology and a linguistic model specifically adapted for describing guidance routines. This typology is used to understand the rhetorical strategies generally found in scientific writing to help the reader, and provide an efficient way to make the inventory of reader guidance routines. The describing model provides a detailed analysis of guiding routines linguistic features. Finally, we show that guidance routines allow understanding the rhetorical aspect of scientific writing and their specificities across disciplines.

Keywords: reader guidance routines, scientific writing, rhetorical function, linguistic model