



HAL
open science

Analyse des relations entre concepts scientifiques et structure rhétorique dans des articles de recherche appartenant à un champ disciplinaire des Sciences de l'éducation

Salima Ait Saidi

► **To cite this version:**

Salima Ait Saidi. Analyse des relations entre concepts scientifiques et structure rhétorique dans des articles de recherche appartenant à un champ disciplinaire des Sciences de l'éducation. Linguistique. Université Grenoble Alpes, 2015. Français. NNT : 2015GREAL009 . tel-01400828

HAL Id: tel-01400828

<https://theses.hal.science/tel-01400828>

Submitted on 22 Nov 2016

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.

UNIVERSITÉ GRENOBLE ALPES

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Sciences du Langage

Présentée par

Salima AIT SAIDI

Thèse dirigée par **Francis GROSSMANN** et
codirigée par **Ágnes SÁNDOR**

préparée au sein du **Laboratoire LIDILEM**
dans l'**École Doctorale LLSH**

Analyse des relations entre concepts scientifiques et structure rhétorique dans des articles de recherche appartenant à un champ disciplinaire des Sciences de l'Éducation.

Thèse soutenue publiquement le « **2 décembre 2015** »
devant le jury composé de :

Madame Ingrid GOGOLIN,

Professeur des Universités, Université de Hambourg (Rapporteur et
Présidente du Jury).

Madame Amalia TODIRASCU

Maître de Conférence HDR, Université de Strasbourg (Rapporteur).

Madame Ágnes, SÁNDOR

Ingénieur de Recherche, Xerox Research Centre Europe (Co-encadrante).

Monsieur Francis GROSSMANN

Professeur des Universités, Université de Grenoble Alpes (Directeur de la
Thèse).



Remerciements

Cette thèse n'aurait pu voir le jour sans la contribution de plusieurs personnes à qui je voudrais exprimer toute ma gratitude.

Je tiens tout d'abord à remercier mon directeur de thèse Mr. Francis Grossmann pour le temps inestimable qu'il m'a consacré afin de mener à bien ce travail, pour avoir toujours répondu à mes nombreuses interrogations, pour son suivi régulier, ses précieux conseils, ses encouragements permanents et surtout pour m'avoir aidée à travailler en collaboration avec le centre de recherche européen XRCE et avec Ágnes Sándor.

Mes remerciements vont également à ma co-directrice de thèse, Ágnes Sándor, pour m'avoir chaleureusement accueillie au centre de recherche européen de Xerox (XRCE), pour sa patience, le temps qu'elle a consacré au suivi et corrections de ma thèse, pour m'avoir sans cesse aidée dans ma partie pratique et pour l'ouverture intellectuelle qui a caractérisé nos échanges.

Toute ma reconnaissance est à l'équipe du Lidilem de Grenoble, notamment au professeur Mme. Agnès Tutin pour ses précieuses corrections de mon article paru en 2015 qui ont donné un coup de pouce à l'avancement de ma thèse, aux doctorantes et amies Magda Florez et Chloé Dubois pour leur soutien et les nombreux moments d'échange qui m'ont aidée à avancer dans mon travail.

Je remercie infiniment Driss pour toute l'aide technique qu'il a pu m'apporter.

Je remercie également Djaffer pour ses relectures.

J'exprime toute ma reconnaissance à mon mari, Sofiane, pour m'avoir sans cesse soutenue, rassurée et encouragée dans les moments parfois difficiles qui ont jalonné le chemin de la thèse.

Mon plus grand remerciement revient à ma grand-mère ainsi qu'à mes parents que je chéris tant, sur qui j'ai toujours pu compter, qui m'ont toujours soutenue, encouragée dans tous mes choix et sans qui je n'aurai jamais pu vivre cette expérience extraordinaire en France.

A mon amie de cœur Amanda, pour son soutien infaillible depuis mon arrivée en France et pour avoir toujours su trouver les mots pour me remotiver dans les moments de doute.

A mes frères et ma sœur qui ont toujours répondu présents.

A toute l'équipe du Lidilem de Grenoble, à tous mes amis plus particulièrement Anne et Anne-Claire et collègues, et à tous ceux qui m'ont apporté leur aide, merci d'avoir contribué à ce que mon travail puisse être soutenu.

Notation

Une grande partie des abréviations et notations utilisées couramment dans la rédaction de cette présente thèse, est empruntée à celles utilisées au sein de l'analyseur robuste Xerox Incremental Parser (XIP).

FD = formules de discours

NP = clause nominale (noun phrase),

VP = clause verbale (verb phrase),

PP = clause prépositionnelle (prepositional phrase),

S = phrase (Sentence),

V-ing = verbe sous forme progressive,

Aux = auxiliaire,

V-tns = verbe transitif,

D = déterminant,

N = nom,

P = préposition,

BG= les séparateurs de propositions peuvent être insérés au niveau des éléments étiquetés « BG », qui marquent le début d'une clause,

FV= étiquettes relatives aux groupes verbaux finis,

IV= étiquettes relatives aux groupes verbaux infinitifs,

GV= étiquettes relatives aux groupes verbaux gérondifs,

COORD= Coordination

OP= opération mentale.

SUJ= Sujet

KS= Phrase-clé

KSDEP= Dépendance de phrase-clé

CHANG-COGNI= Changement cognitif

PUBL= Publication

DEICT= Déictique

OP_MENTCONTR= Opération mentale associée à la notion de contraste

AOS= Observable de l'activité scientifique

SLDM= Second Level Discourse Marker (marqueurs de discours de second niveau)

Résumé

Notre travail de recherche vise deux objectifs étroitement liés. Le premier consiste à proposer une aide à l'évaluation des écrits scientifiques et pour cause : le nombre de publications augmente, les limites entre les domaines deviennent floues, et il devient difficile de trouver des publications pertinentes si bien qu'un besoin pratique d'évaluation surgit. Il s'agit aussi de trouver les moyens d'une aide à l'expertise, appuyée sur des indices discursifs permettant d'aider le lecteur à repérer les points clés d'une publication (phase préalable à une évaluation) notamment à travers l'identification du problème de recherche.

Ces indices qui annoncent la formulation du problème de recherche dans les articles scientifiques sont repérables sous forme de « formules de discours ». Notre recherche ne s'étend pas sur la formulation de la problématique scientifique au vu de la complexité de cette notion et de la difficulté de la définir d'un point de vue de l'extraction d'information.

Nous proposons une modélisation de ces formules de discours que nous avons intégrée dans l'analyseur syntaxique Xerox Incremental Parser (XIP) sous forme de règles de reconnaissance. Nous avons utilisé un corpus d'articles de recherche en sciences de l'éducation extraits du corpus Scientext pour y détecter ces formules de discours. Le choix du domaine est motivé par ma participation au projet européen EERQI dont le but est de renforcer et d'améliorer la visibilité mondiale et la compétitivité de la recherche européenne en éducation.

Différentes approches méthodologiques ont été adoptées afin de procéder à une étude linguistique fine de ces formules dites de discours entre autres : l'analyse de discours (M. Pecman, 2004, K. Hyland, 2005, Á. Sándor, A. Kaplan, G. Randeau, 2006, D. Siepmann, 2007, A. Tutin, 2007-2010), robust parsing (S. Aït-Mokhtar, J. Chanod, R. Roux, 2002).

Il s'agit donc de mettre en œuvre une approche applicative en vue de l'aide à la lecture experte à travers l'identification, la typologie et le fonctionnement des associations lexicales véhiculant le problème de recherche.

Abstract

Our research focuses on two interrelated objectives. The first one aims at providing assistance to the evaluation of scientific writings because of; the number of publications which keeps on rising, the boundaries between the areas become blurred and it becomes difficult to find relevant publications so as the practical need for assessments is appearing as unavoidable. We also have to find new ways to help the evaluation of the research work, through a wide range of indications different from those usually used for research assessment, notably, through the identification of the research problem.

These indications that announce the formulation of the research problem in scientific articles can be identified as "speech formulas." Our research does not extend to the formulation of the scientific problem because of the complexity of this concept and the difficulty of defining it from the point of view of information extraction.

We propose a discourse modelling that we have integrated into the parser Xerox Incremental Parser (XIP) in pattern recognition rules. We used a corpus of Educational Sciences Research articles from the Scientext corpus to detect these speeches formulas. The choice of field is motivated by my participation in the European project EERQI which aims to strengthen and enhance the worldwide visibility and competitiveness of European research in education.

Different methodological approaches were adopted to perform a fine linguistic study of these formulas as: discourse analysis (Mr. Pecman, 2004, K. Hyland, 2005, Á, Sándor A. Kaplan, G. Randeau, 2006, D. Siepmann, 2007, A. Tutin, 2007-2010), robust parsing (S. Aït-Mokhtar, J. Chanod, R. Roux, 2002).

Thus, the goal is to implement an applicative approach aiming at helping expert reading through the identification, typology and functioning of lexical associations which convey the research problem.

Table des matières

REMERCIEMENTS	I
NOTATION	III
RESUME	V
ABSTRACT	VI
LISTE DES FIGURES	X
LISTE DES ANNEXES	XIV
INTRODUCTION GENERALE	1
CHAPITRE 1 : CADRE THEORIQUE DE L'ANALYSE DES ECRITS SCIENTIFIQUES	6
Introduction	6
1. La notion de problématique et de problème de recherche	6
1.1. Comment définir la problématique scientifique ?	6
1.1.1. Les composants d'une problématique	7
1.2. Généralité et écrit scientifique	10
1.3. Les différentes dimensions des écrits scientifiques	12
1.3.1 Dimension pragmatique	12
1.3.2. Dimension rhétorique des écrits scientifiques	13
1.3.3. Dimension énonciative	14
1.3.4. Dimension lexicale des écrits scientifiques	15
1.4. Langue Scientifique Générale (LSG)	23
1.4.1. Genre de l'article et typologie du discours scientifique	25
1.4.2. Genre de l'article de recherche	27
Conclusion	29
CHAPITRE 2 : FORMULES DE DISCOURS, COLLOCATIONS, MARQUEURS POLYLEXICAUX, MARQUES DU METADISOURS	31
Introduction	31
2.1. La notion de formule de discours	31

2.1.1. En quoi les formules de discours diffèrent-elles des collocations ?	32
2.2. Les marqueurs de discours polylexicaux	44
2.2.1. Paramètres définitoires des marqueurs polylexicaux	44
2.2.2. La réalisation syntaxique des marqueurs de discours	48
2.2.3. Fonctions discursives des marqueurs discursifs polylexicaux	50
2.3. Métadiscours	51
2.3.1. Le métadiscours : définitions et questionnements	52
2.3.2. Le contenu propositionnel vs. niveau métadiscursif	55
2.3.3. Modèles du métadiscours	59
2.3.4. Les facteurs qui influencent l'usage et la distribution des marques métadiscursives	71
2.3.5. Vers une définition des formules de discours	73
Conclusion	81
CHAPITRE 3 : METHODOLOGIE ET ANALYSE DE CORPUS	83
Introduction	83
3.1. Méthodologie et Concept-matching	83
3.1.1. Le Concept-matching	85
3.2. Développement de l'analyseur	88
3.2.1. Démarche suivie	91
3.2.2. Actions pour développer cette grammaire	93
3.2.3. Méthodologie d'intégration dans XIP et applications	95
3.3. Approche lexicale adoptée pour l'analyse des associations donnant lieu aux concepts constituants	98
3.4. Classification grammaticale des éléments lexicaux	99
3.5. Structure syntaxique des associations lexicales (AL)	116
3.6. Contenu conceptuel	118
3.6.1. La sphère de la scientificité	120
3.6.2. La sphère de la modalité	122
3.6.3. La sphère de la discursivité	124
3.7. Hiérarchie des concepts constituants	125
3.8. Structuration des phrases traduisant un problème de recherche	129
3.8.1. Structure de phrase déclarative	129
3.8.2. Structure de phrase interrogative	131

3.8.3. Structure de phrase impérative	131
3.9. Temps verbaux	132
3.9.1. Présent de l'indicatif	132
3.9.2. Le passé composé	133
3.9.3. Le futur	133
3.9.4. L'imparfait	134
3.10. Les marques de personne et de non-personne relevant de l'inscription de son auteur	134
Conclusion	137
CHAPITRE 4 : LIMITES & EVALUATION DU SYSTEME	139
4.1. Evaluation du système	139
4.2. Problèmes rencontrés	145
4.3. Evaluation du système	146
4.3.1. Comparaison avec une étude antérieure	146
CHAPITRE 5 : CONCLUSION GENERALE ET PERSPECTIVES	148
5.1. Conclusion	148
5.2. Perspective	149
5.2.1. Perspectives à court terme	150
5.2.2. Perspectives à long terme	151
REFERENCES	153

Liste des figures

Toutes les représentations (tableaux, graphiques, schémas) sont désignées comme figures, et numérotées. Les numéros correspondent à leur ordre d'apparition dans le texte.

FIGURE 1 : LE CYCLE DE LA RECHERCHE SELON O. AKTOUF (1987, 39)	8
FIGURE 2: LEXIQUE TRANSDISCIPLINAIRE QUI A POUR TRAIT OPERATION_MENTALE ET PORTEE.....	17
FIGURE 3: EXEMPLE DE LEXIQUE PROPRE AUX ECRITS SCIENTIFIQUES QUI A POUR TRAIT PORTEE EXTRAIT D'UN ARTICLE DE LINGUISTIQUE.....	17
FIGURE 4 : EXEMPLE DE LEXIQUE PROPRE AUX ECRITS SCIENTIFIQUES QUI A POUR TRAIT PORTEE EXTRAIT D'UN ARTICLE DE MEDECINE	17
FIGURE 5 : EXEMPLE DE LEXIQUE PROPRE AUX ECRITS SCIENTIFIQUES QUI A POUR TRAIT PUBLICATION EXTRAIT D'UN ARTICLE DE BIOLOGIE	18
FIGURE 6 : LEXIQUE ABSTRAIT NON-SPECIALISE	19
FIGURE 7: EXEMPLE DE LEXIQUE ABSTRAIT NON SPECIALISE QUI A POUR TRAIT CONTRASTE EXTRAIT D'UN ARTICLE DE TRAITEMENT AUTOMATIQUE DES LANGUES.....	19
FIGURE 8: EXEMPLE DE LEXIQUE ABSTRAIT NON SPECIALISE QUI A POUR TRAIT CONTRASTE EXTRAIT D'UN ARTICLE DE TRAITEMENT AUTOMATIQUE DES LANGUES.....	19
FIGURE 9: EXEMPLES DE LEXIQUE METHODOLOGIQUE DISCIPLINAIRE (A. TUTIN, 2007A)	20
FIGURE 10 : EXEMPLES DE LEXIQUE TERMINOLOGIQUE EXTRAIT DE NOTRE CORPUS.....	21
FIGURE 11: LEXIQUE DE LA LANGUE GENERALE/COMMUNE.....	21
FIGURE 12 : LEXIQUE DES ACTIVITES INTELLECTUELLES	22
FIGURE 13 : EXEMPLES DE LEXIQUE DES ACTIVITES INTELLECTUELLES ASSOCIE AU TRAIT OP_MENTALE	22
FIGURE 14: SPHERES NOTIONNELLES DE LA LSG (M. PECMAN, 2007)	24
FIGURE 15 : CONCEPTS CONSTITUANT LA LSG (M. PECMAN, 2007)	25
FIGURE 16 : TYPOLOGIE DE COMBINAISON DE MOTS SELON F.J. HAUSSMANN (1989) SYNTHETISEE PAR U. HEID (199, 553)	33
FIGURE 17 : EXPRESSIONS POLYLEXICALES/COLLOCATIONS A TRAIT SPECIFIQUE.....	39
FIGURE 18 : COLLOCATIONS EN DEPENDANCE SYNTAXIQUE AVEC D'AUTRES CONSTITUANTS DONNANT LIEU AUX FORMULES DE DISCOURS.....	40

FIGURE 19 : MARQUEURS DE DISCOURS POLYLEXCAUX DE SECOND NIVEAU (D. SIEPMANN, 2005)	47
FIGURE 20 : REALISATION LEXICO-GRAMMATICALES DES MARQUEURS DE DISCOURS EN FRANÇAIS	48
FIGURE 21 : MODELE DU METADISOURS DE A. CRISMORE (1989, 91).....	54
FIGURE 22 : L'APPROCHE LARGE DU METADISOURS	60
FIGURE 23 : LE MODELE INTERPERSONEL DU METADISOURSE (K. HYLAND, 2005, 49)	63
FIGURE 24 : SYSTEME DE CLASSIFICATION DU METADISOURS PAR W.J. VANDE KOPPLE (1985,82-92).....	64
FIGURE 25 : L'APPROCHE ETROITE DU METADISOURS	66
FIGURE 26 : MODELE REFLEXIF DU METADISOURS (A. MAURANEN, 1993).....	67
FIGURE 27 : ILLUSTRATION DU LEXIQUE DES CONCEPTS CONSTITUANTS	80
FIGURE 28 : MODELE RHETORIQUE DU METADISOURS DANS LES RESUMES DE BIOLOGIE (A. SANDOR 2007).....	86
FIGURE 29 : ANALYSE SYNTAXIQUE A TRAVERS LA CONCEPTUALISATION DES LEXIES ET ASSOCIATIONS LEXICALES.	87
FIGURE 30 : ARBRE DE CONSTITUANTS.....	90
FIGURE 31 : IMPLEMENTATION DU CONCEPT-MATCHING AU SEIN DE L'ANALYSEUR AUTOMATIQUE XEROX INCREMENTAL PARSER.....	91
FIGURE 32 : ANNOTATION SOUS XML DU CORPUS DE REFERENCE	92
FIGURE 33 : DEPENDANCES SYNTAXIQUES COMPRENANT LES MOTS D'UNE FORMULE DE DISCOURS FOURNIES PAR XIP	93
FIGURE 34 : DEVELOPPEMENT DU LEXIQUE DES CONCEPTS CONSTITUANTS.....	94
FIGURE 35 : CONJONCTIONS TYPIQUES DE CONCEPTS CONSTITUANTS LIES PAR DES DEPENDANCES SYNTAXIQUES QUI REALISENT LES DEUX FD QUI VEHICULENT LE PROBLEME DE RECHERCHE.....	96
FIGURE 36 : CLASSIFICATION TRADITIONNELLE FONDEE SUR LES CATEGORIES SYNTAXIQUES D'ELEMENTS LEXICAUX DE NATURE DIFFERENTE (M. PECMAN, 2004).....	100
FIGURE 37: CLASSIFICATION DES ADJECTIFS SUBJECTIFS PAR C. KERBRAT-ORECCHIONI (1980)	101
FIGURE 38 : CLASSIFICATION DES ADJECTIFS DES ECRITS SCIENTIFIQUES BASEE SUR LES TRAVAUX DE C. KERBRAT-ORECCHIONI (1980) & A.TUTIN (2008)	101

FIGURE 39 : CLASSIFICATION DES NOMS TRANSDISCIPLINAIRE INSPIREE DES TRAVAUX DE A.TUTIN ET G. BLOQUET, 2004	102
FIGURE 40: OCCURRENCE DES NOMS LOCALISATEURS IDENTIFIES DANS LES PHRASES-CLE	104
FIGURE 41 : OCCURRENCE DE NOMS CONSTRUITS ET COMPLEXES, ET CEUX PRODUIT PAR L'ACTIVITE SCIENTIFIQUE.....	105
FIGURE 42 : CLASSIFICATION DES VERBES DU LEXIQUE SCIENTIFIQUE INSPIREE DES TRAVAUX DE K. HYLAND (2005) & A.TUTIN (2010).....	106
FIGURE 43 : EXEMPLES DE VERBES DU METADISOURS INTERACTIONNEL EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS DU LEXIQUE SCIENTIFIQUE.	107
FIGURE 44: EXEMPLES DE VERBES DE POINT DE VUE QUI SE TROUVENT EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS DU LEXIQUE SCIENTIFIQUE	108
FIGURE 45: OCCURRENCE DES VERBES D'APPORTS SCIENTIFIQUE EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS LEXICAUX DU LEXIQUE SCIENTIFIQUE.....	109
FIGURE 46 : OCCURRENCE DU VERBE ILLUSTRE R EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS LEXICAUX DU LEXIQUE SCIENTIFIQUE.....	110
FIGURE 47: OCCURRENCE DU VERBE TENTER EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS LEXICAUX DU LEXIQUE SCIENTIFIQUE.....	111
FIGURE 48: OCCURRENCE DU VERBE METTRE EN EVIDENCE EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS LEXICAUX DU LEXIQUE SCIENTIFIQUE.....	112
FIGURE 49 : OCCURRENCE DU VERBE REPONDRE EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS LEXICAUX DU LEXIQUE SCIENTIFIQUE.....	113
FIGURE 50 : CLASSIFICATION DES ADVERBES DU LEXIQUE SCIENTIFIQUE INSPIREE DES TRAVAUX DE	114
FIGURE 51: OCCURRENCES DES ADVERBES DE MANIERE EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS DU LEXIQUE SCIENTIFIQUE	115
FIGURE 52 : OCCURRENCES DES ADVERBES D'ILLOCUTION ET DE JUGEMENT DANS LES PHRASES-CLE.....	116
FIGURE 53: ILLUSTRATIONS DE LA CLASSIFICATION DES AL SELON LE CRITERE SYNTAXIQUE	117
FIGURE 54: CLASSIFICATION DES AL SELON LE CRITERE CATEGORIEL	117
FIGURE 55 : CLASSIFICATION DES FD SELON LE CONTENU CONCEPTUEL	119
FIGURE 56: EXPRESSION DE LA POSSIBILITE	123

FIGURE 57: CONCEPTS CONSTITUANTS ET CORRESPONDANCE AVEC LES CATEGORIES D'AUTRES CLASSEMENTS.....	126
FIGURE 58 : FORMULES DE DISCOURS EXPRIMANT DIVERSES STRATEGIES DISCURSIVES QUI VEHICULENT DES CHANGEMENTS COGNITIFS.....	128
FIGURE 59 : EVALUATION DU SYSTEME	141
FIGURE 60 : EXTRACTION D'UN EXEMPLE DE PHRASE SURLIGNEE PAR LES EXPERTS ET IDENTIFIEE PAR XIP.....	142
FIGURE 61 : ILLUSTRATION DE PHRASES EXPRIMANT LE CONCEPT DE <i>RESUME</i> ET DE <i>CHANGEMENT COGNITIF</i> AU SEIN DE LA MEME PHRASE	143
FIGURE 62 : EXTRAIT D'UN ARTICLE ANALYSE : LES FORMULES DE DISCOURS SONT SURLIGNEES, EN ITALIQUES ET EN GRAS, LES PHRASES <i>RESUME</i> SONT SURLIGNEES EN GRIS CLAIR, ET LES PHRASES <i>CHANGEMENT COGNITIF</i> EN GRIS FONCE.....	144

Liste des annexes

Nous avons dressé une liste de documents dont nous nous sommes servis tout au long de la thèse proposés en annexes.

Les 51 articles analysés sous XIP sont consultables en Tome à disposition en Tome II de la thèse.

ANNEXES	162
ANNEXE 1.....	162
ANNEXE 1.1. EXEMPLES D'OCCURRENCES DE NOMS CONSIDERES COMME OBJETS FABRIQUES PHYSIQUE ET NON PHYSIQUE DONNENT LIEU AUX CONCEPTS DE RESUME ET DE CHANGEMENT COGNITIF	162
ANNEXE 1.2. EXEMPLES DE VERBES DE POINT DE VUE QUI SE TROUVENT EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS DU LEXIQUE SCIENTIFIQUE	170
ANNEXE 1.3. EXEMPLES DE VERBES DU METADISOURS INTERACTIONNEL EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS DU LEXIQUE SCIENTIFIQUE.	172
ANNEXE 1.4. OCCURRENCE DES VERBES D'APPORTS SCIENTIFIQUE EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS LEXICAUX DU LEXIQUE SCIENTIFIQUE.....	174
ANNEXE 1.5. OCCURRENCE DU VERBE TENTER EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS LEXICAUX DU LEXIQUE SCIENTIFIQUE.....	176
ANNEXE 1.6. OCCURRENCE DU VERBE METTRE EN EVIDENCE EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS LEXICAUX DU LEXIQUE SCIENTIFIQUE.....	178
ANNEXE 1.7. OCCURRENCE DE NOMS CONSTRUITS ET COMPLEXES, ET CEUX PRODUIT PAR L'ACTIVITE SCIENTIFIQUE.....	180
ANNEXE 2 : LISTE DE LEXIQUE.....	182
ANNEXE 3.....	190
ANNEXE 3.1 : W.J. VANDE KOPPLE'S CLASSIFICATION SYSTEM FOR METADISOURSE (1985,82-92).....	190
ANNEXE 3.2 : A.CRISMORE'S MODAL OF METADISOURSE (1989, 91).....	191
ANNEXE 3.3 : METADISOURSE CATEGORIZATION BY A. CRISMORE ET AL. (1993,47-54).....	192
ANNEXE 3.4 : AN INTERPERSONAL MODEL OF METADISOURSE (K. HYLAND, 2005, 49).....	193
ANNEXE 3.5 : PERSONAL AND IMPERSONAL CONFIGURATIONS OF METATEXT AND WRITER-READER INTERACTION (A. ADÈL, 2006, 38)	194

ANNEXE 4: ARTICLE TRANSMIS AUX EXPERTS POUR SURLIGNER LES PHRASES VEHICULANT UN PROBLEME DE RECHERCHE	195
ANNEXE 4.1. : PHRASES EXPRIMANT UN PROBLEME DE RECHERCHE SURLIGNEES PAR LES EXPERTS AU SEIN DU MEME ARTICLE DE RECHERCHE	211
ARTICLE 1	211
ARTICLE 2	211
ARTICLE 3	212
ARTICLE 4	214
ARTICLE 5	214
ARTICLE 6	215
ARTICLE 7	216
ANNEXE 4.2 : PHRASES EXPRIMANT UN PROBLEME DE RECHERCHE SURLIGNEES PAR LES EXPERTS AU SEIN DE 5 ARTICLES DE RECHERCHE DIFFERENTS.....	217
ARTICLE 1	217
ARTICLE 2	217
ARTICLE 3	217
ARTICLE 4	218
ARTICLE 5	219

Introduction générale

Cette thèse relève de l'analyse des relations entre concepts scientifiques et structure rhétorique dans des articles de recherche appartenant à un champ disciplinaire assez vaste et mouvant, celui des Sciences de l'Éducation. L'étude de ces relations est motivée par la construction d'un prototype permettant le repérage automatique des segments textuels à travers lesquels s'exprime la question de recherche que veut traiter un auteur scientifique dans un article, c'est-à-dire des éléments clés fournis par l'auteur scientifique pour définir la question traitée par sa recherche.

Le présent travail résulte de recherches menées au sein du Laboratoire de Linguistique des Langues Étrangères et Maternelles Appliquée (LIDILEM) de l'Université de Grenoble 3 - STENDHAL, dans le cadre du projet SCIENTEXT¹ en collaboration avec le centre de recherche européen Xerox Research Center Europe (XRCE) dans le cadre du projet EERQI (Educationnal research quality indicators). Le premier projet vise à la constitution d'un large corpus de textes scientifiques en français et en anglais comprenant chacun plusieurs millions de mots et étiquetés à l'aide de l'analyseur syntaxique SYNTAX². Les travaux du projet Scientext ont porté principalement sur le corpus en français. Ce projet est dédié à l'élaboration d'une ressource textuelle contemporaine facilement exploitable qui traite du positionnement et du raisonnement scientifique, à travers la phraséologie, les marques énonciatives et les marques syntaxiques liées entre autres au lexique exprimant la problématique scientifique. Le second projet se propose de fournir des indicateurs de qualité pour les recherches européennes en éducation dans le cadre entre autres de l'évaluation des écrits scientifiques.

Ce projet de recherche en sciences socio-économiques et humaines (SSH) a pour but de renforcer et d'améliorer la visibilité mondiale et la compétitivité de la recherche européenne en éducation.

¹ Projet ANR qui, initialement, a pour date de début janvier 2007 et se clôture en décembre 2010. Toutefois, celui-ci se prolonge aujourd'hui hors financement ANR. Parmi les ajouts en cours, il y a celui de nouvelles fonctionnalités notamment statistiques, ainsi que celui de nouveaux corpus.

² L'analyseur de dépendance Syntax a été développé par Didier D. Bourigault (2007). « *Il permet d'extraire d'un corpus des listes de syntagmes verbaux, nominaux, adjectivaux, structurées par des relations de dépendance syntaxique... Ces relations de dépendance permettent d'effectuer automatiquement des regroupements distributionnels : par exemple la liste de tous les compléments de tel verbe ou la liste des adjectifs modificateurs de tel nom, qui constituent des amorces de classes sémantiques.* » (D. Bourigault, 2000, 131-151).

L'objectif commun de ces projets, en ce qui concerne plus spécifiquement l'identification et la modélisation des associations lexicales exprimant un problème de recherche est double. D'une part, il consiste en une exploration des possibilités qu'offrent celles-ci pour le développement d'outils et moyens pour l'évaluation des écrits scientifiques et l'aide à la lecture experte. D'autre part, il s'agit également d'apporter des éléments de réponse aux grandes problématiques qui sont au cœur de notre champ de recherche : tout d'abord l'identification des formulations qui véhiculent les problèmes de recherche, ensuite leur organisation dans une typologie ainsi que l'analyse de leur fonctionnement.

S'inscrivant dans les domaines du TAL, de la linguistique de corpus et de la recherche documentaire, notre étude opère plus précisément dans la perspective d'une analyse rhétorique, conceptuelle et syntaxique de la structuration discursive des formulations discursives du problème de recherche. Plus spécifiquement, nous proposons une analyse syntaxico-conceptuelle des structures rhétoriques de ces associations lexicales.

Il est crucial, du point de vue de la recherche d'information mais aussi d'un point de vue linguistique, d'identifier les éléments à travers lesquels la problématique scientifique se manifeste. La notion de problématique, toutefois, reste un concept très complexe, peu facile à saisir et très difficile à définir d'un point de vue de l'extraction d'informations d'où le besoin d'une notion intermédiaire notion permettant plus facilement l'identification des structures rhétoriques qui nous intéressent.

C'est pourquoi, nous nous sommes focalisés sur la formulation de ce que nous préférons appeler, plus modestement, le « problème de recherche » en nous intéressant aux indices linguistiques qui permettent de le repérer. De manière conventionnelle, nous nommerons *formules de discours* ces indices, qui apparaissent sous forme d'associations lexicales au sein de la même phrase et qui se traduisent par exemple de la manière suivante :

....ce cadre théorique ... nous proposons, ... notre problématique... l'étude des questions....

Nos observations empiriques nous conduisent à penser que certaines associations spécifiques (les « *formules de discours* ») jouent le rôle de signaux permettant d'alerter le lecteur sur le fait que l'auteur scientifique est en train de formuler son problème de recherche.

Nous tenterons de montrer tout au long de cette étude en quoi les formules de discours, telles que nous tentons de les définir, divergent des associations lexicales étudiées jusqu' alors. En effet, en passant en revue plusieurs travaux portant sur les associations lexicales, nous nous sommes aperçus que les formules de discours ne s'identifient pas dans tous les cas de figure aux collocations, appelées également unités polylexicales ou phraséologiques, telles qu'elles sont définies notamment par M. Pecman (2004), A. Tutin (2007), car elles ne sont pas nécessairement contiguës, et les éléments qui les constituent ne se trouvent pas toujours en relation de dépendance syntaxique (voir partie 2.1). Elles divergent également parfois des autres ensembles phraséologiques notamment les marqueurs de métadiscours tels qu'ils sont définis par K. Hyland (2005) ou encore les marques de discours telles qu'elles sont définies par F. Boch (2010). Toutefois, elles s'en rapprochent sur certains points (voir partie 2.2).

Cette étude vise, d'une part, à mettre en œuvre une approche applicative en vue de l'aide à la lecture experte, d'autre part, à procéder à l'identification, la typologie et le fonctionnement des associations lexicales véhiculant le problème de recherche. Notons, que le second objectif dépend de la réussite du premier car si notre approche n'est pas pertinente nous ne pourrions pas proposer une méthode qui vise l'aide à la lecture experte.

Notre travail s'articule autour de cinq axes principaux.

Les chapitres 1 et 2 présentent le cadre théorique dans lequel a été mené notre travail tout en exposant la problématique de recherche. Les troisième et quatrième chapitres portent sur la méthodologie adoptée, l'analyse de notre corpus et le traitement de nos résultats. Le dernier chapitre porte sur la conclusion générale et les perspectives à court et long terme par rapport à notre sujet d'étude.

Dans le premier chapitre, nous nous attardons, avant toute chose, sur les notions de problème de recherche et de problématique scientifique dans le but de mieux délimiter notre sujet d'étude. Nous revenons brièvement sur les travaux qui ont porté sur cette question (O. Aktouf, 1987, B. Dionne, 1998, R. Tremblay et Y. Perrier 2006).

Puis, nous abordons les généralités liées à l'écrit scientifique. Nous nous intéressons aux différentes dimensions des écrits scientifiques, notamment à travers la question de la spécification des caractéristiques génériques de l'article et de celle des relations entre genre et typologie du discours scientifique.

Étant donné le fait que, même si notre corpus comporte des disciplines spécifiques, les formules discursives que nous cherchons à identifier sont pour beaucoup d'entre elles repérables dans la plupart des disciplines scientifiques, il nous a paru important également de nous intéresser à ce que l'on nomme la langue scientifique générale (LSG).

Dans le second chapitre, nous aborderons la notion de formules de discours telles que nous la définissons et nous préciserons en quoi celles-ci diffèrent des collocations, marqueurs polylexicaux et marques du métadiscours. Dans cette partie, nous allons procéder à une analyse de l'écrit scientifique en particulier dans une perspective linguistique. L'intérêt de cette analyse est de mettre en évidence les divergences entre des associations lexicales telles que les collocations, les marqueurs polylexicaux et les formules dites discursives. Il est également important de distinguer le métadiscours, notamment le lexique métadiscursif, du contenu asserté dans le discours, en trouvant des indices contextuels permettant la désambiguïsation. Rappelons que notre travail porte sur les marqueurs de discours véhiculant un problème de recherche à un méta-niveau. D'où l'importance de faire la distinction entre ces deux niveaux de discours ; le métadiscours et le contenu asserté. Nous partons de l'hypothèse selon laquelle la notion de métadiscours est très abstraite car il existe seulement des indices ou des marques, entre autres celles de métatexte³, les évaluatifs, les marques d'atténuation (*enclosures*⁴) ou les marques de prise de position qui permettent d'orienter le récepteur vers l'idée qu'une expression exprimant un problème de recherche, dans un contexte donné, joue un rôle métadiscursif. En effet, cette notion ne se laisse appréhender qu'à travers des marques qui n'ont pas, à proprement parler, de fonction informative, mais qui ont pour fonction de décrire le texte dans lequel il figure. Ainsi, à travers cette analyse nous tenterons, d'une part, d'identifier et de désambiguïser ces éléments qui permettent de signaler une fonction métadiscursive, d'autre part, de montrer en quoi ce concept peut s'associer à nos formules de discours.

La troisième partie sera consacrée à la méthodologie adoptée et au développement du système auquel nous avons fait appel afin d'identifier les associations lexicales traduisant un problème de recherche. Nous adopterons une analyse syntaxico-conceptuelle à l'aide

³ Qui n'ont pas de fonction informative, plus précisément celles des éléments lexicaux qui ont pour fonction de décrire le texte dans lequel il figure.

⁴ Le terme utilisé dans les études anglo-saxonnes pour qualifier ces marqueurs est *hedges*.

de la méthode du Concept-matching⁵ ce qui nous permettra de tester la validité de nos hypothèses de départ. Nous développerons l'approche lexicale adoptée pour l'analyse de notre liste de lexique des associations donnant lieu aux concepts constituants. Nous avons également porté un intérêt aux caractéristiques linguistiques des phrases-clé en abordant notamment les structures de ces phrases, les temps verbaux utilisés ainsi que l'inscription de l'auteur dans son texte.

Dans la quatrième partie, nous procéderons à l'évaluation de notre système à l'aide d'une comparaison entre une identification manuelle et une analyse automatique des phrases véhiculant le problème de recherche. En effet, nous avons transmis 15 articles de recherche qui ont pour domaine les sciences de l'éducation à des enseignants de l'Université de STENDHAL, Grenoble 3 qui pour la plupart sont des relecteurs notamment pour la revue LIDIL afin qu'ils puissent surligner les phrases qui traduisent le problème traité dans l'article. En parallèle, nous avons transmis un seul et même article à ces sujets avec la même consigne. L'objectif est de vérifier si notre système peut être pertinent dans une perspective d'aide à la lecture experte. Nous aborderons également les limites de notre système et proposons des solutions dans la partie 5.

En effet, nous recensons quelques-unes des difficultés sur lesquelles bute encore cette approche, d'où la nécessité de préciser rigoureusement les critères permettant d'identifier les formules de discours afin de résoudre les problèmes rencontrés.

Au sein de cette même partie, nous avons fait le choix de revenir sur une étude antérieure réalisée avec Á. Sándor (2015) où nous testons la pertinence de notre système en le comparant avec des résumés d'articles rédigés par des experts en sciences de l'éducation toujours dans le cadre du Projet européen EERQI. Nous avons tenté de voir dans quelle mesure les résumés étaient comparables à une ou plusieurs phrases donnée(s) figurant dans l'article.

En guise de bilan, nous résumons brièvement les apports de l'étude, et revenons sur les difficultés rencontrées lors de la mise en place de notre système d'identification de formules de discours exprimant les problèmes de recherche traités dans les articles de sciences de l'éducation et mettons en avant les perspectives dans le cadre de l'amélioration des problèmes rencontrés.

⁵ Le Concept-matching, développé par Á.Sándor (2007), permet d'effectuer une analyse syntaxique et conceptuelle. Nous revenons plus en détail dans le chapitre 3 sur la méthode du Concept-matching.

Chapitre 1 : Cadre théorique de l'analyse des écrits scientifiques

Introduction

Notre étude traite de l'analyse syntaxico-conceptuelle associée à la représentation conceptuelle des formules de discours traduisant le problème de recherche dans l'écrit scientifique. Rappelons que notre objectif est de parvenir à identifier automatiquement et à classer ces formules dites de discours dans l'écrit scientifique. Nous nous intéressons principalement à l'expression du problème de recherche dans l'écrit scientifique car il est au cœur de toute démarche scientifique.

Au début de cette étude dans le cadre du stage effectué au sein de XRCE, notre intérêt s'est porté sur l'expression de la « problématique » en général. Ce sujet a été soulevé par la construction d'un prototype pour l'identification de la problématique de recherche à travers les éléments clés fournis par l'auteur scientifique pour définir les orientations, hypothèses et questions traitées dans sa recherche. Toutefois, au fur et à mesure de l'évolution de notre travail, nous nous sommes aperçus que la notion de « problématique » est difficile à saisir ce qui portait à confusion au sein de cette étude. C'est pourquoi, nous avons opté pour une notion intermédiaire comme indiqué dans notre introduction. Rappelons que nous nous intéressons donc, plus modestement, à l'identification du « problème de recherche ».

Nous proposons dans ce qui suit de revenir sur la distinction entre les notions de problème de recherche et problématique scientifique, afin de mieux délimiter et présenter notre sujet d'étude. Nous ne prétendons en aucun cas établir un état de l'art exhaustif sur ces deux notions (qui serait un sujet d'étude à part entière), et renvoyons à plusieurs ouvrages pour approfondir ce domaine (O. Aktouf 1987, B. Dionne 1998, R. Tremblay et Y. Perrier 2006).

1. La notion de problématique et de problème de recherche

1.1. Comment définir la problématique scientifique ?

Selon R. Tremblay et Y. Perrier (2006) un problème de recherche appartient à une problématique particulière.

En effet, « *la problématique de recherche est l'exposé de l'ensemble des concepts, des théories, des questions, des méthodes, des hypothèses, et des références qui contribuent à clarifier et à développer le problème de recherche* » (R. Tremblay et Y. Perrier 2006).

La problématique est la mise en perspective de l'ensemble des liens qui existent entre les faits, les acteurs et les composantes d'un problème donné (B. Dionne, 1998). C'est un ensemble de questionnements, d'hypothèses et de problèmes envisagés dans une théorie. Elle s'élabore donc autour des questions principales traitées, d'hypothèses établies et des théories où elle s'inscrit.

1.1.1. Les composants d'une problématique

La problématique est composée d'un ensemble d'éléments, à commencer par le thème qui représente l'intitulé et le sujet sur lequel porte la recherche. Toute problématique traite d'un problème. Le problème trouve sa formulation appropriée à l'intérieur d'un champ de recherche à l'aide de concepts, de théories et de méthodes d'investigation qui lui sont propres et qui découlent de la difficulté et du manque de connaissance auquel fait face tout chercheur (R. Tremblay et Y. Perrier, 2006).

En effet, tout chercheur part d'un élément, d'un fait précis qui se présente sous forme d'une interrogation, ou d'une méconnaissance et qui se traduit par un problème à traiter.

« Ce point de départ sera, dans toute recherche convenablement menée, le point d'aboutissement sous l'aspect d'une nouvelle formulation, d'un progrès dans la connaissance du sujet traité... Cet aboutissement doit montrer clairement ce que le chercheur apporte de plus ou de nouveau à ce qui lui a servi de « problème » de départ. (O. Aktouf, 1987, 38) ».

La figure 1, tirée des travaux d'O. Aktouf (1987), reprend le cycle de recherche depuis la genèse du problème jusqu'au retour à ce même problème traité au sein de chaque recherche.

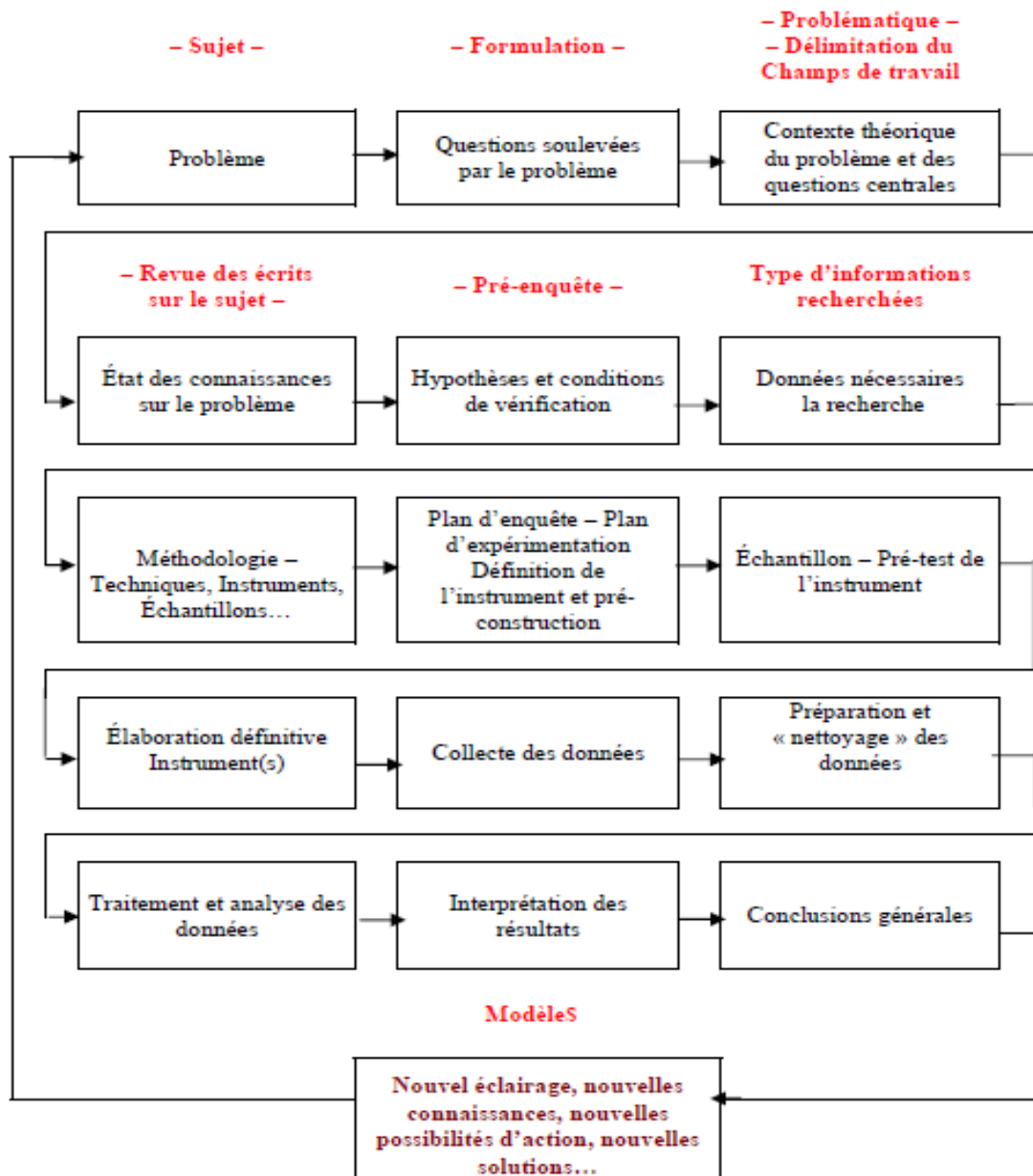


Figure 1 : Le cycle de la recherche selon O. Aktouf (1987, 39)

Comme nous pouvons le constater dans la figure 1, il y a une suite logique depuis le départ du problème jusqu'au retour à ce premier problème. Notre intérêt se porte plus précisément sur la formulation de ce problème dans l'article de recherche qui donne naissance à la problématique scientifique de chaque travail scientifique.

Il est difficile de ne poser qu'un seul type de problèmes tout au long de la recherche. En effet, ils peuvent être de différentes natures ; pratique, empirique, conceptuel et/ou théorique comme détaillé par R. Tremblay et Y. Perrier, 2006 dans ce qui suit.

- Le problème pratique relève d'une situation sociale difficile voire technique (conditions sociales des familles monoparentales) :
- Le problème empirique peut-être soulevé par un manque de connaissances des faits qu'une expérimentation ou observation peut permettre de résoudre (par ex. estimation de l'efficacité des politiques d'incitations financières à la fécondité) :
- Le problème conceptuel vise notamment à clarifier la définition exacte d'un concept (voir partie 3.6).
- Le problème théorique relève de l'explication d'un phénomène voire de l'évolution d'une théorie (explication des inégalités entre les hommes et les femmes).

Le problème se traduit donc par l'énonciation de ce qui intrigue le chercheur, ce qui se traduit notamment à travers la ou les questions de recherche qu'il pose. Ce sont ces questions qui entraînent la formulation des hypothèses du chercheur à travers un énoncé qui prédit les résultats.

Face à la difficulté que représente le concept de problématique qui reste très complexe et difficile à délimiter du point de vue l'extraction d'information, comparée à celle du problème traité dans la recherche scientifique, rappelons que nous avons opté pour une notion intermédiaire que nous convenons d'appeler tout au long de notre étude « *problème de recherche* ».

Selon nos observations empiriques, l'expression du problème de recherche se traduit par l'usage d'associations lexicales ;

- qui se trouvent être en relation syntaxique,
- qui ne sont pas nécessairement contiguës,
- qui interviennent à méta-niveau au sein de l'écrit scientifique en général et de l'article de recherche en particulier pour les besoins de cette présente étude,
- qui sont modélisées comme les réalisations des différentes configurations des concepts constituants⁶.

Ces associations lexicales qui véhiculent le problème de recherche dans l'écrit scientifique sont dénommées pour les besoins de la présente étude *formules de discours*.

⁶ Nous reviendrons plus en détail sur la notion de concepts constituants dans la partie 2.3.5.2 qui s'intitule « Lexique des concepts constituants ».

Nous formulons l'hypothèse selon laquelle ces formules discursives, auxquelles nous avons associé les fonctions de *résumé* et *changement cognitif* (voir section 2.1 pour une définition de ces deux fonctions) sont porteuses au sein des articles, de certains des contenus qui expriment le ou les problèmes de recherche qui composent la problématique scientifique en général.

Pour situer notre objet d'étude, les *formules de discours*, dans leur contexte linguistique, nous devons tout d'abord nous attacher à décrire le paysage dans lequel elles s'inscrivent.

Rappelons que notre corpus porte sur l'écrit scientifique. C'est pourquoi, nous nous intéressons à cet écrit en général et aux articles de recherche en particulier tout en décrivant les aspects linguistiques de la langue utilisée.

1.2. Généralité et écrit scientifique

J-M. Berthelot (2003), philosophe et sociologue, parmi bien d'autres, s'interroge sur ce qu'est un écrit scientifique. Il a été l'un des premiers à aborder le statut textuel de ce genre d'écrit. Il s'agit pour lui de recenser les différentes figures utilisées pour transmettre et explorer les connaissances, en montrant les fonctions que remplit le texte scientifique.

Ce genre d'écrit, destiné à un public spécifique, produit sous différentes formes, fait l'objet de pratiques sociales et de processus cognitifs. Il doit répondre à certaines exigences. Celles-ci seraient entourées par des codes et seraient soumises à des variations culturelles et disciplinaires tout en ayant en commun le système de la preuve et le rationalisme qui serait le dénominateur commun de tout écrit scientifique (I. Chavez, 2008, 13).

Signalons, toutefois, que la notion d'« écrit scientifique » est très générale et recouvre des sous-genres scientifiques et des pratiques disciplinaires très variées ; articles scientifiques, communications écrites, thèses ou mémoires d'habilitation à diriger des recherches.

Comme le souligne P. Bourdieu (1982), l'écrit scientifique est un « produit normé ». Le positivisme logique du début du 20^{ème} siècle contribue à faire du texte scientifique un genre destiné à un public particulier, garantissant la cohérence scientifique et linguistique du discours à travers des normes qui concernent aussi bien l'écriture (état de l'art,

référencement des travaux cités, etc.) que la manière d'administrer la preuve par la mise en scène des éléments d'observation ou de constat fournis au lecteur.

Les recherches sur le discours scientifique sont aujourd'hui en plein essor. De nombreux travaux ont porté sur l'analyse de discours des écrits scientifiques à travers la description linguistique ou dans une perspective rhétorique.

Parmi ces études, on peut citer :

- l'étude pionnière de B. Latour et Fabbri (1977) dont le travail a porté sur la rhétorique de la science ;
- celle de L.J. Prelli (1989) qui à l'issue de son travail considère le discours scientifique comme un discours tourné vers l'action, et qui se fonde sur le principe de coopération ;
- celle d'A. G. Gross (1996) qui applique les principes de la rhétorique à l'interprétation des textes scientifiques classiques et contemporains pour montrer la façon dont ils arrivent à persuader le lecteur. Il s'agit de mettre en évidence la façon dont les scientifiques - de Copernic à Darwin en passant par Newton et James Watson - établissent l'autorité et arrivent à convaincre les uns les autres et comment cette stratégie peut conduire à un remodelage de notre compréhension de la science et de sa place en société ;
- les études de K. Hyland (1998), K. Hyland (2002), K. Hyland (2005) avec des travaux portant notamment sur l'étude systématique du métadiscours scientifique ;
- celle de F. Grossmann et F. Rinck, 2004 dont la recherche porte sur la « surénonciation » comme norme du genre où il s'agit d'identifier les procédés énonciatifs par lesquels l'auteur scientifique construit une posture d'autorité, en montrant le caractère évident ou prouvé de telle ou telle assertion ;
- les contributions de K. Fløttum et al. (2006), et de F. Rinck (2006) qui soutiennent l'idée selon laquelle l'écrit scientifique est un texte argumentatif où la dimension rhétorique est très présente ;
- Les travaux de K. Fløttum (2007) qui portent sur le genre de l'article de recherche en français dans le domaine de la linguistique et qui compare des articles publiés par des doctorants et d'autres par des chercheurs plus expérimentés ;

- Ceux d'A. Tutin (2007) à travers l'analyse des noms transdisciplinaires des écrits scientifiques et l'étude sur les collocations dans l'écrit scientifique;
- L'étude de F. Grossmann & F. Wirth (2007) qui ont abordé la question de l'évidentialité dans les écrits de recherche à travers des marqueurs tels que *en fait* ;
- Les recherches de F. Rinck, et al. (2007) qui ont procédé à l'étude des marqueurs de convergence/divergence dans l'article scientifique.

Ces travaux abordent les différents aspects rhétoriques, énonciatifs ou encore lexicaux de l'écrit scientifique. Notre travail s'inscrit dans cette lignée et utilisera certains des résultats obtenus par ces recherches en les adaptant à notre propre objet d'étude.

Nous passerons en revue, les différentes dimensions des écrits scientifiques tout en nous concentrant plus particulièrement sur la dimension lexicale qui est centrale dans cette étude.

1.3. Les différentes dimensions des écrits scientifiques

1.3.1 Dimension pragmatique

L'écrit scientifique a fait l'objet de plusieurs études notamment à travers celles portant sur le genre emblématique qui est l'article. En effet, ce dernier a suscité l'intérêt de plusieurs chercheurs qui l'ont envisagé en tant que genre dans leurs travaux. Plusieurs aspects ont été abordés dans ces derniers, entre autres ; les styles individuels (U. Reutner), la variation disciplinaire (A. Tutin) et le statut de l'auteur (K. Fløttum et E. T. Vold).

Notre travail s'inscrit dans la continuité de ces travaux qui considèrent l'article comme produit présentant des propriétés linguistiques régulières et conventionnelles (F. Boch, et al. 2007). En effet, nous nous intéressons ici à cet instrument qui structure les interactions sociales au sein de la communauté du discours scientifique.

La dimension pragmatique est importante dans le cadre de cette étude car nous considérons que les FD expriment un problème de recherche sur un plan pragmatique, c'est-à-dire comme un élément mettant en évidence pour le lecteur le fait que la question de recherche va être explicitée dans l'énoncé. Nous reviendrons plus en détail sur ce point dans la partie 2.1 du chapitre 2 consacrée à la définition des FD.

1.3.2. Dimension rhétorique des écrits scientifiques

L'approche rhétorique du discours scientifique trouve son origine dans les travaux menés dans les années 1960 basés sur une revendication récurrente qui avait pour but de promouvoir une conception constructiviste de la science, autour de laquelle ont convergé des approches historiques, sociologiques et ethnographiques (F. Boch, 2010).

Les travaux portant sur cet aspect sont arrivés à la conclusion selon laquelle ces textes scientifiques visant à construire et à véhiculer des savoirs à travers des travaux de recherche ont recours à une forme de subjectivité car les auteurs de ces écrits, quelles que soient leurs disciplines, ont souvent recours à différentes stratégies d'écriture dans le but d'atteindre leurs objectifs.

A.G. Gross (2006), se fondant sur les catégories aristotéliennes, distingue deux types de stratégies rhétoriques auxquelles les auteurs font appel afin de renforcer leurs arguments : le *logos* et le *pathos*. Ces stratégies rhétoriques peuvent être trouvées dans de nombreux articles. Malgré le caractère objectif que les textes scientifiques doivent revêtir, la présence de ces stratégies rhétoriques traduit l'engagement subjectif de l'auteur dans ses écrits, surtout en ce qui concerne la seconde catégorie. Le *logos* est un appel à la logique grâce à l'utilisation de faits, y compris en mathématique, scientifiques ou statistiques des données pour appuyer un argument. Le *pathos* est un appel à l'émotion.

Ces deux stratégies rhétoriques sont notamment utilisées dans l'écrit de P. Piot et al. (1988) dans son article pour la revue *Scientific American* qui s'intitule⁷ : «Le sida : Une réponse mondiale ». Il tente de convaincre ses lecteurs que leur attention doit être réorientée afin de stopper la propagation du VIH dans les pays en voie de développement en s'appuyant sur des données statistiques. Ensuite, à travers le *pathos*, il communique un sentiment d'urgence à ses lecteurs et les implique personnellement dans un problème dont ils sont géographiquement détachés. On constate donc que ce genre d'écrit scientifique s'ouvre à des arguments subjectifs qui permettent à l'auteur d'infléchir son écrit afin d'atteindre ces objectifs.

Dans les travaux de K. Hyland (2004), d'autres stratégies rhétoriques utilisées différemment par les auteurs des écrits académiques et ceux des textes éditoriaux ont été mises en évidence notamment celles liées à la persuasion. L'analyse de ces textes a

⁷ "AIDS : A Global Response".

montré que chacun vise à persuader les lecteurs, en utilisant les marques du métadiscours différemment. L'usage des marqueurs d'évidentialité, dans les articles académiques a pour but de référer à des travaux antérieurs ou à des travaux dans le même domaine, tandis que dans les textes éditoriaux dont *Le Monde*, ces marques sont utilisées afin de mettre l'accent sur le sérieux, la crédibilité du journal, son élitisme ainsi que sa liberté d'esprit (K. Hyland, 2005, 87-89).

Le modèle du métadiscours introduit par K. Hyland (2005) qui consiste à impliquer le lecteur-potentiel dans le texte en exposant son point de vue explicitement et en tentant d'engager le lecteur-potentiel dans un dialogue en anticipant ces réponses vis-à-vis du texte met également en évidence la dimension rhétorique des écrits scientifiques. L'auteur fait appel à une stratégie rhétorique à travers l'usage de différentes marques du métadiscours, notamment aux « enclosures » et aux « emphatiques » qui ont pour but de renforcer la certitude et d'engager un dialogue proche, aux « marques d'attitudes » qui expriment l'attitude de l'auteur vis-à-vis de la proposition (*malheureusement, nous adhérons*), aux (auto)marqueurs personnels (*je, nous, mon, notre*), ainsi qu'aux marques d'engagement qui tente d'établir explicitement une relation avec le lecteur potentiel (*considérez, notez, vous pouvez constater/voir que*).

Les écrits scientifiques comportent donc une forme de subjectivité et une visée persuasive auxquelles les auteurs font appel afin d'atteindre leur objectif.

1.3.3. Dimension énonciative

Le concept d'énonciation a vu le jour dans les travaux de Benveniste dans les années 1970. Celui-ci considère l'énonciation comme une « *mise en fonctionnement de la langue par un acte individuel d'utilisation* » (É, Benveniste, 1970, 80). En effet, l'énoncé est le résultat d'un acte d'énonciation réalisé dans un contexte particulier. Le sens d'un énoncé dépend d'une interprétation globale et ne peut être saisi qu'à travers son contexte d'énonciation. L'approche énonciative envisage la communication linguistique de façon dynamique car elle envisage l'acte de communication, la prise de parole contrairement à l'approche référentielle. F. Jacques (1983) propose de dépasser la définition de Benveniste de l'énonciation en tant qu'acte individuel au profit « mise en communauté de l'énonciation », idée que rejoint également J. J. Authier-Revuz (1982), qui propose de qualifier d'hétérogénéité constitutive du discours ce double principe constitutif des

pratiques discursives que sont le dialogisme interdiscursif et interlocutif. Nous retrouvons une autre définition qui consiste à penser qu'« au lieu d'englober la totalité du parcours communicationnel, l'énonciation est alors définie comme le mécanisme d'engendrement d'un texte, le surgissement dans l'énoncé du sujet d'énonciation, l'insertion du locuteur au sein de sa parole.» (Ibid, 34). Cette insertion de l'énonciateur en qualité de sujet dans le discours implique un aspect qui peut parfois être subjectif c'est-à-dire qui a trait au sujet de l'énonciation. En effet, cet aspect de l'énonciation qui est la subjectivité et qui a trait au sujet de l'énonciation «...est la capacité du locuteur à se poser comme « sujet ». [...] elle n'est que l'émergence dans l'être d'une propriété fondamentale du langage. Est « ego » qui dit « ego ». Nous trouvons là le fondement de la subjectivité qui se détermine par le statut linguistique de la personne.» (É. Benveniste, 1970).

D'autres aspects énonciatifs des écrits scientifiques ont été abordés et ont fait l'objet de plusieurs études. C'est notamment le cas de la polyphonie (H. Nølke, K. Fløttum. C. Noren, 2000) cité in F. Boch (2010), ou encore les études réalisées dans le cadre du projet KIAP⁸ (K. Fløttum et al. 2006 ; 2007) qui a comparé des écrits dans plusieurs langues et plusieurs disciplines et celle du projet Scientext à travers le positionnement énonciatif.

Notons que notre intérêt se porte plutôt sur d'autres dimensions – à savoir principalement la dimension lexicale, ce qui ne nous empêchera pas de prendre en compte ces aspects énonciatifs. En effet, les *formules de discours* telles que nous les abordons dans ce présent travail sont directement liées à la dimension énonciative à travers les déictiques en particulier. Cette unité linguistique qui est inséparable du lieu, temps et sujet de l'énonciation contribue à identifier nos phrases-clé représentées par les concepts généraux de *changement cognitif* et *résumé* dès lors où celles-ci se trouvent en relation syntaxique avec d'autres éléments du lexique scientifique.

1.3.4. Dimension lexicale des écrits scientifiques

A. Tutin (2007b, 5) considère qu'une des caractéristiques de ce genre d'écrits est de posséder un lexique partagé par les différentes disciplines, d'une part, parce que ces écrits ont en commun des procédures et des objectifs, d'autre part, parce qu'ils renvoient

⁸ Voir le site de l'équipe : <http://kia.uib.no/>

à des genres codifiés par des communautés de discours. Les écrits scientifiques, quelles que soit leurs disciplines de référence, mobiliseraient alors plusieurs types de lexiques, allant du lexique général au lexique terminologique (A. Tutin, 2007b, 6-7).

Toutefois, le lexique propre aux écrits scientifiques reste difficile à délimiter. En effet, certains termes se rapprochent plutôt du lexique général alors que d'autres sont propres à certaines disciplines (C. Cavalla & F. Grossmann, 2005). Ce point a été soulevé dans l'article publié dans la *Revue de linguistique appliquée* par A. Tutin (2007).

L'auteur y fait une distinction entre différents types de lexique qu'elle caractérise comme des 'ensembles lexicaux'. La catégorisation établie dans la partie 1.3.4.1. illustrée par des exemples extraits de notre corpus, est inspirée et basée sur cet article.

1.3.4.1. Le lexique propre aux écrits scientifiques

Ce type de lexique, appelé lexique transdisciplinaire, également appelé transversal par A. Tutin (2010) est commun à tous les écrits scientifiques,

« il renvoie aux procédures ou aux concepts génériques de l'activité scientifique, probablement présentés dans nombre d'écrits du même type et dans des disciplines variées » (A. Tutin, 2010, 150).

En effet, il renvoie au discours sur les objets et les procédures scientifiques et non aux objets spécifiques des domaines de spécialité. Ce lexique est spécialisé dans la description de l'activité et de la réflexion scientifique dans son ensemble.

Il est utilisé dans notre corpus de sciences de l'éducation comme nous pouvons le voir ci-dessous. Nous lui avons attribué les traits OPERATION_MENTALE et PORTEE. Nous revenons plus en détail dans la partie 3 sur ce que nous entendons par *traits* et sur ce qui justifie nos appellations de PORTEE, OP_MENTALE au sein de ce présent travail.

Processus de l'activité scientifique	noms extensifs	analyse, travail, projet, recherche, etc.
Objet construit par l'activité scientifique (OAS)	noms non-extensifs	approche, modèle, système, théorie, idée, etc.
Observable de l'activité scientifique	noms examinés par l'activité scientifiques	données, résultat, point, etc.
Noms supports de la rédaction scientifique	Concrets et abstraits non-extensif (ne se combinent pas avec <i>lors</i> , <i>durant</i> , etc.)	articles, texte, documents, etc.

Figure 2: Lexique transdisciplinaire qui a pour trait OPERATION_MENTALE et PORTEE

La figure 2 propose une illustration des noms de notre lexique des écrits scientifiques et de leur classification: associés à d'autres éléments linguistiques, ils donnent lieu a des FD dites de *résumé* et de *changement cognitif*.

Quel que soit le domaine de recherche considéré, l'écrit scientifique a en partage ce type de phraséologie. Illustrons notre propos par quelques exemples extraits de différents domaines du corpus Scientext.

Quel que soit le domaine de recherche, l'écrit scientifique partage cette phraséologie. Le rôle du corpus est enfin fondamental dans notre **approche** parce que le corpus est non seulement l'objet du traitement pour l'analyseur, et le lieu d'interprétation pour l'analyste, mais c'est aussi une source d'information pour l'acquisition par l'analyseur d'informations de complémentation, comme nous allons le montrer dans la section suivante.

Figure 3: Exemple de lexique propre aux écrits scientifiques qui a pour trait PORTEE extrait d'un article de Linguistique

Ces **résultats** ont été confirmés en utilisant des fibroblastes isolés de souris dans lesquelles le gène RIG-I était inactivé [7]

Figure 4 : Exemple de lexique propre aux écrits scientifiques qui a pour trait PORTEE extrait d'un article de Médecine

L'objectif de cet **article** est de montrer le formidable potentiel des puces à ADN dans ce domaine, mais aussi d'analyser les raisons pour lesquelles cette technologie tarde à s'implanter dans les laboratoires d'analyse microbiologique.

Figure 5 : Exemple de lexique propre aux écrits scientifiques qui a pour trait PUBLICATION
extrait d'un article de Biologie

Comme nous pouvons le constater à travers les trois figures ci-dessus, cette phraséologie est représentée le plus fréquemment par des noms extensifs, non-extensifs⁹, examinés et traités par l'activité scientifique et concrets et abstraits non-extensifs est commune à tous les écrits scientifiques quelles que soient les disciplines. Prenons pour exemple *approche*, ce nom incarne un nom d'objet construit par l'activité scientifique. Il n'est pas extensif et a un agent humain. L'occurrence *article* incarne un nom de support de la rédaction scientifique, nom qui est à la fois concrets et abstraits non extensifs.

1.3.4.2. Le lexique abstrait non spécialisé

Comme son nom l'indique, ce lexique n'est pas exclusif de l'écrit scientifique. En effet, il se trouve dans divers autres genres d'écrits tels l'écrit littéraire, informatif ou argumentatif (A. Tutin, 2010). Ce lexique abstrait non spécialisé relève d'un méta-niveau et n'est pas associé à un genre ou à une discipline en particulier.

Dans la mesure où nos formules de discours mobilisent également un lexique non spécialisé, relevant d'un méta-niveau, cette catégorie nous semble particulièrement intéressante pour notre recherche.

Nous pouvons voir à travers ce qui suit une illustration du lexique relevant de ce niveau l'abstrait, auquel nous avons attribué entre autres le trait CONTRASTE. Cette appellation est justifiée, comme nous le verrons dans la suite (voir section), par le fait que ce trait est associé à tous les éléments lexicaux qui traduisent le questionnement initial et les interrogations du chercheur.

⁹ Les noms extensifs qui sont associés au processus de l'activité scientifique se combinent avec *lors*, *durant*, des verbes phasiques, souvent avec *faire*.

Les noms non-extensifs sont des noms concrets et/ou abstraits avec un agent humain. Ils se combinent avec des verbes comme élaborer, construire. Parmi les noms non-extensifs de l'écrit scientifique, nous retrouvons ; Les noms d'observables de l'activité scientifique (cas, données, facteur, etc.), les noms de supports de la rédaction scientifique (article, chapitre, figure, etc.).

Nous revenons plus en détail dans le chapitre trois sur la définition et l'attribution de ce trait qui, associé à d'autres éléments lexicaux, donne lieu au concept général de *changement cognitif*.

<p>Nom exprimant une problématique (également OAS)</p>	<ul style="list-style-type: none"> • noms examinés par l'activité scientifiques • verbes exprimant une opinion négative 	<ul style="list-style-type: none"> - question, hypothèse, etc. - contester, rejeter, etc.)
---	---	--

Figure 6 : Lexique abstrait non-spécialisé

Cette question pose bien entendu le problème classique en TAL des mots hors vocabulaire, mais peut concerner, tout simplement, la fréquence d'usage des mots spécifiques aux productions de l'utilisateur

Figure 7 : Exemple de lexique abstrait non spécialisé qui a pour trait CONTRASTE extrait d'un article de traitement automatique des langues

Nous relevons dans les figures qui suivent quelques exemples tirés de la base de données Scientext de l'occurrence de ce lexique abstrait non spécialisé extraits d'articles de traitement automatique des langues.

Ce niveau correspond plus précisément à la probabilité d'erreur quant au fait de rejeter l'hypothèse stipulant que la différence entre les résultats testés n'est pas significative.

Figure 8 : Exemple de lexique abstrait non spécialisé qui a pour trait CONTRASTE extrait d'un article de traitement automatique des langues

Comme nous pouvons le constater dans les exemples précédents, le lexique abstrait non spécialisé intervient à un méta-niveau et il est utilisé dans plusieurs genres d'écrits autres que l'écrit scientifique.

1.3.4.3. Le lexique méthodologique disciplinaire

Le lexique méthodologique disciplinaire « renvoie aux procédures scientifiques d'un ensemble de sciences » telles les sciences appliquées ou encore les sciences dures ou

expérimentales. Ainsi, la phraséologie, dont ce lexique est extrait, peut être considérée comme étant disciplinaire ou relevant d'une famille de disciplines (A. Tutin, 2007a, 3).

Ci-dessous un exemple de texte extrait de cet article de A. Tutin (2007a) qui illustre les occurrences de lexique méthodologique disciplinaire.

L'évaluation du cout de l'enfant en coupe instantanée pose le problème de l'hétérogénéité des préférences entre les ménages selon qu'ils ont ou non des enfants. La difficulté théorique de réaliser des comparaisons de bien-être interpersonnelles de joue cependant plus lorsque l'on dispose de données longitudinales. Celle-ci permet une observation directe de l'influence de l'arrivée ou du départ des enfants sur le niveau de bien-être tel qu'il est perçu par les ménages. Le panel européen fournit ainsi des éléments subjectifs sur l'aisance financière des agents, mais sous forme de réponse polytomique.

Figure 9: Exemples de lexique méthodologique disciplinaire (A. Tutin, 2007a)

Nous écartons cette catégorie car rappelons que nous nous intéressons uniquement au lexique relevons d'un méta-niveau indépendamment de la discipline abordée. En effet, Il pourrait exister une catégorie de lexique méthodologique transdisciplinaire (par exemple « méthode », « démarche méthodologique », « données », etc.), mais ce lexique là non plus ne semble pas correspondre au lexique recensé, dans la mesure où il ne relève pas d'un méta-niveau.

1.3.4.4. Le lexique terminologique

Avant d'aborder le lexique terminologique, il convient de revenir sur ce que nous entendons par terminologie. M. Le Guern (2003) fournit une précision utile sur ce qu'est le lexique (qui considère les mots indépendamment des choses) et sur la terminologie, ces « *termes de toutes les sciences et des arts, savoir la philosophie, logique, et physique (...)* », où les mots sont liés aux choses. Ainsi, le lexique terminologique diffère d'un domaine à un autre étant donné qu'il traite un lexique spécifique qui fait l'objet de l'étude traitée.

Etude des relations entre les stratégies d'adaptation aux sentiments dépressifs, la symptomatologie dépressive et les idées suicidaires chez l'adolescent.

Figure 10 : Exemples de lexique terminologique extrait de notre corpus

Nous ne nous intéressons pas à ce lexique spécifique dans le cadre de cette étude car il est propre à une discipline en particulier et ne relève pas non plus d'un méta-niveau.

1.3.4.5. *Le lexique de la langue générale/commune*

Ce genre de lexique, également appelé lexique de base, n'est propre à aucune discipline ni genre en particulier. Il est peu spécialisé et intègre des lexies qui ont une fonction grammaticale. Parmi les traits que nous avons attribués à ce lexique, nous retrouvons DEICTIQUE et CONTRASTE. Ce dernier représente les questionnements et interrogations initiaux du chercheur. En effet, comme nous pouvons le constater à travers la figure suivante, ce lexique est utilisé par les auteurs quel que soit la discipline dans la construction, et enchaînement des idées véhiculées.

Pronoms personnels		Je, nous
Adjectifs démonstratifs		ce, cet, etc.
Conjonction de coordination	de	Mais, cependant

Figure 11: Lexique de la langue générale/commune

Nous nous intéressons plus particulièrement à cette catégorie et la prenons en compte dans notre liste de lexique des écrits scientifiques car elle peut relever d'un méta-niveau et traverse les disciplines. Elle est donc susceptible de donner lieu à nos concepts généraux de *résumé* et *changement cognitif* si les termes considérés se trouvent en relation syntaxique avec d'autres éléments linguistiques de notre lexique scientifique.

1.3.4.6. *Le lexique des activités intellectuelles*

M. Pecman (2007, 83) introduit une catégorie supplémentaire permettant de caractériser un certain type de lexique, celui qui « renvoie aux activités intellectuelles ». Selon elle,

ce type de lexique « *sous-tend le raisonnement scientifique* » et traverse les genres et les disciplines. Dans la mesure où nous nous intéressons à la « question de recherche », qui mobilise des formes de raisonnement, il nous semble particulièrement important pour notre sujet, et nous l'avons recensé en lui attribuant le trait conceptuel OPERATION_MENTALE. Nous reviendrons plus en détail dans le chapitre 3 sur les définitions de ces traits.

La figure ci-dessus illustre quelques exemples extraits de notre corpus auxquels nous avons attribué le trait OPERATION_MENTALE.

Noms et verbes qui renvoient à une opération mentale de réflexion/raisonnement	Noms et verbes du champ sémantique de l'étude	<i>analyser, étudier, etc.</i>
---	--	--------------------------------

Figure 12 : Lexique des activités intellectuelles

Comme observé dans la figure 12, cette catégorie de verbes et de noms, qui renvoient à une opération mentale de réflexion et ou de raisonnement est représentée par le trait OP_MENTALE.

Etudier le problème de la modélisation de relations conceptuelles à partir de l'observation de relations lexicales en corpus permet d'aborder toutes les facettes de la sémantique de ces relations.

Figure 13 : Exemples de lexique des activités intellectuelles associé au trait OP_MENTALE

L'illustration précédente tirée d'un article de linguistique de la base de données Scientext met en évidence un exemple de phrase où apparaît le lexique des activités intellectuelles qui est au centre de notre étude.

Outre la réponse aux besoins en didactique des langues étrangères qui a motivé de nombreux travaux sur la question du lexique propres aux écrits scientifiques (Phal 1971, M. Pecman 2004, Coxhead & Hirsh 2007), la délimitation de ce lexique permet également de mieux comprendre les spécificités du discours scientifique, et plus particulièrement la langue codifiée, à lexique divergent¹⁰ et spécifique, liée à l'expression du problème de recherche, qui est au centre de notre étude.

¹⁰ Comme nous avons pu le constater, l'écrit scientifique a plusieurs lexiques à son actif hormis le lexique spécifique.

1.4. Langue Scientifique Générale (LSG)

La notion de LSG se fonde sur le postulat d'une langue commune partagée par la communauté scientifique, quelle que soit la discipline.

Les études menées sur la LSG se situent dans la lignée des travaux menés sur le *Vocabulaire d'Orientation Général Scientifique* (VOGS).

A. Phal (1971) est l'auteur des premiers travaux, sur le VOGS, relatifs à la langue écrite dans les domaines de mathématique, physique, chimie et langue naturelle (P. Drouin, 2007, 46). Il définit ce vocabulaire spécifique de la manière suivante :

« Le vocabulaire scientifique général est (...) commun à toutes les spécialités. Il sert à exprimer les notions élémentaires dont elles ont toutes également besoin (mesure, poids, rapport, vitesse, etc.) et les opérations intellectuelles que suppose toute démarche méthodique de la pensée (hypothèse, mise en relation, déduction et induction, etc.) ».

Quant au concept de LSG, il est défini par M. Pecman (2004, 136) comme une :

« pratique langagière spécifique à une communauté de discours composée de chercheurs en sciences exactes dont les objectifs communicatifs poursuivis émanent des préoccupations partagées par des scientifiques à travers le monde et indépendamment de leurs spécificités disciplinaires ».

Pour appuyer sa définition, M. Pecman donne l'exemple de l'anglais, langue partagée par la communauté linguistique, dont l'emploi répond aux besoins communicatifs de cette communauté. Soulignons un aspect important du fait que l'anglais n'est pas seulement la langue d'une communauté scientifique mais également la langue de communication privilégiée en sciences (au moins dans les sciences « dures »).

M. Pecman (2004,8) considère également que la Langue Scientifique Générale (LSG) comporte un ensemble de formules préfabriquées qui véhiculent des notions fondamentales et des raisonnements communs aux différentes sciences exactes (chimie, physique, biologie, mathématique, informatique, etc.). Ses travaux (2004, 2007) ont porté sur les caractéristiques de la LSG afin de rendre compte du panorama linguistique de cette langue à travers un corpus bi-textuel recueillant des documents scientifiques de trois domaines connexes (chimie, physique et biologie) qui a été mis en place. Le

dépouillement de ce corpus a permis de collecter et de modéliser les FD selon une méthode dont nous nous sommes partiellement inspirés pour la présente étude.

Il a également permis de dégager quatre grandes sphères notionnelles dans lesquelles sont classées l'ensemble des associations préfabriquées : les sphères de la scientificité, de l'universalité, de la modalité et de la discursivité.

Sphères	NOTIONS
Scientificité	EXPERIMENTATION, VARIATION, OBSERVATION, EVALUATION, TRANSFORMATION
Universalité	CONSEQUENCE, TEMPORALITE, QUANTITE
Modalité	NECESSITE, PROBABILITE, FAISABILITE, OPINION
Discursivité	POSITION, TRANSITION, PRESENTATION

Figure 14: Sphères notionnelles de la LSG (M. Pecman, 2007)

M. Pecman (2004, 293-294) décrit ces quatre sphères notionnelles comme suit :

- La sphère de la scientificité regroupe toutes les notions de nature purement scientifique, qui ont une valeur spécifique dans le domaine scientifique ou encore qui se réfèrent à une action, un objet, une entité, une abstraction, une qualité, etc.,
- La sphère de l'universalité regroupe des notions dont l'usage est significatif dans le discours scientifique et qui est plus large que celles représentées dans la LSG
- La sphère de la modalité est représentée par les unités phraséologiques qui expriment un avis, un jugement ou une appréciation de la part de l'auteur à l'égard du contenu de son discours. Parmi les notions recensées, nous retrouvons la modalité épistémique (cf. |PROBABILITE|), la modalité déontique (cf. |NECESSITE|), ou encore l'expression de la faisabilité |FAISABILITE|, du point de vue appréciatif ou dépréciatif de l'auteur |OPINION|, etc.
- La sphère de la discursivité fait référence à tous les aspects argumentatifs, tels que la présentation du sujet d'étude (cf. |PRESENTATION|), le positionnement de l'auteur (cf. |POSITIONNEMENT|), l'introduction des exemples (cf. |EXEMPLE|), etc.

La figure suivante reprend ces différentes sphères introduite par M. Pecmann (2004). Cette représentation est importante dans le cadre de cette étude car nous en avons besoin dans l'analyse des FD dans le chapitre 3.

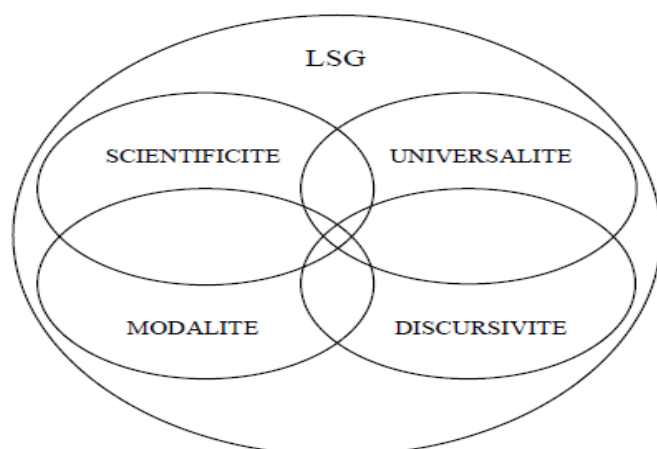


Figure 15 : Concepts constituant la LSG (M. Pecman, 2007)

Nos travaux s'inscrivent dans la lignée des travaux de M. Pecman (2004) qui portent donc sur le lexique et la phraséologie des écrits scientifiques, plus particulièrement, sur la *langue scientifique générale* (LSG) qui est caractérisée par un *lexique transdisciplinaire* (A. Tutin 2007) entre autres. Toutefois, notre objet d'étude est plus large car les FD telles que nous les définissons dans le présent travail, appartiennent à des catégories ne faisant pas partie du lexique transdisciplinaire, comme nous pourrons le constater plus loin.

Nous recenserons donc le lexique propre à cette langue scientifique au sein des articles de recherche, constitué de texte relevant du domaine des sciences de l'éducation. En effet, nous partons de l'hypothèse que dans le discours scientifique, les auteurs emploient des associations lexicales donnant lieu à des *formules de discours* véhiculant les problèmes traités dans ces écrits. Celles-ci seraient employées pour traduire un problème de recherche et leur identification permettrait d'aider à la lecture experte. Toutefois, ces formules spécifiques relèveraient d'un méta-niveau. Elles ne seraient donc pas associées à un genre en particulier ou à une discipline donnée.

1.4.1. Genre de l'article et typologie du discours scientifique

L'article décrit la recherche d'une personne ou d'une équipe de recherche - comme c'est notamment le cas dans les sciences expérimentales - effectuée sur un sujet particulier. Il

inclut généralement un état de l'art sur les recherches accomplies. Quel que soit le sujet traité, l'article de recherche contient des faits, des données et des explications sur une recherche donnée.

Tout article de recherche s'appuie « sur des connaissances préalablement admises, sur des principes reconnus, sur des faits évidents. Il faut dire sur quoi nous nous basons, manifester la valeur et la pertinence de cette source et montrer en quoi elle éclaire l'énoncé en question » (Thibaudeau, 1997, 320) afin de rendre compte des idées véhiculées. Ainsi, afin d'appuyer ses propos notamment pour formuler son problème de recherche, l'auteur recourt à des procédés variés : explication, réfutation, discussion, démonstration, justification, démonstration, comparaison, etc.

Les articles de recherche sont la principale source d'information scientifique des chercheurs.

Depuis l'institutionnalisation de la science, le genre de l'article de recherche est considéré comme étant le plus étudié en raison de son statut emblématique dans l'activité scientifique (C. Bazerman, 1988). En se référant à la distinction de Maingueneau, F. Rinck (2007,33) considère que l'article de recherche

« correspond à un genre institutionnalisé, en raison de son inscription dans le champ scientifique, et notamment du dispositif éditorial impliqué dans sa production..... les genres institutionnalisés se caractérisent par un ensemble de routines, ce sur quoi s'accordent un certain nombre d'auteurs à propos du discours scientifique ; comme le montre D. Maingueneau, la routinisation concerne en premier chef les rôles discursifs, qui dans ces genres sont pré-établis, très codifiés, et, partant, relativement stables, à la différence de ce qui se joue dans le cadre de genres conversationnels » .

L'article est l'indicateur majeur de la production des chercheurs et de la visibilité de leurs travaux.

L'essentiel d'un article scientifique peut se résumer par ses buts, problèmes et conclusions. Ces éléments remplissent au sein des articles des fonctions rhétoriques que les auteurs peuvent formuler, entre autres, par des associations lexicales véhiculant le problème de recherche scientifique. Prenons pour exemple l'association lexicale *on conclut que* telle qu'elle est étudiée par T. Dahl (2004, 1811-1812) : elle permet au lecteur potentiel d'identifier un des actes rhétoriques de l'auteur dans son processus

d'argumentation. En effet, dans l'exemple précédent l'usage d'un marqueur de cadre utilisé pour rendre explicite l'acte discursif et fait référence au niveau du texte.

Plusieurs travaux ont traité de l'article de recherche. J. Heslot (1983) et De Nuchèze (1998), parmi bien d'autres, rappellent que celui-ci est bâti selon une certaine norme canonique et D. Maingueneau ajoute qu'il comporte un ensemble de routines scripturales. Il est considéré comme étant un genre institutionnalisé et donc contraignant par D. Maingueneau (1995).

En ce qui nous concerne, il nous faudra définir plus précisément comment s'effectue la formulation des problèmes de recherche dans l'article scientifique, et ce, plus particulièrement en sciences de l'éducation. D'autre part, nous tenterons de proposer un outil d'aide à la lecture experte à travers l'identification automatique des problèmes traités dans les articles de recherche. Rappelons que la réussite du second objectif dépend de celle du premier. Ces derniers ne sont pas dissociables.

1.4.2. Genre de l'article de recherche

La notion de genre a suscité de nombreux débats quant au type d'objet qu'elle prétend définir, ce qui a conduit les auteurs qui l'emploient à des définitions relativement différentes. Cela est dû à la diversité des approches auxquelles les auteurs font appel et la diversité de leurs fondements théoriques (F. Rinck 2006). En effet, certains des travaux de recherche sur la question traitent parfois des « genres de discours », tandis que d'autres s'intéressent plutôt aux « genres de textes ». C'est notamment le cas de J.M. Adam (1999) qui traite la question de *genres* en l'opposant aux *types de textes*, ou encore de J.-P. Bronckart (1996) qui mobilise quant à lui à la fois la notion de *genres de textes* et celle de types de discours. Quant à D. Maingueneau (1998), ses travaux relèvent de l'analyse du discours, et les *genres* de discours y occupent une place centrale.

Pour les besoins de cette étude, il convient de souligner la différence qui a été mise en évidence entre « texte » et « discours ».

J.- P. Bronckart (1996) considère que les notions de texte et de discours correspondent à des perspectives différentes alors que D. Maingueneau (2001, 186) les fait correspondre à deux modes d'appréhension des unités transphrastiques donc à deux types d'approches d'un objet commun (cité in Rinck, 2006, 7).

La première approche, celle dont se réclame la linguistique textuelle de J.M. Adam (1999) relève d'une approche de type interne et formel.

Au sein de ce courant, le texte est considéré comme un objet théorique propre, dont la structure rhétorique et linguistique peut être étudiée en tant que telle. La seconde approche, de type externe, est attentive au cadre de l'énonciation et envisage le texte en lien avec la pratique sociale dont il est à la fois un produit et un des éléments structurants. (F. Rinck, 2006, 6).

Pour les besoins de cette étude, nous nous positionnons plutôt dans le cadre de cette seconde approche. En effet, bien que notre étude relève principalement de l'analyse textuelle, nous nous intéressons également aux pratiques discursives utilisées par l'auteur qui ne se réduisent pas à ce qui est observable au niveau textuel.

1.4.2.1. La nature des textes traités

Le domaine des sciences de l'éducation est un champ disciplinaire des sciences humaines et sociales qui tente de répondre à des demandes socio-professionnelles qui émanent des systèmes éducatifs. Aujourd'hui son champ d'étude est lié à d'autres disciplines telles que la psychologie, la philosophie, l'économie ou encore la sociologie. Au départ, les sujets traités relevaient entre autres de phénomènes précis tels que les interrogations portées sur les pratiques d'enseignement alors que de nos jours ils relèvent de phénomènes sociaux, institutionnels, etc.

Divers sujets relevant principalement des pratiques de l'enseignement mais également de la psychologie de l'éducation sont traités dans les articles de recherche qui constituent notre corpus à savoir :

- Si les enfants en cours d'apprentissage de la lecture perçoivent-ils la syllabe à l'écrit ? Comment s'expriment les émotions dans le dessin d'un homme chez l'enfant de 5 à 11 ans ?
- Les orateurs ont-ils intérêt à s'épiler les sourcils afin de comprendre l'influence d'indices non verbaux sur les processus persuasifs.
- Comment peut-on réorganiser l'espace conceptuel au cours des états émotionnels ?
- Comment peut-on maintenir dans le temps des effets d'un apprentissage implicite chez des enfants et des adolescents avec retard mental.

- Quels sont les effets de la fréquence du voisinage orthographique interlangue lors de la reconnaissance visuelle de mots chez les bilingues.
- Si le changement d'attitude et fausse attribution est un effet de la centration sur le comportement de soumission.
- Connaissance du contexte et anticipation posturale chez l'enfant au cours du développement.

Ces articles extraits de la base de données Scientext ont été collectés en fonction de leur disponibilité en ligne. Il n'y a donc eu aucune sélection des sujets traités.

Conclusion

Nous avons tenté tout au long de ce chapitre de situer notre objet d'étude, après avoir distingué entre les notions de problématique et de problème de recherche. Il nous a semblé important dans le cadre de la présente étude de mettre en évidence les différences entre ces deux notions afin de ne pas les confondre.

Nous avons passé en revue quelques-uns des travaux consacrés à l'analyse de discours des écrits scientifiques et qui abordent les différents aspects rhétoriques, énonciatifs ou encore lexicaux de l'écrit scientifique, entre autres ceux de K. Hyland (1998-2002-2005), de F. Grossmann et F. Rinck (2006), de A. Tutin, (2007-2008-2010). Nous avons volontairement laissé de côté d'autres études, davantage orientées vers l'analyse du discours car elles étaient moins centrales pour notre propos. Nous voulons en effet nous concentrer sur la dimension lexicale, plus directement liée aux FD et avons commencé, à l'aide d'exemples extraits de notre corpus, à rendre compte des différents aspects impliqués dans la délimitation de ces formules. En effet, il nous paraît important de bien catégoriser les différentes sortes de lexique que l'on peut repérer dans les écrits scientifiques. Cela nous permet de bien situer le lexique propre à l'expression du problème de recherche, en le référant à des catégories précises, fondées sur les recherches antérieures.

En ce qui concerne l'angle d'approche choisi pour traiter cet aspect lexical, rappelons les références majeures sur lesquelles s'appuie notre réflexion : elle s'inscrit en premier lieu dans la lignée des travaux de M. Pecman (2004) qui s'efforce de modéliser le lexique et la phraséologie des écrits scientifique, avec un focus spécifique sur la LSG ; elle prend appui en second lieu sur ceux d'A.Tutin (2007a-2007b-2008-2010) qui a visé plus

spécifiquement la définition du lexique scientifique transdisciplinaire aussi appelé transversal. Toutefois, comme nous le verrons, notre lexique ne se limite pas au lexique transdisciplinaire et exige de prendre en compte des aspects plus généraux, qu'il nous faudra préciser. Nous tenons à signaler pour finir l'appui majeur pour ce travail que représentent les travaux d'Á. Sándor (2007, 2009, 2010, 2015), qui sont au fondement même de toute cette présente recherche.

Pour les besoins de ce travail, nous recenserons, compilerons et classerons donc le lexique propre à cette langue scientifique au sein des articles de recherche, constitué de textes relevant du domaine des sciences de l'éducation (voir annexe 2).

Chapitre 2 : Formules de discours, collocations, marqueurs polylexicaux, marques du métadiscours

Introduction

Dans ce chapitre, notre objectif est de proposer une définition des formules de discours traduisant le problème de recherche en partant des différentes associations lexicales du discours scientifique étudiées dans la littérature (D.A. Cruse, 1986, Van Kopple, 1986, Haussman, 1989, A. Crismore et al., 1993, I. Mel'čuk, 1998, M. Pecman, 2004, T. Dahl, 2004, De K. Hyland, 2005, A. Adél , 2006, A. Tutin, 2010).

La notion de formule de discours fait donc l'objet de ce chapitre, qui propose d'analyser en quoi celles-ci peuvent se rapprocher et /ou se dissocier des associations lexicales telles que les collocations, les marqueurs polylexicaux ou en encore les marqueurs du métadiscours et dans quelle mesure celles-ci s'inscrivent dans la continuité des travaux portés sur les différents marqueurs lexicaux cités précédemment.

A partir des travaux de D.A. Cruse (1986), F.J. Haussmann (1989), et en référence aux travaux plus récents sur les discours et les textes (K. Hyland, 2004-2005-2007, M. Pecman, 2005, F. Rinck, 2006, A.Tutin, 2010) entre autres, nous allons explorer les rapprochements possibles entre les FD telles que nous venons de commencer à les définir dans notre premier chapitre et les différents types d'associations lexicales étudiées par ces auteurs.

Notre but, à l'issue de ce chapitre, est d'arriver à proposer une définition opérationnelle de la notion de formule de discours afin de permettre un traitement pour la description linguistique et les applications TAL.

2.1. La notion de formule de discours

Le fait d'identifier la problématique scientifique et de définir le problème de recherche traité, fait partie intégrante de toute démarche scientifique. De ce fait il est crucial, du point de vue de la recherche d'information mais aussi d'un point de vue linguistique, d'identifier les éléments à travers lesquels le problème se manifeste. Cependant, comme nous l'avons déjà rappelé, la notion même de « problématique » n'est pas toujours très claire, et il est parfois difficile d'identifier avec suffisamment de précision les formes

linguistiques à travers lesquelles elle s'incarne. C'est pourquoi, nous avons proposé de nous en tenir à un niveau inférieur, en cherchant à délimiter les éléments linguistiques à travers lesquels le ou les problème(s) de recherche traduisant la problématique générale se manifestent.

Nous partons alors de l'hypothèse de travail que certaines associations spécifiques – que nous convenons d'appeler *formules de discours* – jouent le rôle de signaux permettant d'alerter le lecteur sur le fait que l'auteur scientifique est en train de formuler son problème de recherche. Ces FD sont mentionnées dans les phrases où l'auteur résume l'article entier ou bien une partie de l'article mais également dans les phrases où l'auteur signale le problème traité par l'expression d'un *changement cognitif*: une contradiction, une tension, une lacune, une surprise, etc. c'est-à-dire toute expression d'un contraste au niveau de la réflexion scientifique d'où l'attribution des fonctions *résumé* et *changement cognitif* aux formules de discours.

Nous allons montrer dans ce qui suit en quoi les FD telles que nous tentons de les définir dans notre travail, diffèrent des expressions appelées des « marqueurs de métadiscours » (par ex. K. Hyland 2007, voir chapitre), des collocations ou unités phraséologiques - un des objets d'étude du projet Scientext – (voir A. Tutin et F. Grossmann 2014), ou encore les marqueurs de discours (D. Siepmann 2007, F. Boch et.al. 2007), bien qu'elles ne constituent pas pour autant des propositions autonomes.

2.1.1. En quoi les formules de discours diffèrent-elles des collocations ?

Notre objet d'étude, les *formules de discours*, sont des suites de mots ou d'expressions syntaxiquement cohérentes, mais pas nécessairement contiguës, aux structures syntaxiques et composantes lexicales variées. Parmi certains composants constitutifs de ces associations lexicales, nous retrouvons des collocations figurant dans notre corpus d'analyse.

2.1.1.1. Qu'est-ce qu'une collocation ?

La notion de *collocation* a été introduite dans les années 30 par Firth, membre fondateur de l'école contextualiste britannique, qui a développé l'une des conceptions majeures de

la collocation. Ce terme a été introduit pour caractériser les phénomènes linguistiques de cooccurrence qui relèvent de la compétence des locuteurs natifs (G. Williams, 2003, 5).

Le courant contextualiste également pionnier en lexicographie considère les collocations comme un outil d'analyse statistique, permettant d'analyser le style de certains auteurs (B. Orliac, 2004, 17). Ces analyses statistiques se basent sur la grammaire fonctionnelle (développée par M.A.K. Halliday). Cette dernière vise à expliquer le fonctionnement de la langue en contexte dans une situation donnée.

Cette conception sera développée par F.J. Hausmann (1989, 1010) qui considère les collocations comme des combinaisons de deux mots ou de deux lexies. En effet, pour F.J. Hausmann (1989), la collocation est une « *combinaison polaire de deux lexèmes qui a un caractère conventionnel à l'intérieur d'un groupe linguistique* » (cité in J-G. Ceballos-Escalera, 2012). Cette combinaison est responsable du fait que l'un des deux lexèmes (la base) détermine ceux avec lesquels il peut être combiné (le collocatif).

Dans sa classification des combinaisons lexicales, F.J. Hausmann (1989) fait d'abord une distinction entre les combinaisons « fixées » où il répertorie les locutions idiomatiques, et les combinaisons « non fixées ». Celles-ci se divisent en trois types : la co-création, la collocation et la contre-création. La co-création est une cooccurrence libre. La collocation, quant à elle, se distingue par l'affinité de ses éléments qui relève de la norme linguistique ; de la convention. Quant à la contre-création, c'est un « effet de style voulu qui combine des éléments dont la combinaison n'est ni usuelle ni co-créativement concevable » (cité dans U. Heid 1992 : 531). Voici le graphique présenté par U. Heid (1992 : 531) pour illustrer le phénomène.

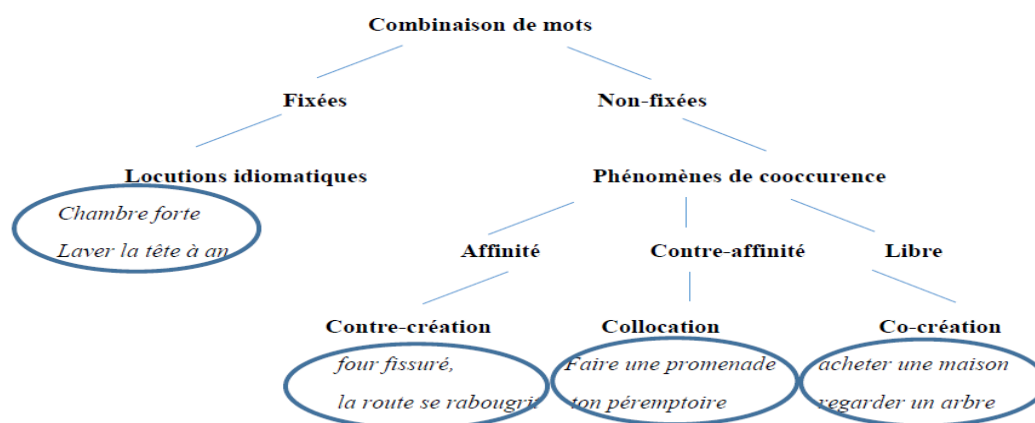


Figure 16 : Typologie de combinaison de mots selon F.J. Hausmann (1989) synthétisée par U. Heid (199, 553)

Notons que F.J. Haussman ne s'intéresse pas à l'aspect statistique ni I. Mel'čuk dont nous exposerons brièvement dans ce qui suit, la conception de la collocation. Car tous deux partent d'une conception linguistique de la collocation, et non d'une conception statistique. Nous nous inscrivons également dans cette lignée car notre objectif n'est pas de rendre compte de l'aspect quantitatif des collocations figurant comme FD au sein de notre corpus. Il s'agit plutôt de parvenir à une définition opérationnelle des FD en traitant l'aspect linguistique.

I. Mel'čuk (1997) présente une définition formelle aux collocations, en remettant en cause l'impossibilité d'un traitement linguistique, souvent alléguée par la tradition :

« Toutes ces bizarreries d'usage sont connues depuis longtemps : ce sont des collocations– des expressions phraséologiques d'un certain type, qu'on ne peut pas prévoir et que la linguistique ne savait pas présenter de façon logique et facilement calculable » (1997, 23).

En partant de la théorie Sens-Texte, I. Mel'čuk (2003) classe les collocations lexicales en fonction de leur sens, mais aussi en prenant en compte, à travers la structure actancielle, certaines de leurs caractéristiques syntaxiques.

Au sein de son Dictionnaire explicatif et combinatoire (DECFC 1984-1992), il présente un modèle qui permet de décrire de façon systématique les unités lexicales de la langue générale à travers les "Fonctions Lexicales".

D'autres courants de la tradition continentale se fondent plutôt sur la notion de 'cooccurrence lexicale restreinte'(F-J. Haussmann & P. Blumenthal, 2006, 3). Ce courant considère lui-aussi que la collocation est un syntagme lexicalisé où deux éléments lexicaux récurrents entretiendraient une relation syntaxique (A. Tutin, 2008-2010). Le caractère binaire ainsi que le statut grammatical de la collocation sont les deux aspects majeurs dans cette définition (*inter alia*, S. Bartsch 2004, A.Tutin, 2008-2010).

Nous observons à travers ce panorama que divers paramètres sont pris en compte dans la définition des *collocations*.

2.1.1.2. Paramètres définitoires des collocations

Parmi les définitions les plus marquantes des collocations, on trouve les paramètres définitoires suivants, mis au jour par F. Grossmann et A. Tutin (2002) :

- *La fréquence de la cooccurrence des éléments ;*

Le problème empirique de la fréquence acceptable a été introduit dans l'approche contextualiste. Selon, D.A. Cruse (1986, 40), on utilise ce terme *collocations* pour désigner la cooccurrence d'une séquence lexicale d'item.

- *L'aspect arbitraire (la non prédictibilité) de l'association lexicale ;*

La plupart des collocations sont difficilement prédictibles. Dans la même perspective, Benson (1989, 3) considère que les collocations ne devraient pas seulement être définies comme des combinaisons de mots récurrents (« recurrent word combinations ») mais aussi comme des combinaisons récurrentes arbitraires (« arbitrary recurrent word combination »).

Selon F.J. Hausmann (1989, cité in F. Grossmann & A. Tutin, 2002)

« [la] transparence n'empêche nullement la collocation d'être imprédictible ».

Reprenons les exemples extraits de cet article de A. Tutin & F. Grossmann (2002) qui traite de ce sujet. La cooccurrence de *précipitations* torrentielles paraît comme moins naturelle que celle de *pluie torrentielle*. Ils illustrent ce propos avec un exemple de collocation plus imagée du type *appétit d'ogre* ou *faim de loup* (vs. *appétit de loup* et *faim d'ogre*).

- *La transparence sémantique de la collocation ;*

Selon D.A. Cruse (1986) et F.J. Hausmann (1989), la plupart des collocations sont transparentes et facilement comprises.

En effet, D.A. Cruse¹¹ (1986,40) considère que la collocation est utilisée pour faire référence à des séquences d'éléments lexicaux qui habituellement se trouvent en co-occurrence, mais qui sont néanmoins totalement transparentes dans le sens où chaque constituant lexical est aussi un constituant sémantique.

¹¹ *The term collocation will be used to refer to sequences of lexical items which habitually co-occur, but which are nonetheless fully transparent in the sense that each lexical constituent is also a semantic constituent. (Cruse, 1986, 40).*

Quant à F.J. Hausmann (1989, 1010), il soutient l'idée selon laquelle la collocation se distingue des locutions à travers son caractère de non-figement et de transparence.

Prenons l'exemple de *déborder d'affection* ; cette collocation verbale est transparente et est facilement comprise par un locuteur de français non natif. En effet, le sens peut se calculer à partir de la signification des parties (voir partie 2.1.1.3).

- *Le caractère binaire de la collocation* ;

Selon A. Tutin (2008, 8), pour avoir une association lexicale à caractère binaire, la collocation requiert la structure suivante : 'prédicat-argument' (le prédicat serait le collocatif et l'argument la base). Elle souligne toutefois que dans certaines collocations atypiques, on ne perçoit pas la structure dissymétrique base-collocatif et cela serait dû à l'ambiguïté de la structure prédicat-argument.

Quant au statut grammatical, A. Tutin (2008, 8) considère que les déterminants, conjonctions et autres mots grammaticaux devraient être décrits d'une façon adéquate bien qu'ils ne soient pas des éléments centraux des collocations.

- *La dissymétrie des composants de la collocation (un élément conserve son sens habituel)* ;

Les composants des collocations n'auraient pas le même statut selon certains linguistes.

En reformulant la définition des collocations donnée par I. Mel'čuk, F. Grossmann & A. Tutin (2002) ont abordé ces éléments constituants à statuts divergents qui composeraient les collocations.

Ces composants sont les suivants :

« ...C (le collocatif) est sélectionné en production pour exprimer un sens donné en cooccurrence avec L (la base) ; Le sens L est habituel ».

Ainsi, la base qui garderait son sens habituel, serait stable sur le plan sémantique contrairement au collocatif dont le sens changerait.

- *La notion de cooccurrence restreinte, de sélection lexicale* ;

I. Mel'čuk et Hausmann, considèrent tous deux que les collocations doivent être envisagées dans le cadre de la production, l'encodage. Selon F.J. Hausmann (1989), la base impose la sélection du collocatif. Quant à I. Mel'čuk, il considère que le choix d'une

des expressions n'est pas libre, mais imposé par l'autre élément dans les collocations (F. Grossmann & A. Tutin, 2002).

- La relation syntaxique entre les deux composantes de la collocation

Ce point a été soulevé par la tradition continentale qui considère que les composants des collocations doivent entretenir une relation syntaxique. On retrouve ce critère de sélection dans la définition donnée par S. Bartsch, (2004, 77) qui considère que les éléments de la collocation doivent avoir une relation syntaxique directe¹²

2.1.1.3. Classification des collocations

Quelques associations lexicales considérées comme étant des FD tirées de notre corpus et qui traduisent un problème de recherche se trouvent être des collocations en relation syntaxique avec d'autres éléments du lexique scientifique. Ces collocations sont alors recensées et décrites selon les critères sémantiques de transparence et de régularité tels qu'ils ont été définies par F. Grossmann et A. Tutin (2002, 26) :

- **Collocations opaques** : le sens est non-compositionnel (il ne peut se calculer à partir du sens des éléments constituant la collocation) car le collocatif perd son sens habituel.

Ex : *Peur bleue*

- **Collocations transparentes** : la compositionnalité du sens (le sens peut se calculer à partir de la signification des parties) et le collocatif est généralement un syntagme.

Ex : *Déborder d'affection.*

- **Collocations régulières** : elles sont transparentes (avec compositionnalité du sens), le collocatif contient le sens de la base ou a un sens générique.

Ex : *Grande tristesse.*

2.1.1.4. L'approche théorique des collocations dans le cadre de notre recherche

Les paramètres définitoires cités au-dessus, nous semblent les plus appropriés pour classer et décrire les collocations figurant dans notre corpus d'analyse. Rappelons que

¹² « must be in a direct syntactic relation with each other ».

les unités analysées se laissent dans certains cas décrire comme des collocations au sens strict (Voir figure 18). Cette analyse lexicale exige des critères tels que les critères syntaxiques et sémantiques pour une analyse linguistique plus fine. Notons, toutefois, que nous n'accordons pas d'importance au critère *statistique* introduit par Firth.

La dimension statistique aurait pu être utile à un stade ultérieur de la recherche, pour identifier celles qui sont réellement productives dans les textes. Or, cet aspect n'est pas traité dans le cadre de cette étude. Pour revenir sur cette dimension, la probabilité statistique des combinaisons lexicales ne prouve pas l'existence des collocations, ainsi, selon l'exemple présenté par Coseriu (Cité in J-G. Ceballos-Escalera, 2012), on peut trouver « mouette blanche » répétée plusieurs fois, mais il s'agit d'une combinaison très fréquente parce que normalement les mouettes sont blanches.

Nous avons donc procédé au recensement, classification et à la description de ces collocations en attribuant un trait conceptuel à la *base* uniquement de la collocation comme nous pouvons le voir à travers la figure 17 ci-dessous qui illustre quelques exemples issus de notre corpus.

Nous revenons plus en détail sur cette notion de « trait » dans la partie 3.2 qui s'intitule développement de l'analyseur au sein modèle du concept-matching dans laquelle cette notion s'inscrit.

Expressions polylexicales/collocations	Catégorie	Trait
Cadre d'analyse	Nominale	PORTEE
Cadre d'étude	Nominale	PORTEE
Cadre explicatif	Nominale	PORTEE
Cadre expérimental	Nominale	PORTEE
Concept fondamental	Nominale	OP_Mentale
Considération théorique	Nominale	OP_Mentale
Enquête de terrain	Nominale	PORTEE
Equipe de recherche	Nominale	PORTEE
Erreur type	Nominale	Contrast
Mettre l'accent	Verbale	OP_Mentale
Mettre en avant	Verbale	OP_Mentale
Mettre en évidence	Verbale	OP_Mentale
Mettre en perspective	Verbale	OP_Mentale
Mise en relation	Nominale	OP_Mentale
Mode de pensée	Nominale	PORTEE
Notion fondamentale	Nominale	OP_Mentale
Nombreux points	Adjectival	Importance

Figure 17 : Expressions polylexicales/collocations à trait spécifique

Parmi les collocations que nous relevons dans notre corpus et qui se trouvent être en dépendance syntaxique avec d'autres concepts constituants donnant lieu à nos concepts généraux de *résumé* et de *changement cognitif*, nous trouvons, entre autres, les exemples ci-dessous.

En effet, nous présentons dans la figure suivante quelques-unes de ces collocations (sans que cette liste ne soit exhaustive) tirées de notre corpus qui ont pour entrées des verbes et noms qui appartiennent à la phraséologie visée.

Base	Collocations
Cadre	<i>Cadre d'analyse</i> <i>Cadre d'étude</i> <i>Cadre théorique</i> <i>Cadre expérimental</i> <i>Cadre commun</i>
Champ	<i>Champ d'étude</i> <i>Champ conceptuel</i> <i>Champ de recherche</i>
Mettre	<i>Mettre l'accent</i> <i>Mettre en avant</i> <i>Mettre en jeu</i> <i>Mettre en évidence</i> <i>Mettre en œuvre</i> <i>Mettre en perspective</i> <i>Mettre au jour</i>
Mise	<i>Mise en relation</i>
Notion	<i>Notion fondamentale</i>
Nombre	<i>Nombre de difficultés</i> <i>Nombre de résultat</i> <i>Nombre de savoir</i>
Idée	<i>Idée centrale</i> <i>Idée nouvelle</i>
Analyser	<i>Analyser ce débat</i>

Figure 18 : Collocations en dépendance syntaxique avec d'autres constituants donnant lieu aux formules de discours

Comme nous pouvons le constater à travers les exemples précédents, toutes ces collocations

sont transparentes ou régulières et leur sens se calcule à partir de la signification des parties.

Ces collocations sont des expressions qui peuvent connaître des alternances et insertions syntaxiques. C'est pourquoi, celles-ci ne s'accompagnent pas d'une opacité sémantique.

Par exemple, *mettre en évidence*, (=mettre d'abord en évidence), *cadre d'étude* (=cadre de cette étude), *nombre de difficulté* (=nombre important de difficultés), etc.

Nous allons aborder dans ce qui suit la formation syntaxique de ces FD associées aux collocations et expressions polylexicales présentes dans notre corpus.

2.1.1.5. Types de constituants des formules de discours

2.1.1.5.1. Constituants sub-phrastiques des formules de discours

Les FD associées à des collocations transparentes et expressions polylexicales figurant au sein de notre corpus, apparaissent plus couramment sous forme de constituants sub-phrastiques.

Dans la tradition des grammaires de constituants, celle-ci correspondent à ce que l'on peut analyser comme des syntagmes. La formation syntaxique de la séquence polylexicale correspond à une expansion obéissant aux règles habituelles de la grammaire : une séquence lexicale associable à une structure XP, pourra être analysée comme ayant la catégorie X.

Par exemple, le syntagme *étude de texte* est construit comme une expansion nominale à une structure classique de type N1 de N2 et la catégorie syntaxique, que l'on peut associer à cet élément est le nom. On parle souvent de structure endocentrique pour ce type de formation des séquences polylexicales.

C'est ainsi le cas d'expressions :

Expansions du nom :

- Champs d'étude
- Cadre théorique

Expansions de l'adjectif

- Nombreux points

Expansions du verbe

- Analyser le débat
- Mettre en évidence

2.1.1.6. *Sur le plan sémantique*

Le critère majeur à prendre en considération pour définir les propriétés sémantiques est la *compositionnalité*. Cette notion, centrale quant à la définition de la polylexicalité, met en jeu deux éléments majeurs qui sont le calcul du sens sémantique et la prédictibilité. Nous allons analyser dans ce qui suit les différentes expressions relevées en rapport avec leur degré d'opacité et de prédictibilité.

Les Figures 17 & 18 montrent que certaines collocations entretiennent une relation syntaxique directe, combinaison caractérisant les collocations (F.J. Hausmann (1989, 1010).

En effet, en s'appuyant sur la structure morphosyntaxique de ces collocations, on distingue deux types de collocations ; nominale et verbale, et dont la base est un verbe/nom et le collocatif, entre autres, un adverbe qui renforce notamment le jugement « centrale, fondamentale ». Aussi, l'association de ces constituants n'est pas arbitraire.

On constate que ces collocations sont à caractère binaire. Pour F.J. Hausmann et I. Mel'čuk, contrairement à D.A. Cruse, la collocation est effectivement composée de deux mots ou lexies. Pour notre part, nous parlerons d'associations de constituants.

Les collocations apparaissent dans des structures V+ Adv, S+Adv, V+ (Subst+nom), N+(Subst+nom). La définition de ces collocatifs (fondamentale, centrale, etc.) dépend de la base. Ainsi, le collocatif suit des règles régulières d'association sémantique. Ceci correspond aux associations lexicales identifiées dans la partie théorique que F. Grossmann et A. Tutin (2002) ont qualifiées *de collocations régulières* du fait qu'elles soient facilement décodables et prédictibles.

Ces collocations se trouvent être en relation syntaxique avec d'autres éléments appartenant au lexique auquel nous avons attribué des traits spécifiques (voir les détails dans la partie 2.3.6. portant sur le lexique des concepts constituants).

Nous expliquons brièvement dans ce qui suit, sur quels principes ont été sélectionnés les segments nécessaires à la réalisation de l'expression du problème de recherche (voir exemple 1 et 2) spécifiques.

De façon prototypique, le problème de recherche est mentionné dans les phrases où l'auteur exprime un *contraste* - une contradiction, une tension, une lacune, une surprise dans sa réflexion scientifique à travers l'usage de FD données comme nous pourrions le

voir plus en détail dans le chapitre 3. Nous appellerons cet acte rhétorique dans sa globalité, *changement cognitif*. La phrase suivante en est un exemple, et le contraste dans la réflexion scientifique est véhiculé par la partie en italique :

(1)-CC¹³ *La recherche a mis au jour la réalité d'un effet-maître (Bressoux, 1994) mais n'est pas encore en mesure de l'expliquer* de façon précise.

Ces FD sont modélisées comme les réalisations des différentes configurations des concepts constituants¹⁴. Ainsi, la formule générale *changement cognitif* est construite à travers des différentes configurations des concepts constituants OPERATION_MENTALE, PORTEE et CONTRASTE comme nous pouvons le voir dans l'exemple qui suit.

(2)-CC *En nous inspirant de ce cadre théorique, nous proposons de préciser notre problématique en termes de gestion des distances aux différents savoirs pour l'étude des questions économiques et sociales*

(1)-CC *La recherche a mis au jour ... mais n'est pas ... l'expliquer*

(2)-CC *...ce cadre théorique, nous proposons, ... notre problématique... l'étude des questions....*

(1)-CC DEPENDANCE_SYNTAXIQUE(mis au jour[OP_MENT],
recherche[PORTEE]) DEPENDANCE_SYNTAXIQUE(n'est
pas[COTRASTE],expliquer[OP_MENT])

(2)-CC DEPENDANCE_SYNTAXIQUE(cadre théorique[PORTEE],ce[DEICTIC])
DEPENDANCE_SYNTAXIQUE(nous[DEICT],proposer[OP_MENT])
DEPENDANCE_SYNTAXIQUE(notre[DEICT],problématique[CONTRAST])
DEPENDANCE_SYNTAXIQUE(étude[DEICT],questions[CONTRAST])

¹³ Précisons que « CC » fait référence à la notion de *changement cognitif*.

¹⁴ Nous reviendrons plus en détail dans la partie 2.3.6 sur le lexique des concepts constituants.

Comme nous pouvons le voir à travers l'analyse précédente des deux phrases (1)-A et (1)-B, la détection des collocations seules, même si celles-ci sont bien étiquetées comme associations lexicales appartenant à notre lexique scientifique, ne suffit pas à identifier la fonction rhétorique de *changement cognitif*. En effet, il est nécessaire, en outre d'avoir premièrement une réalisation des différentes configurations des concepts constituants OP_MENTALE, PORTEE, CONTRASTE, DEICTIC que nous avons attribués à chaque lexème et association lexicale ; deuxièmement, il faut encore que ces concepts soient en relation syntaxique afin d'obtenir notre fonction rhétorique.

Ainsi, les FD telles que nous tentons de les définir dans ce présent travail véhiculent donc des contenus plus larges que les collocations – l'un des objets d'étude du projet Scientext (F. Grossmann & A. Tutin 2014). En effet, celles-ci ne constituent pas à elles seules nos FD. Néanmoins, elles ne peuvent être isolées car certaines associations lexicales de notre corpus se trouvent être des collocations en dépendance syntaxique avec d'autres éléments qui constituent notre lexique à traits spécifiques propres à cette étude et qui donne lieu aux FD véhiculant un problème de recherche.

2.2. Les marqueurs de discours polylexicaux

Comme nous avons l'avons déjà signalé, les FD, telles que nous tentons de les définir dans cette étude diffèrent également des marqueurs de discours polylexicaux décrits dans des travaux antérieurs notamment dans ceux de D. Siepmann (2005,2007).

Selon G. Bordet (2011, 4), les marqueurs discursifs sont des éléments syntaxiques dont le choix et les modalités d'usage marquent une stratégie rhétorique au service des objectifs du genre. Ces marqueurs, qui incluent notamment démonstratifs, conjonctions et connecteurs logiques, ont pour fonction de marquer la progression logique et rhétorique d'un texte, garantissant les conditions d'une interprétation cohérente par le lecteur (S. Carter-Thomas, 2008).

2.2.1. Paramètres définitoires des marqueurs polylexicaux

Parmi les différents critères définitoires des marqueurs discursifs présentés dans les travaux de D. Siepmann (2007,124), nous retrouvons un premier critère d'ordre

pragmatique. Il a été explicité par R. Risselada et Spooren (1998,132)¹⁵ qui considèrent comme marqueurs de discours polylexicaux, toutes les expressions en langage naturel dont la fonction principale est de faciliter le processus d'interprétation des relations de cohérence entre différents segments textuels et/ou des aspects de la situation d'énonciation.

Ils mettent en évidence le lien entre l'interprétation de la cohérence du discours et l'adéquation de ce discours avec la situation de communication et ses objectifs contextuels (G. Bordet, 2011,4).

D. Siepmann introduit un autre critère, d'ordre lexicologique cette fois. Il considère qu'il est relativement facile de distinguer les unités polylexicales des combinaisons libres car les premières, contrairement aux secondes, comportent toujours un certain degré de figement. D. Siepmann (2005, 49)¹⁶ observe que les marqueurs polylexicaux peuvent être décrits comme étant des collocations souvent discontinues ou des expressions figées. Cela n'est pas le cas des FD véhiculant un problème de recherche telles que nous les entendons. En effet, les éléments qui les constituent sont entre autres des associations lexicales régulières (voir 2.1.1.6. *Sur le plan sémantique*) qui ne sont pas figées.

Enfin, le dernier critère relève de la fréquence. Ce critère a permis entre autres d'intégrer des connecteurs fréquents tels que *par conséquent*, bien que composés de deux mots, parmi les connecteurs monolexicaux tels que *mais*, *pourtant*, *néanmoins*, etc. Ces marqueurs ont été appelés *marqueurs de premier niveau* (« first-level discourse markers »). Ces marqueurs de premier niveau se trouvent être présents dans notre liste de lexique à traits spécifiques. Nous reviendrons plus en détail sur la description de ce lexique dans la section suivante de ce chapitre.

Quant aux marqueurs polylexicaux moins fréquents, ils ont été qualifiés de *marqueurs de deuxième niveau*, « second-level discourse markers » (SDLM). Ces derniers ont été définis¹⁷ comme des expressions figées ou collocations de fréquence moyenne,

¹⁵ “Discourse markers can be defined as those natural language expressions whose primary function is to facilitate the process of interpreting the coherence relation(s) between a particular unit of discourse and other, surrounding units and/or aspects of the communicative situation.” (Risselada et Spooren (1998, 132)).

¹⁶ “...in lexicographic terms, multi-word discourse markers can be described as collocation or fixed expressions.” D. Siepmann (2005,49)

¹⁷ “Second-level discourse markers,... are medium-frequency fixed expressions or collocations composed of two or more printed words acting as single units. Their function is to facilitate of interpreting

composées de deux mots ou plus mais qui fonctionnent comme segments textuels. Leur fonction consiste à faciliter le processus d'interprétations des relations de cohérence entre différents segments textuels et/ou des aspects de la situation d'énonciation.

Rappelons que les FD telles que nous tentons de les définir sont des cooccurrences entre autres régulières et transparentes.

Toutefois, nous allons nous intéresser aux marqueurs de second niveau et voir si ce seul critère parfois de figement les dissocient des associations lexicales traitées au sein de cette étude. Nous traduisant dans la figure 19, les exemples de D. Siepmann (2005) de l'anglais concernant la taxonomie des marqueurs de second niveau.

coherence relation (s) between elements, sequences or texts segments and/or aspects of communication situation." (D. Siepmann, 2005, 52).

Marqueurs de discours de deuxième niveau (SLDM)
<i>C'est une chose Mais c'est une autre chose....</i>
<i>Ça serait une erreur de....., il est également intéressant de...</i>
<i>Considérons l'exemple....</i>
<i>Cela est du....., l'explication provient de...</i>
<i>Etroitement défini</i>
<i>Premièrement, nous devons ...</i>
<i>Il ne me reste, comme point final</i>
<i>Cela n'implique pas que....., donc il se trouve que....</i>
<i>Il y a deux raisons principales pour cela....</i>
<i>A présent, nous allons brièvement décrirel'examen de ...doit-être laissé jusqu'à....</i>
<i>Il est souvent dit que,à présent, considérons que....</i>
<i>Ce n'est pas le lieu...</i>
<i>Il convient de mentionner en passant que,</i>
<i>La question qui se pose...., la prochaine interrogation évidente....</i>
<i>Il faut souligner que....., notons que.....</i>
<i>Il convient de reconnaître que...</i>
<i>Mais ce n'est pas le point.... , le point est que....</i>
<i>Une chose est sûre, ...</i>
<i>Il est suggéré que...., imaginons que....</i>
<i>Pour confondre encore l'image.....</i>
<i>[X] considère que....., il a été noté que....</i>
<i>En d'autres termes.....</i>

Figure 19 : Marqueurs de discours polylexcaux de second niveau (D. Siepmann, 2005)

D. Siepmann (2005) se base sur la fonction textuelle des marqueurs de second niveau afin d'établir sa taxonomie (voir partie 2.2.3.). Les fonctions discursives qu'il attribue se basent sur les relations de cohérence entre différents segments textuels et/ou des aspects de la situation d'énonciation. Nous nous intéressons dans ce qui suit aux réalisations lexico-grammaticales de ces marqueurs.

2.2.2. La réalisation syntaxique des marqueurs de discours

A travers la figure suivante, nous mettons en évidence les différentes réalisations lexico-grammaticales des marqueurs de discours en français en les illustrant avec quelques exemples.

Catégorie	EXEMPLES
<i>Expressions figées (phrasèmes)</i> - Ensembles d'expressions structurellement incomplètes - Ensembles d'expressions structurellement complètes	- <i>Soit disant en passant, quoiqu'il en soit, de ce fait</i> - <i>et j'en passe</i>
<i>Fragment de phrase</i> - Anticipatif <i>il/ce/cela</i> + syntagme verbal + adjectif - Pronoms personnels (<i>je/on/nous</i>) + syntagme verbal + (<i>inf+</i>) + <i>que</i> + (<i>à/de</i>) Syntagme nominal - Première personne au pluriel à l'impératif + Complément - Syntagme adverbial - Syntagme infinitif introduit par <i>pour</i> ou <i>sans</i> - <i>Clauses adverbiales</i> - Syntagme nominal + être + <i>que</i> (syntagme) Syntagme nominal + <i>c'est que</i> - Syntagme non-verbal	- <i>Il est admis que</i> - <i>On sait que, croyons que nous le verrons</i> - <i>Voyons la chose de près, disons donc</i> - <i>Comme nous le verrons plus loin</i> - <i>Pour être complet, sans exagérer</i> - <i>mieux encore</i> - <i>Le problème est que</i> <i>Le problème c'est que</i> - <i>Autre remarque, fait plus grave</i>
<i>Marqueurs de phrases intégrés</i> - Syntagme nominal + syntagme verbal - Phrase comme segments	- <i>Nous fournit un bon exemple</i> - <i>On pourrait multiplier les exemples</i>

Figure 20 : Réalisation lexico-grammaticales des marqueurs de discours en français

...cette interprétation.....mène à penser...), notre objet d'étude diffère de ces marqueurs de discours en général. Toutefois, les éléments qui constituent ces expressions discursives peuvent parfois appartenir à notre phraséologie comme c'est notamment le cas des catégories *Syntagme nominal+être+que* (ex. le problème est que) et *Syntagme nominal+c'est que* (ex. le problème c'est que). L'exemple (3)-CC illustre cette réalisation lexico-grammaticale dont le sujet est passif. Nous retrouvons également des éléments lexicaux présents dans notre lexique à trait spécifique à savoir les connecteurs

monolexicaux de type *par, mais, pourtant, néanmoins*, etc. auxquels nous avons attribué le trait CONTRAST.

(3)- CC **Le problème** est que **cette interprétation** de la transposition didactique **mène à penser** que la transformation **opère** sur un objet pour produire un nouvel objet.

(3)- CC

GROUPE {NP{Le NOUN {problème_CONTRAST}} FV {est} BG {que} NP {ce_DEICTIC NOUN {interprétation_ PORTEE}} PP{de NP{la transposition}} NP{AP{didactique} mène} NP{à} IV{penser_OP_MENTAL} BG{que} NP{la transformation} NP{opère } PP{sur NP{un objet}} IV{pour produire} NP{un AP{nouvel} objet} .}

La réalisation syntaxique des associations lexicales étudiées sont plus étendues que celles des marqueurs de discours également. Voici d'autres exemples d'associations lexicales appartenant au domaine de l'écrit scientifique qui n'ont pas été consignées.

(1)- CC *La recherche a mis au jour mais n'est pas en mesure de l'expliquer.....*

(2)- CC *.....ce cadre théorique, nous proposons notre problématique 'étude des questions*

(3)-CC *Le problèmecette interprétation.....mène à penser...*

La condition pour qu'une formule de discours traduisant un problème de recherche se réalise, serait que ces réalisation syntaxiques ; *Nom + det + nom, Det + nom, Det + nom + adj, V+ prep + V-inf* dont les constituants appartiennent à notre lexique scientifique soient en dépendance syntaxique [voir exemple (4)-CC].

Soulignons le fait qu'à ces associations lexicales peuvent correspondre différents types de structures syntaxiques (interrogatives directes, indirectes ; passives ou actives) mais que notre système doit être en mesure de dépasser ces différences, puisque nous nous fixons pour objectif de repérer toutes les occurrences d'associations lexicales exprimant un problème de recherche. Nous revenons plus en détail sur les structures des phrases dans lesquelles apparaissent ces associations lexicales dans la partie 3.8.

2.2.3. Fonctions discursives des marqueurs discursifs polylexicaux

D. Siepmann (2007, 125) a procédé à la classification de plusieurs catégories fonctionnelles de ces marqueurs de discours. Il soutient qu'un marqueur de discours ne saurait avoir plus d'une fonction primaire. Nous le rejoignons dans le sens où il ne nie pas pour autant le fait que certains marqueurs sont susceptibles d'assumer plusieurs fonctions. Parmi ces fonctions nous retrouvons ; *les marqueurs de comparaison et de contraste* (ex. Où s'arrête...et où commence...), *les marqueurs concessifs* (ex. en admettant que.../encore faut-il que...), *les marqueurs d'exemplifications* (ex. Soit/prenons un seul exemple), *les marqueurs d'explication* (ex. On peut en donner l'explication suivante), *les marqueurs de définition* (ex. Le X est Y si l'on a :), *les marqueurs d'énumération* (ex. Pour commencer,..., Enfin,...), *les résumés et les charnières de terminaison* (ex. nous terminons en indiquant...), *les marqueurs d'inférence* (il ressort de ce qui précède que.../C'est dire que...), *les marqueurs de causalité* (ex. Deux raisons à cela), *les marqueurs d'annonce* (ex. Je me propose de passer en revue...), *les marqueurs d'introduction thématique* (ex. Abordons maintenant...), *les marqueurs d'exclusion* (ex. excluons toutefois le cas de...), *les marqueurs de digression* (ex. Notons au passage que (...)), *les marqueurs d'interrogation et de réponse* (ex. question/ Le tout est de savoir), *les marqueurs de mise en relief* (ex. encore faut-il remarquer que/notons que), *les marqueurs d'information* (pour information, retenons que), *les marqueurs de classification* (ex. Soyons bien clair), *les marqueurs de suggestion* (ex. il va sans dire que.../il est sûr que...), *les marqueurs d'hypothèse* (ex. on peut supposer que.../Supposons que...), *les marqueurs de restriction* (un autre problème concerne.../complication supplémentaire), *les marqueurs de renvoi et d'attribution* (ex. [nom de personne]affirme que), *les marqueurs de reformulation et les récapitulatifs* (ex. Pour dire en d'autres termes, ainsi qu'il est dit plus haut...).

Après avoir survolé les différentes catégories de marqueurs polylexicaux, il est intéressant pour les besoins de cette étude de s'attarder sur les *marqueurs d'hypothèse et de restriction* et de s'interroger pour savoir en quoi ces marqueurs convergent ou divergent de notre objet d'étude.

Prenons pour exemple la phrase suivante tirée de notre corpus.

(4)-CC Concernant les aspects pragmatiques du récit, les résultats obtenus confirment partiellement notre hypothèse de départ.

(4)-CC.. *les résultats....., confirment..... notre hypothèse*

DEPENDANCE_SYNTAXIQUE(**résultats**[PORTEE],**confirmer**¹⁸[PORTEE])

DEPENDANCE_SYNTAXIQUE(**confirmer**[PORTEE],**hypothèse**[CONTRAST])

DEPENDANCE_SYNTAXIQUE(**hypothèse**[CONTRAST], **notre**[DEICT])

Comme nous pouvons le constater, le marqueur d'hypothèse *notre hypothèse* qui se trouve être en dépendance syntaxique avec d'autres éléments (*confirmer, résultat, notre*) du lexique propres aux écrits scientifiques (voir Annexe 2) est recensé dans notre liste. Nous retrouvons d'ailleurs d'autres occurrences notamment de marqueurs d'interrogation au sein de notre corpus que nous avons étiqueté sous l'appellation OBSTACLE/CONTRAST (*la question est, etc.*) ou encore les résumés (*pour résumer, etc.*). Nous remarquerons donc que certaines des catégories fonctionnelles des marqueurs de discours avancées par D. Siepmann sont couvertes par un ou plusieurs de nos concepts constituants notamment celui de CONTRAST. Toutefois, les autres marqueurs de discours doivent être écartés car ils ne sont pas couverts par nos concepts constituants traduisant un problème de recherche¹⁹.

2.3. Métadiscours

Il est important d'aborder cette question du métadiscours car nous situons nos associations lexicales à un méta-niveau de discours. Ce qui justifie qu'elles ne soient pas communes à une discipline en particulier et qu'elles soient susceptibles d'être utilisées par les auteurs des écrits scientifiques afin de formuler leur problème de recherche.

Dans cette partie, nous commencerons par passer en revue les travaux consacrés au métadiscours entre autres ceux de K. Hyland, (1998 ; 2004 ; 2005), T. Dahl (2004), A. Adél (2006), N. Toumi (2009). Nous procéderons à une analyse des marques

¹⁸ Confirmer a été associé au trait conceptuel _PORTEE (voire définition partie 3.6) mais également au trait POSITION. En effet, en plus du fait que ça soit un verbe d'apport scientifique, il peut également représenter une prise de position.

¹⁹ Voir détail dans la partie 3.6 & 3.7. Hiérarchie des concepts constituants.

métadiscursives des écrits scientifiques en passant en revue les approches énonciatives et rhétoriques des textes en lien avec l'activité scientifique (F. Boch & F. Rinck, 2010, 6).

Nous nous intéressons à cette notion de métadiscours en particulier car les FD telles que nous tentons de les définir tout au long de notre travail relèvent bien de ce méta-niveau. Le but, dans le cadre de ce travail est d'arriver à identifier les éléments linguistiques appartenant au métadiscours et constituant nos FD. Cela nous permettra d'établir une typologie lexicale afin de procéder à une classification des différentes marques métadiscursives auxquelles les auteurs font appel dans leur formulation des problèmes de recherche. Rappelons que notre tâche consiste à détecter automatiquement puis à classer ces formules du discours notamment celles qui sont utilisées pour véhiculer le problème de recherche.

2.3.1. Le métadiscours : définitions et questionnements

La notion de métadiscours a été introduite par Harris (1959) puis reprise et approfondie par A. Crismore (1982, 1983, 1984a, 1989, 1990) et W.J. Vande Kopple (1985, 1990). Depuis, de nombreuses études ne cessent de s'intéresser à ce concept (A. Mauranen (1992; 1993), R. Markkanen et al. (1993), K. Hyland (1998; 2004; 2005), T. Dahl (2004), A. Adél (2006).

La dénomination de métadiscours est en général propre à l'écrit car dans le cas d'un discours oral, on parle plutôt de *metatalk*.²⁰ Cette notion réservée à l'écrit, est définie comme étant le discours du/sur le discours (W.J. Vande Kopple, 1985) où l'auteur commente explicitement son propre discours (A. Adél, 1996, 20).

L'auteur fait appel au métadiscours en tant qu'outil linguistique qui lui permet d'aborder le texte au-delà de son contenu propositionnel. Précisons : l'auteur a la capacité d'aborder, d'une part, le sujet de son écrit - en d'autres termes le contenu asserté -, et d'autre part, la situation de communication, grâce à une propriété de réflexivité propre aux langues dites naturelles. Le métadiscours est considéré comme étant l'un des indications de réflexivité du langage. Nous reviendrons dans ce qui suit sur cette notion.

Il est également considéré comme le discours qui fait référence à la manifestation linguistique de l'auteur dans son écrit au moment d'interagir avec son lecteur potentiel et

²⁰ Soulignons toutefois que ce terme est propre aux études portant sur l'anglais dans la mesure où en français ce terme n'est guère utilisé.

ce afin d'aider ce dernier à organiser, classer, interpréter, évaluer et réagir aux écrits (W.J. Vande Kopple, 1985).

Les auteurs mettent généralement l'accent sur deux dimensions du métadiscours, la dimension sociale et la dimension réflexive et cognitive.

2.3.1.1. La dimension sociale

Cette dimension se fonde sur l'idée selon laquelle la communication ne comprend pas uniquement un échange d'information mais implique d'autres éléments, entre autres les personnalités, attitudes et les suppositions/hypothèses des participants de l'échange (K. Hyland, 2005, 3). Ainsi, le métadiscours implique une dimension sociale du sens en révélant la personnalité et l'identité de l'auteur (N. Toumi, 2009, 64) : le métadiscours se fonde sur une vision de l'écrit en tant que forme d'engagement social et communicatif, ce qui se traduit par une interaction sociale entre l'auteur et son lecteur potentiel. Le métadiscours, en ce sens, fait donc référence aux aspects dans le texte qui font apparaître la présence de l'auteur/lecteur, ce qu'A. Adél (2006, 2)²¹ explicite de la manière suivante :

« Dans un sens plus large, le métadiscours fait référence à des éléments linguistiques qui révèlent la présence de l'auteur et celle du lecteur dans le texte, et ce soit en se référant à l'organisation du texte ou en commentant le texte d'une autre manière ».

K. Hyland (1997) ajoute que le métadiscours fait référence aux aspects qui organisent le discours, attirent l'attention du lecteur, et signalent l'attitude de l'auteur de manière explicite. Il est également admis que l'usage des traits/marques métadiscursifs aident à transformer un texte brut et à mettre en évidence la capacité de l'auteur à utiliser les moyens nécessaires afin d'assurer la compréhension et l'acceptation de son contenu propositionnel (K. Hyland, 2004). A. Crismore (1989, 64) nous rejoint pour dire que cet outil linguistique aide l'auteur à guider et à diriger et non à informer en proposant le modèle suivant :

²¹ Voici la définition donnée par A. Adél (2006, p2) : « *In a wider sense, it refers to linguistic items which reveals the writer's and the reader's (or speakers and hearers) presence in the text, either by referring to the organisation of the text or by commenting on the text in another way* ».

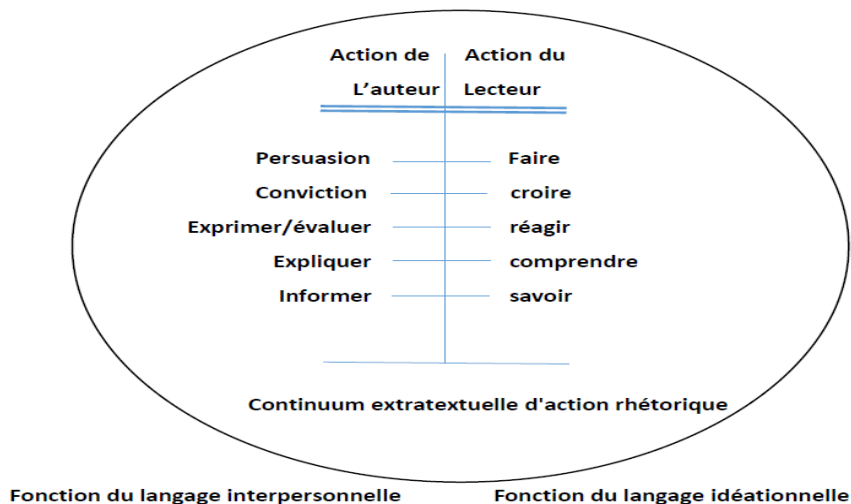


Figure 21 : Modèle du métadiscours de A. Crismore (1989, 91)

Dans ce modèle, le *metadiscours* a pour but de remplir une fonction textuelle et interpersonnelle dans le discours écrit. D'ailleurs, l'auteur rejoint W.J. Vandekopple (1985) dans ce sens. Il considère que les actions rhétoriques de l'auteur à travers les marqueurs du métadiscours textuel et interpersonnel se situent dans un continuum. Nous revenons plus en détail sur ce point dans la partie 2.3.3.1, qui prend appui sur l'approche large du métadiscours, courant dans lequel s'inscrit A. Crismore.

En effet, l'auteur n'a pas pour unique but de véhiculer des informations et des connaissances à travers son écrit, mais également de faire en sorte, à travers divers moyens linguistiques, de faire accepter le contenu asserté par son lecteur potentiel en interagissant avec celui-ci. C'est pourquoi l'écrit est perçu comme un engagement dans lequel l'auteur et son lecteur potentiel interagissent à travers des codes de communication sociale.

2.3.1.2. La dimension réflexive et cognitive

Pour revenir à cette autre notion qu'est la *réflexivité*, notons qu'elle est considérée comme un trait universel et propre au système de communication humaine (C. F. Hockett 1977, 173 cité in A. Adél, 2006, 1). Cette propriété est utilisée dans toutes les langues naturelles par l'auteur ou le sujet parlant, afin de commenter son propre discours, autrement dit afin de parler du langage. M. Silverstein (1976,16) considère que la réflexivité contribue à faire du langage humain ce système à caractère unique à travers

tous les codes culturels de la communication sociale. R. Jakobson (1980) soutient l'idée que c'est une propriété centrale du langage humain dans notre usage quotidien. A. Adél (2006, 1) partage ce point de vue et considère que cet acte cognitif n'est pas seulement propre aux discours des linguistes ou d'autres professionnels de la société dont la tâche consiste à s'interroger sur le langage. En effet, cette propriété propre à chaque individu permet, notamment aux auteurs, d'aborder leur propre discours à travers divers moyens linguistiques, et ce, afin d'interagir avec leur lecteur-potentiel dans un but particulier et stratégique. Aussi, cette propriété permet à tout auteur d'opérer sur son objet linguistique.

2.3.2. Le contenu propositionnel vs. niveau métadiscursif

Comme nous l'avons laissé entendre précédemment au moins deux niveaux discursifs peuvent être identifiés dans un écrit en général (W.J. Vande Kopple, 1985 ; A. Crismore et al, 1993) ; le niveau propositionnel et le niveau métadiscursif. Toutefois, les tentatives de délimitation entre ces deux niveaux ont rapidement montré leurs limites. De nombreux auteurs ont tenté d'aborder ce problème en essayant de distinguer entre les deux niveaux. Pour ce faire, ils tentent de spécifier et d'opposer les traits qui les caractérisent. C'est notamment ce que cherchent à faire G. Williams (1981) et Dillon (1981) par exemple. Ces derniers considèrent que le niveau propositionnel se signale au lecteur avec des informations concernant le sujet traité. Alors que le niveau métadiscursif requiert de l'attention quant à l'acte d'écriture, d'où les éléments et indices linguistiques divers qui aident entre autres à l'organisation et classification du discours.

Ce point de vue est partagé par W.J. Vande Kopple (1985 et 2002) qui soutient l'idée que le premier niveau, « idéationnel », qui véhicule des idées, se manifeste à travers des informations concernant le sujet de l'écrit, c'est-à-dire le contenu asserté.

Quant au second, il correspond aux éléments/marqueurs/ indices linguistiques qui permettent d'organiser, classifier, interpréter, évaluer et réagir aux écrits, d'où sa définition comme discours du/sur le discours.

Concernant le niveau métadiscursif, les auteurs s'accordent à lui attribuer un statut non-propositionnel. A. Crismore et al. (1993, 40) par exemple, justifient la nature de ce statut en soutenant qu'à ce niveau les auteurs n'utilisent les outils linguistiques métadiscursifs que dans le but d'ajouter quelque chose au contenu propositionnel. Car il considère que c'est un moyen utilisé par l'auteur pour s'immiscer dans le discours que cela soit de

manière implicite ou explicite afin d'aider le lecteur-potential à comprendre comment il peut organiser, évaluer et interpréter les informations véhiculées, plutôt qu'à l'informer.

Si l'on s'en tenait là, la démarche entreprise par les auteurs pour aborder la distinction entre ces deux niveaux ne se révélerait pas fructueuse car ils se contentent de relever les traits caractérisant chacun d'eux. M.A.K. Halliday (1994) tente d'aller plus loin en partant du niveau propositionnel. Il part de l'idée que le matériel propositionnel représente ce qu'on peut discuter en terme de validité, qu'on peut affirmer, nier, mettre en doute, tempérer, regretter, sur lequel on peut insister, etc.²² (M.A.K. Halliday, 1994, 40 cité in K. Hyland et P. Tse, 2004, 160). Il propose alors un moyen afin de délimiter ces deux niveaux à travers un test de falsifiabilité qui vise à identifier les propositions. Mao (1993) remet en cause ce test car il considère que les deux niveaux peuvent être falsifiés (Mao, 1993, 267 cité in K. Hyland, 2005, 19). K. Hyland (2005) remet également en doute ce test et argumente en prenant pour appui un exemple donné par G. Myers (1990), qui traite des écrits académiques. G. Myers (1990) donne un exemple où il est montré comment un texte paru dans le journal *Science* a été réécrit et adapté pour un public plus large et non spécialisé dans la revue *Scientific American*. Il constate alors que lorsqu'un texte est réécrit par un éditeur pour un public autre que celui principalement visé, son contenu reste identique alors que le sens qu'il véhicule peut radicalement changer à travers la re-textualisation. Il arrive alors à la conclusion que le sens d'un texte ne dépend pas uniquement du matériel propositionnel car l'on a plutôt affaire à un paquet de « contenu propositionnel et métadiscursif ».

Il ajoute que le métadiscours est un élément essentiel et indispensable qui contribue à la construction du sens (K. Hyland, 2005, 22).

Ainsi, cette séparation de « niveau de sens métadiscursif » soulevée par W.J. Vandekopple et autres semble remise en doute car le sens d'un texte dépend de chacun des éléments, métadiscursif et propositionnel. La seule raison pour distinguer entre ces deux niveaux est motivée par des fins de recherche et d'exploration (K. Hyland & P. Tse, 2004).

Un autre problème est soulevé dans les études portées sur le métadiscours. Ce problème relève de l'ordre syntaxique ou fonctionnel. K. Hyland (2005) et A. Adél (2006)

²² "propositional material is something that can be argued about, affirmed, denied, doubted, insisted upon, qualified, tempered, regretted and so on" (M.A.K. Halliday, 1994:40 cited in K. Hyland and P. Tse, 2004:160).

s'accordent à penser que la fonction métadiscursive est attribuée aux marques textuelles en fonction du co-texte dans lequel celles-ci apparaissent. Ainsi, une marque textuelle peut se voir attribuer dans certains cas de figure la fonction métadiscursive et pas dans d'autres en fonction de l'énoncé dans lequel elle apparaît et ce comme nous pouvons le constater dans l'exemple qui suit traduit des travaux de K. Hyland (2005, 22-24).

(a) Je pense qu'elle est folle. Premièrement, elle m'a crié dessus. Deuxièmement, elle a déchiré le mail.

(b) Quand je lui ai appris la nouvelle, premièrement, elle m'a crié dessus. Deuxièmement, elle a déchiré le mail.

A travers cet exemple, on constate que les marques de séquentialité (*premièrement, deuxièmement*) sont utilisées dans le premier exemple pour lister les arguments et persuader le locuteur que la personne cible est folle. Alors que dans le second exemple ils sont utilisés pour montrer le déroulement de l'évènement. Ainsi, le co-texte est un élément important car il détermine la nature de la fonction.

Cet élément est primordial dans l'établissement des règles formelles qui nous permettent d'identifier les phrases qui véhiculent le problème de recherche. En effet, pour les besoins de la présente étude, notre démarche consiste à observer la structure syntaxique de chaque phrase notamment l'occurrence et la dépendance des lexies représentant nos concepts spécifiques dans leur co-texte (ces dépendances sont à la base du concept général de *résumé* et de *changement cognitif*). Toutefois, suite à des problèmes de référence au sein du co-texte, des phrases pouvant parfois avoir un rôle important ne sont pas identifiées. Nous revenons plus en détail dans le chapitre 4 et présentons la solution avancée pour résoudre ce problème.

2.3.2.1. Les fonctions textuelles et interpersonnelles du métadiscours

Certains auteurs s'accordent à dire que le métadiscours a deux fonctions distinctes. La distinction entre ces deux niveaux trouve son origine dans les études d'M.A.K. Halliday (1973). Ce dernier détermine deux macrofonctions du langage au niveau métadiscursif, textuelle et interpersonnelle (T. Dahl, 2004, 1811) et une fonction idéationnelle correspondant au niveau propositionnel.

Au niveau métadiscursif, nous retrouvons, d'une part, le métadiscours textuel considéré comme comprenant tous les éléments qui visent à organiser le texte dans le but d'aider le lecteur potentiel (W.J. Vande Kopple, 1985). A ce niveau, l'auteur s'assure d'organiser les informations proportionnelles de manière cohérente et appropriée à un contexte donné (Hylliday, 1994 cité in K. Hyland, 2005, 25). Ainsi, à ce niveau, la langue est utilisée afin d'organiser le texte de manière cohérente en reliant ce qui est transmis au monde et au lecteur potentiel (K. Hyland, 2005, 7). Ce type de métadiscours, appelé *métatexte* (*metatext*) par certains auteurs (A. Mauranen, 1993; Valero-Garce's, 1996; Moreno, 1997; Bunton, 1999) contribue à l'organisation du texte en apportant un cadre qui met en lumière la structure schématique de ce dernier (N. Toumi, 2009, 65)²³. En effet, cette première fonction discursive vise à guider le lecteur à travers le texte. L'intérêt est porté sur la structure, l'action discursive ainsi que la phraséologie du texte (Adël, 2006, 20). La fonction textuelle vise alors à organiser le contenu propositionnel de manière cohérente. Parmi les exemples, nous retrouvons entre autres : *dans cet article, pour conclure, pour résumer...* on considère ces exemples comme des marques métatextuelles.

D'autre part, on retrouve le métadiscours interpersonnel également appelé *interaction auteur-lecteur* (A. Adël, 2006,20)²⁴. Comme son nom l'indique, ce second type de métadiscours est utilisé par l'auteur afin d'interagir avec son lecteur potentiel concernant le contenu asserté. Le but étant de maintenir une relation avec celui-ci afin de l'influencer et de lui faire mieux accepter ce contenu propositionnel ; exemple : *comme vous pouvez le voir, vous allez probablement penser, etc.*

Ainsi, ce métadiscours exprime le « persona²⁵ » de l'individu qui est l'auteur. Cette fonction requiert une démarche cognitive de l'auteur et témoigne d'un effort de *réflexivité* de sa part.

²³ Voici la définition donnée par N. Toumi (2009, 65) : '*textual metadiscourse*', *contribute to the deployment of rhetorical strategies used to express a theory of experience in a coherent way. It provides a framework which clarifies the schematic structure of the tex..*

²⁴ Voir modèle de configuration personnelle et impersonnelle du métatexte et de l'interaction auteur-lecteur en annexe 3.

²⁵ C'est-à-dire une personnalité créée mise en avant lors de l'acte de communication (N. Campbell 1975, 394).

2.3.3. Modèles du métadiscours

Depuis que la linguistique appliquée s'est intéressée à l'étude du métadiscours, il y a une vingtaine d'années, deux courants distincts se sont développés. A. Mauranen (1993, 145-155) et A. Adél (2006, 167-179) sont revenues sur ces deux traditions. La première, appelée 'intégrative' par A. Mauranen (1993) ou 'modèle intégratif' par A. Adél (2006), fait appel à une approche large du métadiscours et considère que l'interaction textuelle est fondamentale pour élaborer une typologie du métadiscours. Quant à la seconde, appelée 'non-intégrative' (A. Mauranen 1993) 'modèle réflexif' (2006), elle se base sur une définition du métadiscours plus étroite car elle considère que la *réflexivité* est le fondement de cette notion. Selon W.J. Vande Kopple (1985), la différence entre ces deux approches, large et étroite, se trouve dans les traits qui caractérisent chacune d'entre elles, à savoir *l'attitude* et les *marqueurs de validité* (N. Toumi, 2009, 65) qui constituent les marqueurs de positionnement.

2.3.3.1. L'approche large ou intégrative

Les nombreuses études qui se sont intéressées aux métadiscours textuel et interpersonnel, s'inscrivent dans un des deux courants. Les études menées par W.J. Vande Kopple (1985), A. Crismore et al. (1993), R. Markkanen et al. (1993) et K. Hyland (1998b; 2004) s'inscrivent dans la perspective la plus large du métadiscours, où ce dernier est perçu comme un moyen de rendre explicite la présence de l'auteur, soit à travers des attitudes ou des commentaires vis-à-vis du texte, ou bien en montrant la manière dans lequel le texte s'organise (A. Adél, 2006, 168), c'est-à-dire un moyen d'aider le lecteur potentiel à organiser, évaluer et interpréter les informations du texte cible. Ce modèle est illustré de manière générale dans la figure 22 (traduit de la figure d'Adél, 2006, 168).

Objet d'étude	Éléments linguistiques qui montrent comment le texte s'organise ou bien qui font référence au texte lui même	Éléments linguistiques qui représentent l'attitude de l'auteur par rapport à ce qui est dit dans le texte
Exemples	<i>premièrement, pour résumer, en d'autres termes</i>	<i>certainement, peut-être</i>
Fonction Dominant	TEXTUELLE	INTERPERSONNELLE

Figure 22 : L'approche large du métadiscours

W.J. Vande Kopple (1985) est l'un des premiers à tenter de proposer une typologie du métadiscours. Comme illustré dans la figure précédente, la première catégorie constitue le métadiscours textuel. Cette catégorie vise à montrer comment un texte s'organise et comment les différentes parties sont reliées les unes aux autres. On retrouve au sein de cette catégorie une sous-division de quatre autres catégories dont les « connecteurs », utilisés pour montrer comment les parties sont reliées les unes aux autres. Cette catégorie inclut les marques de séquentialité (*premièrement, on va aborder, dans la deuxième partie nous allons voir, etc.*), de rappels (*l'argument décrit auparavant*), d'annonces (*comme nous pourrions le voir dans le dernier chapitre*). Puis, on retrouve une seconde catégorie appelée les marqueurs de gloses (« code glosses ») utilisée pour aider le lecteur potentiel à saisir le sens visé par l'auteur (*cela signifie que*) et afin de lui donner des indices pour l'aider dans son interprétation (*ce qui est contradictoire*). Viennent alors les « marqueurs de validité » à travers lesquels est exprimé l'engagement de l'auteur vis-à-vis du degré de vérité de ces énoncés. Parmi ces marqueurs, nous avons les atténuateurs ou enclosures (*peut, peut-être*), les emphatiques (*clairement, sans doute*), et les attributeurs²⁶ qui renforcent le positionnement de l'auteur à l'aide d'appui d'autres auteurs (*citations*). Pour finir avec cette première catégorie textuelle, on retrouve les « narrateurs » qui visent à informer le lecteur potentiel quant à ceux qui ont dit ou écrit quoi (ex. forme passive, et les guillemets).

Quant à la catégorie « interpersonnel », elle est constituée des « marques d'illocution » qui ont pour but de mettre en évidence un acte discursif spécifique tel qu'une hypothèse, exemples, conclusion (*pour conclure*). La seconde catégorie s'intitule les « marqueurs

²⁶ Appelé en anglais *attributers*.

d'attitude ». Elle est utilisée pour exprimer l'attitude de l'auteur vis-à-vis du contenu propositionnel qu'il présente (*surprenant, malheureusement*). Les « commentaires », qui constituent la dernière catégorie appartenant à la classe « impersonnel », consistent à s'adresser directement au lecteur potentiel en tentant d'établir un dialogue implicite en commentant son éventuelle réaction vis-à-vis du texte (*vous allez certainement adhérer à, vous ne devrez pas être d'accord avec ça*).

Toutefois ces catégories restent vagues et confuses en termes de fonction car si on reprend l'exemple donné par K. Hyland (2005) sur les citations, on constate que celles-ci sont effectivement utilisées pour renforcer un positionnement en prenant appui sur d'autres auteurs ce qui se traduit par l'usage des marqueurs de validités. Cependant, les citations peuvent également être utilisées comme sources d'informations ce qui les classerait alors dans la catégorie des *narrateurs*.

A. Crismore et al. (1993) reprennent cette typologie et réorganisent ce modèle comme nous pouvons le voir à travers la figure²⁷ en annexe 3.

Ils commencent par ajouter une catégorie appelée « interprétative » à la catégorie de marqueurs textuels et ce dans le but de procéder à la division entre la partie qui vise à aider le lecteur-cible à organiser le discours et celle qui va l'aider pour une meilleure compréhension et interprétations du sens-cible et des stratégies de l'écrit (A. Crismore et al. 1993).

Quant au modèle introduit par K. Hyland (2005)²⁸ (voir figure 23), il y distingue deux catégories spécifiques ; « interactive » et « interactionnelle ». Ce modèle propose une approche plus précise en introduisant les marques de positionnement et d'engagement. D'une part, la catégorie « interactive » vise à guider le lecteur-potential à travers le texte. Cette catégorie textuelle se constitue des « connecteurs logiques » aussi appelé « transitions » (*mais, de plus, donc*), des « marqueurs de cadre » qui font référence à l'acte discursif, aux séquences ainsi qu'aux étapes (*Pour conclure, enfin, mon but est*), les « marqueurs endophoriques » (*noté précédemment, dans la section 3, voir Figure x*),

²⁷ Voir figure qui se traduit en français par Catégorisation du Metadiscourse par A. Crismore et al. (1993, 47-54) en annexe 3.

²⁸ Voir figure 23 qui se traduit en français par Le Modèle Interpersonnel du metadiscourse (K. Hyland, 2005, 49) en annexe 3.

les « marqueurs évidentiels²⁹ » qui renvoient aux informations d'autres textes (*d'après X, Y*) et les « codes glosses » qui contribue à élaborer le sens propositionnel (*tel que, en d'autres termes*).

D'autre part, la catégorie « interactionnelle » consiste à impliquer le lecteur-potential dans le texte et ce en exposant son point de vue explicitement et en tentant d'engager le lecteur-potential dans un dialogue, en anticipant ces réponses vis-à-vis du texte. Cette catégorie se compose des « enclosures », « emphatiques/boosters » qui ont pour but de renforcer la certitude et d'engager un dialogue proche, des « marques d'attitudes » qui exprime l'attitude de l'auteur vis-à-vis de la proposition (*malheureusement, nous adhérons*), (auto)marqueurs personnel (*je, nous, mon, notre*), et les marques d'engagement/relationnelle qui tente d'établir explicitement une relation avec le lecteur potential (*considérez, notez, vous pouvez constater/voir que*). A travers cette catégorisation, on constate qu'effectivement il y a eu une révision du modèle proposé par A. Crismore et al. (1993). En revanche, il est évident que ce modèle proposé par K. Hyland reste très similaire à celui proposé par W.J. Vandepol comme nous pouvons clairement le constater à travers les deux figures suivantes (figure 23 & figure 24).

²⁹ Les marqueurs évidentiels sont appelés « evidentials » par K. Hyland (2005). Nous avons emprunté cette appellation en français à F.Grossmann et A. Tutin (2010) qui rappellent que « ces marqueurs permettent de signaler les éléments empiriques sur lesquels se fonde une assertion, et la manière dont ils ont été obtenus (par témoignage direct, par inférence, ou à travers les propos d'autrui) ».

Catégorie	Fonction	Exemples
Interactive	Aide le lecteur en le guidant à travers le texte	
Transitions / Connecteurs logique	Expriment une relation sémantique entre les clauses principales	<i>En outre, mais, donc, et, etc.</i>
Marqueurs de Cadre	Font référence aux actes discursifs, séquence et niveau de texte	<i>Pour conclure..., le but ici est de...</i>
Marqueurs endophoriques	Font références à des informations à d'autres parties du texte	<i>Noté précédemment, voir figure section X</i>
Marqueurs évidentiels	Font référence à des sources d'information dans d'autres textes	<i>Selon XX (1998)</i>
Marqueurs de gloses	Elabore un sens propositionnel	<i>En d'autres termes,</i>
Interactionnel	Permet à l'auteur d'interagir avec son lecteur potentiel	
Enclosures	aide le lecteur potentiel à saisir le sens visé par l'auteur	<i>Peut-être, possible</i>
Emphatiques	Souligne la certitude et un dialogue étroit avec le lecteur	<i>Il est clair que, en effet,</i>
Marqueurs d'attitudes	Exprime l'attitude de l'auteur vis-à-vis de la proposition	<i>Malheureusement, nous adhérons, étonnamment</i>
Marqueurs d'engagement	Exprime la référence à l'auteur	<i>Considérer, noté que, vous pouvez voir que,</i>
Marqueurs personnel	Explicite la construction d'une relation avec l	<i>Je, nous, mon, notre</i>

Figure 23 : Le Modèle Interpersonnel du metadiscourse (K. Hyland, 2005, 49)

Catégorie	Fonction	Exemples
Metadiscours textuel		
Connecteurs	Permettent de voir comment les différentes parties sont reliées entre elles. Permettent de faire des rappels. Permettent d'attirer l'attention sur le sujet d'un segment de texte.	<ul style="list-style-type: none"> - <i>Premièrement, dans la deuxième partie</i> - <i>Comme indiqué dans le chapitre XX,</i> - <i>en ce qui concerne, dans le cadre de</i>
Marqueurs de glose	Utilisé pour aider les lecteurs à comprendre l'orientation du sens de l'auteur Selon l'évaluation de l'auteur de la connaissance du lecteur, ces dispositifs, expliquent, définissent ou précisent le sens d'un usage.	<ul style="list-style-type: none"> - <i>Pour conclure..., le but ici est de...</i>
Marqueurs de validité	Utilisé pour exprimer l'engagement de l'auteur à la probabilité ou la vérité d'une déclaration. Ceux-ci comprennent des enclosures, emphatiques, et les attributeurs qui renforcent une position en affirmant le soutien d'un autre auteur	<ul style="list-style-type: none"> - <i>peut-être, pourrait, peut (enclosures)</i> - <i>clairement, sans aucun doute (emphatiques)</i> - <i>selon Einstein (attributeurs)</i>
Narrateurs	Utilisé pour informer les lecteurs de la source de l'information présentée- qui a dit ou écrit quelque chose	<ul style="list-style-type: none"> - <i>le premier ministre a annoncé que</i>
Metadiscours interpersonnel		
Marqueurs d'illocution	Utilisé pour rendre explicite l'acte discursif d'un auteur qu'il est entrain d'effectuer.	<ul style="list-style-type: none"> - <i>pour conclure, je fais l'hypothèse, pour résumer, on prédit</i>
Marqueurs d'attitude	Utilisé pour exprimer l'attitude de l'auteur face aux propositions qu'il présente	<ul style="list-style-type: none"> - <i>malheureusement, fait intéressant, je souhaite que</i>
Commentaires	Utilisé pour traiter directement les lecteurs, les entraînant dans un	<ul style="list-style-type: none"> - <i>vous serez certainement</i>

Figure 24 : Système de classification du métadiscours par W.J. Vande Kopple (1985,82-92)

Malgré la différence d'appellation attribuée aux différentes catégories déterminées par chacun de Van Kopple (1985) et K. Hyland (2005), ce dernier a repensé la distinction entre les deux niveaux de métadiscours qu'il a appelés *interactif* et *interpersonnel*, et a affiné la classification avec les catégories de *marqueurs personnels et d'attitude*. Le modèle de K. Hyland (2005) se rapproche donc de celui de Van Kopple.

Cette conception large du métadiscours a longtemps été utilisée dans les études. D'autres, en revanche, conçoivent la fonction métadiscursive différemment car ils considèrent que cette conception n'est pas suffisamment rigoureuse et requiert une sous-classification.

C'est notamment le cas de R. Markkanen et al. (1993, 143 cité in A. Adél, 2006, 171) qui considèrent qu'une catégorisation plus stricte est nécessaire car ils soutiennent que le fait de dire que certains items dans un texte sont métadiscursifs reste insuffisant. Un autre modèle, se fondant sur une conception plus étroite, s'est alors développé comme nous pourrions le voir dans ce qui suit.

2.3.3.2. L'approche étroite ou non-intégrative

Les études portées sur le métadiscours réalisées entre autres par ; A. Mauranen (1992 ; 1993a; 1993b), Bunton (1999), T. Dahl (2004), Valero-Garces (1996), Moreno (1997; 2004) et Peterlin (2005) s'inscrivent en effet dans une vision plus étroite du métadiscours, c'est-à-dire dans l'approche non intégrative. Ces auteurs considèrent que la délimitation entre le métadiscours textuel et interpersonnel reste floue. C'est pourquoi ils considèrent qu'il faut revoir cette catégorisation en termes sémantique et pragmatique.

Le modèle introduit par A. Mauranen (1992, 1993b) et A. Adél (2006) appartient à ce courant car ils considèrent que tout métadiscours est interpersonnel. D'après ces auteurs, une distinction entre le métadiscours et d'autres catégories linguistiques devient nécessaire, notamment, celles entre le métadiscours et le positionnement. Car dans la tradition large du métadiscours, l'attitude de l'auteur vis-à-vis de ce qui se dit est un élément essentiel. Son positionnement se traduit à travers les marques de validité et d'attitude (W.J. Van Kopple, 1985). Comme nous l'avons mentionné précédemment, cette première catégorie présente le contenu propositionnel de l'énoncé de l'auteur en termes de vérité probable. Les marques de validité (enclosures, emphatiques, attributeurs) traduisent le degré d'engagement de l'auteur par rapport à la vérité de son énoncé. La seconde catégorie correspond à l'attitude de l'auteur vis-à-vis du contenu propositionnel

qu'il présente (*pour conclure, en d'autres termes, etc.*). Ainsi, la différence entre ces deux types de marques discursives à travers lesquelles le positionnement peut s'exprimer, se trouve dans le fait que les marqueurs de validité expriment une modalité épistémique alors que les marques d'attitude expriment une modalité déontique, i.e. marquent l'attitude de l'auteur vis-à-vis des facteurs sociaux d'obligation, de permission et de responsabilité (A. Adél, 2006, 174). Ainsi, l'approche étroite du métadiscours, illustrée dans la figure 25, exclut le positionnement en tant que forme du métadiscours.

Objet d'étude	Éléments linguistiques qui montrent comment le texte s'organise ou qui font référence au texte lui-même.
Exemples	<i>Pour conclure, en d'autres termes, brièvement, deuxièmement, etc.</i>
Fonction Dominante	TEXTUELLE

Figure 25 : L'approche étroite du métadiscours

Contrairement à l'approche large, l'approche étroite exclut la fonction interpersonnelle. De plus, l'appellation utilisée par ce courant pour aborder ce niveau discursif n'est plus métadiscours mais *métatexte*, notion introduite par Enkvist (1975). R. Markkanen et al. (1993) définissent ce métatexte comme tout élément qui a pour première fonction de décrire le texte dans lequel il figure (R. Markkanen et al. (1993), 142 cité in A. Adél, 2006, 175). Le métatexte sous cette acception décrit la structure et organisation du texte plus qu'il n'informe sur son contenu.

La pionnière de ce courant non-intégratif est A. Mauranen (1993). Cependant, elle ne qualifie pas cette fonction discursive de métatextuelle mais elle la caractérise plutôt par rapport à la « réflexivité du langage ». Elle met l'accent sur les commentaires explicites des auteurs dans leur texte. Elle introduit une échelle d'explication où les expressions réflexives sont classées suivant leur degré d'explicitation. Elle identifie ainsi d'une part, une classe traduisant une « réflexivité ayant un degré élevé d'explication ». Celle-ci se divise à son tour en trois sous classes ; référence au texte, étiquettes discursives, adresse au lecteur potentiel. D'autres part, elle définit une catégorie dont la réflexivité a « un degré bas/moins d'explicitation ». Cette catégorie se divise en « connecteurs internes » (*premièrement, deuxièmement, etc.*), étiquettes discursives (*nos présentes données*

montrent), référence au texte (*comme première étape*), éléments permettant de s'adresser au lecteur potentiel comme illustré dans la figure 26.

Réflexivité d'un niveau élevé d'explication	
Référence au texte	<i>dans cet article, dans la section suivante</i>
Etiquette de discours	<i>pour illustrer, comme indiqué précédemment, cet argument</i>
Référence au lecteur	<i>notez que, rappel au lecteur</i>
Réflexivité d'un niveau bas d'explication	
Connecteurs internes	<i>cependant, deuxièmement, aussi</i>
Etiquettes de discours	<i>question, il est raisonnable de penser, notre présente donnée montre</i>
Référence au texte	<i>maintenant, pour première étape</i>
Référence au lecteur	<i>il y a une raison pour se rappeler</i>

Figure 26 : Modèle réflexif du métadiscours (A. Mauranen, 1993)

Ce modèle non-intégratif de A. Mauranen exclut les catégories qui relèvent de questions rhétoriques car elle considère que celles-ci participent de la relation entre auteur et lecteur plutôt que de celle de l'auteur vis-à-vis de son texte. En revanche, la dernière catégorie qui prend en compte les adresses au lecteur-potentiel, est considérée comme étant métadiscursive. D'ailleurs Baklund (1998) et Valero-Garcès (1996) reprennent cette catégorie dans leur typologie bien que les défenseurs de ce courant stipulent que les marques linguistiques dont la fonction est interpersonnelle, sont exclues car ils prennent en compte uniquement les références métatextuelles au texte cible (cité in A. Adél, 2006, 177). On constate alors la difficulté à vouloir délimiter ce qui est métadiscursif de ce qui ne l'est pas. Mais surtout, nous rejoignons A. Adél (2006) pour soulever l'incohérence qui consiste à considérer que les marques linguistiques faisant référence au lecteur potentiel se trouvent être métadiscursives alors que celles référant à l'auteur ne le sont pas.

T. Dahl (2004) s'inscrit dans l'approche étroite. Toutefois, elle n'adopte pas le modèle introduit par A. Mauranen dans son traitement du métadiscours dans les articles de recherche. Ces derniers sont publiés en français, anglais et norvégien dans les disciplines d'économie et médecine. Elle détermine deux catégories métatextuelles, « locationnelle »

et « rhétorique ». Dans la première est intégré tout élément linguistique faisant référence au texte lui-même (ex. *la présente étude, dans la section 3*).

Quant à la seconde, elle traite des éléments qui permettent au lecteur potentiel d'interpréter les actes rhétoriques de l'auteur au cours du processus d'argumentation (T.Dahl, 2004, 1811-1812), grâce à des marqueurs tels que *on conclut que*, etc.

A. Adél (2006) adhère également à cette tradition du métadiscours. Elle partage l'idée qu'il est préférable de s'en tenir à une conception étroite du métadiscours. Cependant, elle est consciente des limites de ce modèle, du fait qu'il ne prend pas en compte l'auteur et le lecteur potentiel dans la situation de l'écrit. C'est pourquoi, elle propose un autre modèle basé sur la vision étroite mais, où les phénomènes métadiscursifs sont définis comme faisant référence au texte cible et/ou au processus d'écriture, prenant ainsi en compte également l'auteur et le lecteur potentiel. Elle définit pour ce faire une autre catégorie qu'elle nomme « métadiscours auteur-lecteur » potentiel. Son modèle se base sur les trois fonctions jacobsoniennes (1998) ; *métalinguistique, expressive et directive*³⁰.

La catégorie métatextuelle relèverait de la fonction métalinguistique (code/texte) en prenant en compte l'organisation du texte. Cette catégorie inclut la réflexivité métatextuelle explicite à un degré élevé et moindre. C'est-à-dire, dans le premier cas, ce type de réflexivité du langage inclut les expressions qui font référence explicitement au texte, à son écriture et à l'organisation du langage. Concernant la *référence au texte*, elle se traduit dans le processus de l'écrit à travers l'usage de noms, d'adverbes de temps ou de manière qui ont pour fonction de faire référence à tout le texte où uniquement à des parties du texte comme nous pouvons le voir dans les exemples ci-dessous :

- (1) *Dans cet article,*
- (2) *Pour tous les résultats de cette section, une signification de 90%, 95%.....*
- (3) *Maintenant, nous avons vu.....*

Pour ce qui est de *l'étiquette discursive*, reprise des travaux de Mauranen dans la catégorie de réflexivité à degré élevé en termes d'explication, elle implique l'usage des adverbes d'illocution (*pour résumer*), de noms d'illocution (*en résumé*), d'adverbes de

³⁰ Toutefois, le métatexte et le métadiscours ne font pas référence dans ce modèle aux fonctions du langage utilisées, mais uniquement aux catégories identifiées. En effet, elle considère que parler de fonctions dans ces cas de figure serait infondé.

manière ainsi que toutes les expressions similaires qui ont pour but de montrer l'acte discursif réalisé soit dans l'environnement immédiat ou dans la progression du discours (N. Toumi, 2009, 69). Certains adverbes ne sont pas illocutionnaires mais montrent une certaine relation avec le texte cible. La relation se manifeste alors à travers la réflexivité sur une partie du texte, ou du lecteur potentiel ou de l'auteur (l'article *est organisé* comme suit).

Quant aux marqueurs phoriques, ils sont considérés comme les éléments permettant au lecteur potentiel de naviguer à travers le texte. Ces expressions ont souvent pour fonction un « changement de thème » (*en ce qui concerne* le résultat obtenu dans ce travail), « aperçus » (mais le système est plus complexe comme on peut le voir *plus tard*), commentaires, synthèses.

D'autre part, on retrouve la réflexivité métatextuelle à un degré moindre d'explication. Cette catégorie concerne les expressions dans le texte qui font référence au texte et aux fonctions des parties qui le constituent de manière implicite. Cette catégorie implique alors des catégories plus vagues, en d'autres termes des expressions qui restent ambiguës dans leur référence (référence au texte ou au contenu propositionnel). Parmi ces marques linguistiques, on retrouve les connecteurs internes, étiquettes discursives, référence au texte et les codes glosses (A. Mauranen, 1992 ; 1993b & N. Toumi, 2009). Ces éléments sont classés dans cette catégorie car ils ne permettent pas de faire référence de manière explicite au texte du/sur le texte ou à son processus d'écrit ou de lecture.

Les connecteurs internes représentent les éléments dont la fonction consiste à exprimer une relation entre les éléments du discours interne tels que les contrastes (*cependant*), addition (*de plus*), généralités (*en général*) et les marques de séquentialité (*premièrement*). Cependant, les connecteurs internes qui reflètent un processus de conceptualisation interne à la *réflexion (mind) de l'auteur* tels que les connecteurs temporels, causal, de conséquence, ne sont pas considérés comme relevant du métadiscours réflexif.

Alors que la catégorie métadiscursive *d'auteur-lecteur potentiel* marque l'interaction entre ces deux acteurs, ce modèle permet d'intégrer l'usage de la réflexivité par chacun de ces deux acteurs « le producteur et le récepteur du texte ».

Ainsi, bien que le métadiscours tend à être défini comme texte du/sur le texte, Mauranen (1993) et A. Adél (2006) font une distinction entre les références « intertextuelles » et

« métatextuelles » car elles considèrent que c'est uniquement cette seconde référence qui relève du métadiscours du fait qu'elle fait référence au texte cible contrairement à la première qui pourraient également faire référence à des textes externes.

C'est notamment le cas du discours rapporté exclu de cette vision et considéré comme étant en dehors du cadre métadiscursif. Comme nous l'avons mentionné précédemment, les marques du positionnement sont également perçues comme appartenant à un cadre externe au métadiscours.

Ainsi, la différence entre l'approche large et étroite du métadiscours réside dans le fait que la première (intégrative) intègre des marques de positionnement, alors que la seconde (non-intégrative) se veut plus étroite en prenant en compte uniquement les marques textuelles faisant référence au texte cible.

Le métatexte sous l'acceptation de l'approche étroite du métadiscours décrit la structure et organisation du texte plus qu'il n'informe. Comme nous avons pu le voir précédemment, cette approche exclut la fonction interpersonnelle. Bien que dans cette même tradition, T. Dahl (2004) détermine une autre catégorie qu'elle qualifie de « rhétorique » qui traite des éléments qui permettent au lecteur potentiel d'identifier les actes rhétoriques de l'auteur dans son processus d'argumentation (*on conclut que*, etc.) ; elle ne prend pas en compte comme A. Adél (2006) l'auteur et le lecteur potentiel dans la situation de l'écrit. Etant donné l'importance que revêt pour notre problématique la relation au lecteur, nous avons fait le choix de nous inscrire dans l'approche large du discours, dans la lignée des travaux de K. Hyland.

Certes, il est important de procéder à une analyse rigoureuse du métadiscours, et l'analyse serait plus simple si nous adoptions l'approche non-intégrative de manière à limiter notre analyse au métatexte ou aux fonctions réflexives du métadiscours. Cependant, l'approche non-intégrative est susceptible de masquer la complexité du phénomène de métadiscours dans le monde du discours.

En outre, A. Adél avance l'argument selon lequel les marqueurs d'attitude et de validité devraient être complètement exclus de l'analyse du métadiscours en raison de leur possibilité de révéler les attitudes de l'auteur à l'égard du monde extérieur. Il convient peut-être d'exclure les entités d'évaluation dont les référents sont dans le monde réel, mais il peut être inapproprié d'exclure les expressions évaluatives de l'analyse lorsque

ces marqueurs du métadiscours montrent les attitudes des auteurs vis-à-vis du contenu asserté.

De plus, les « formules de discours », telles que nous tentons de les définir, et qui se trouvent en dépendance syntaxique avec d'autres éléments linguistiques recensés dans notre lexique des écrits scientifiques peuvent être décrites :

- soit comme des « marqueurs de validité » à travers lesquels est exprimé l'engagement de l'auteur vis-à-vis du degré de vérité de l'énoncé
- soit comme des « marques d'attitude » de l'auteur vis-à-vis de l'énoncé.

Tous ces éléments plaident pour leur intégration au sein de l'approche large du métadiscours. Rappelons qu'au vu de notre postulat de départ à savoir que l'expression du problème de recherche se traduit notamment par l'usage d'associations lexicales qui interviennent à méta-niveau, nous inscrivons donc au sein de courant large car nous recensons des marqueurs de validité et d'attitude parmi les éléments linguistiques constituant nos FD.

2.3.4. Les facteurs qui influencent l'usage et la distribution des marques métadiscursives

Plusieurs facteurs extra-linguistiques peuvent avoir un impact sur l'usage et la distribution des marques métalinguistiques. Nous allons présenter successivement chacun de ces facteurs sur la constitution du métadiscours, à savoir le genre, la communauté de discours, la dimension culturelle et voir dans quelle mesure ceux-ci peuvent influencer l'usage et la distribution des marques du métadiscours.

2.3.4.1. Le métadiscours et le genre

Le premier qui a tenté de préciser la relation entre le métadiscours et la variation de genre est W.J. Vandekopple (1985, 88). Il s'est demandé si certains types de métadiscours étaient plus appropriés que d'autres en fonction du type de textes étudié³¹. A. Crismore (1989, 61) s'est également intéressé à cette question alors qu'il visait à étudier la variation métadiscursive dans les textes. Pour ce faire, il introduit la notion de continuum métadiscursif (« metadiscourse continuum »). Il considère que les marques

³¹ « Are some kind of metadiscourse more appropriate than others-or even necessary- in some kinds of texts ? ».

métadiscursives utilisées varient en fonction du genre. Pour justifier son approche, il prend pour appui l'exemple de deux types d'écrits, les textes pédagogiques et les documents du droit. Il affirme que dans le premier genre d'écrit, il est courant de voir apparaître l'auteur ou le lecteur potentiel alors que dans le second type il n'est pas approprié pour l'auteur d'apparaître. J.M. Swales (1990), considéré, dans le contexte anglo-saxon, comme l'un des pionniers de l'analyse du genre (*Genre Analysis, GA*) définit la notion de genre comme un événement communicatif qui est spécifié par une série de buts communicatifs et de traits reconnus et propre aux membres d'une communauté donnée. En effet, les textes peuvent être classés dans un genre particulier ou un autre en fonction des traits linguistiques et rhétoriques. Une des dimensions clé de l'analyse du genre est le métadiscours. En effet, ce dernier fait partie de ces traits car les textes peuvent être classés et répertoriés en fonction des différents moyens d'interactions qu'on tend à créer avec son lecteur ainsi qu'à travers les différents moyens de persuasion utilisés par l'auteur. Le facteur de persuasion par exemple a été comparé lors d'une étude réalisée en 2004 par K. Hyland entre des écrits académiques et des textes éditoriaux. Cette analyse a montré que bien que chaque texte vise à persuader les lecteurs, les auteurs utilisent le métadiscours différemment.

L'usage des marqueurs évidentiels³², par exemple, dans les articles académiques auxquels font appel les auteurs a pour but de référer à des travaux antérieurs ou à des travaux dans le même domaine. Alors que dans les textes éditoriaux dont *Le Monde*, ces marqueurs sont utilisés afin de mettre l'accent sur le sérieux/crédibilité du journal, son élitisme ainsi que sa liberté d'esprit (K. Hyland, 2005, 89).

2.3.4.2. Le métadiscours et la communauté de discours

La communauté de discours (« discourse community ») est considérée par J.M. Swales (1990, 24-27) comme une communauté dont les membres partagent des objectifs publics communs, ainsi qu'un système communicatif interne. Elle est également perçue comme ayant acquis une terminologie spécialisée. Becher (1989) la considère comme étant une tribu qui détient ces propres normes et ces conventions. On retrouve également cette notion de communauté chez K. Hyland (1999, 5) quand il aborde le métadiscours. Il soutient l'idée que le métadiscours utilisé par un auteur dans son texte est un moyen pour

³² E. Le (2004) considère que ces marqueurs évidentiels indiquent des sources d'information et qu'ils apparaissent dans *Le Monde* dans une variété de formes : Nous, discours rapporté, et d'autres.

représenter les caractéristiques de la communauté à laquelle il appartient car les disciplines académiques ont des façons distinctes, rhétoriques et conventionnelles, de se développer. En effet, les règles, normes et conventions rhétoriques typiques de toute discipline académique ou autres influencent l'auteur dans sa manière d'aborder son texte, de se représenter en tant que « persona » et de représenter son lecteur et ces attitudes. Cet auteur se voit également dans l'obligation de prendre en considération son lecteur potentiel dans son entité. C'est-à-dire qu'il se doit de prendre en considération ces normes, attentes et réponses en fonction de la communauté à laquelle il appartient et ce dans le but de construire un écrit persuasif (K. Hyland, 2005). Ainsi, l'écrit d'un auteur qui se veut persuasif dépend en grande partie du degré de maîtrise des marques métadiscursives et de la connaissance de l'existence de ces différents paramètres.

2.3.4.3. Le métadiscours et la culture

Il est évident que les facteurs culturels sont à prendre en compte car ces derniers influencent notre compréhension. De ce fait, ils influencent probablement notre façon d'écrire et celle d'organiser notre écrit. Les valeurs culturelles affectent notre perception, langage, apprentissage, communication et plus particulièrement l'usage du métadiscours (K. Hyland, 2005, 113-115).

D'ailleurs, K. Hyland (2005) considère que, ce qui fait qu'un texte est bien organisé et cohérent, varie d'une culture à l'autre. Dans une culture d'« écrit responsable » (« a writer responsible culture ») comme l'anglais, les marqueurs métalinguistiques peuvent-être utilisés afin de guider le lecteur potentiel à travers le texte. Alors que dans une culture de « lecteur responsable » (a reader responsible culture) comme le japonais, les connections entre différentes parties du texte restent implicites (A. Adél, 2006, 149). Chaque culture possède, en effet, ses normes, langage, valeurs ainsi que son mode de communication.

2.3.5. Vers une définition des formules de discours

Comme nous avons pu le dire précédemment (page 44), les FD véhiculent des contenus plus larges que les associations lexicales étudiées entre autres par K. Hyland 2007, M. Pecman 2005, A. Tutin 2007, D. Siepman 2007, F. Boch et.al. 2007, mais elles ne sont pas pour autant complètement dissociables de ces expressions définies par ces auteurs.

La notion de *formule de discours* utilisée ici se définit à un triple niveau :

- au plan pragmatique, comme un élément mettant en évidence, pour le lecteur, le fait que la question de recherche va être explicitée dans l'énoncé ; cet élément se traduit au plan cognitif (ou conceptuel) comme la spécification d'une « opération mentale » parmi celles qui sont mobilisées par la rédaction d'un article, comme par exemple ici « *annoncer un problème* », « *exprimer un contraste d'idée* »
- au plan conceptuel, comme un élément constitutif des macro-catégories mobilisées à travers l'opération mentale considérée, comme par exemple le renvoi par l'auteur à son travail de recherche (*étude, texte, publication*) ;
- cette structure conceptuelle s'incarne elle-même dans une structure syntaxique précise au moment de l'actualisation en discours.

La notion de *formule de discours* telle que nous l'entendons, a donc un double statut dans la mesure où elle fait correspondre des unités linguistiques aux étiquettes liées à la nature de l'opération mentale. Par exemple, un début d'énoncé comme celui figurant en (1)

(5)-R³³ *Le but de cet article est de montrer....*

est structuré, suivant les propositions du concept-matching selon les trois niveaux :

- niveau conceptuel : ANNONCE PROBLEMATIQUE
- sous-niveau conceptuel : BUT – PUBLICATION ;
- niveau linguistique et énonciatif : actualisation du BUT de la PUBLICATION (*Le but de la publication.*)

C'est à ce troisième niveau que se fait l'ancrage référentiel et énonciatif, à l'aide notamment de déictiques (ex. *cet article*).

Ces FD ne sont pas nécessairement des collocations car elles ne sont pas forcément contiguës, et leurs éléments constitutifs ne sont pas obligatoirement en relation de dépendance syntaxique.

Dans la plupart des articles de recherche, le problème de recherche n'est pas annoncée par une expression spécifique, mais elle est élaborée au cours d'un processus argumentatif. Par conséquent, elle peut être construite à travers différentes stratégies rhétoriques, comme poser une hypothèse, démontrer une contradiction, etc. Suivant la

³³ Précisons que « R » fait référence à la notion de résumé.

macro-structure rhétorique de l'article scientifique, en formulant sa problématique, l'auteur accomplit des actes rhétoriques variés. Le modèle « Créer un Espace de Recherche » (J.-M. J.M. Swales 1990) – l'un des modèles les plus répandus – segmente le développement des introductions des articles scientifiques en trois « *mouvements*³⁴ » (moves) – *établir son territoire, établir une niche, occuper la niche* – dont chaque *étape* – par exemple *revue de la recherche précédente, poser une question, annoncer les buts* – accomplit des fonctions discursives différentes qui contribuent à la construction de l'« espace de recherche », c'est-à-dire, à la construction de la problématique scientifique.

Inspirés par le modèle de J.M. Swales, dans des travaux empiriques, S. Teufel (1999) et S. Teufel et al. (2002) proposent une méthode pour détecter automatiquement les phrases qui articulent les mouvements rhétoriques et leur fonction *rhétorique* : *but, connaissances préalables, solution de l'auteur, contraste avec d'autres approches, faiblesse d'autres approches, bases de la recherche*. Les modèles d'organisation textuelle axés sur les mouvements rhétoriques de J.M. Swales (1981 ; 1990 ; 2004) se trouvent être des outils permettant d'explorer l'organisation rhétorique interne des textes scientifiques. Quant aux fonctions rhétoriques, elles sont identifiées par un algorithme qui prend en considération des indices variés, tels que la position de la phrase, le lexique, la phraséologie scientifique, etc.

A la différence de cette approche, notre but n'est pas de mettre en évidence le développement rhétorique de la problématique des articles, mais de révéler les indices linguistiques qui renvoient à des problèmes de recherche en général, y compris dans les phrases qui ne sont pas des constituants directs de la macro-structure rhétorique prédéfinie par les modèles de l'argumentation rhétorique.

De façon prototypique, le problème de recherche est mentionné dans les phrases où l'auteur résume l'article entier ou bien une partie de l'article, comme dans l'exemple suivant où la partie en italique constitue une fonction de *résumé* :

(5)-R *Le but de cet article est de montrer, sur la base d'une étude de la construction du jugement des enseignants, les fondements méthodologiques à l'œuvre.*

Cette suite d'associations spécifiques, en relation syntaxique représentant le concept de résumé et annonce le problème de recherche traité.

³⁴ Nous conservons la traduction littérale « mouvement », qui a l'avantage de bien montrer la successivité des opérations considérées. Nous réserverons l'expression « fonction rhétorique » à un niveau plus abstrait, qui peut prendre en compte différents « mouvements ».

En dehors des résumés, on repère également des phrases où l'auteur exprime l'annonce du problème de recherche - une contradiction, une tension, une lacune, une surprise, etc. - dans la réflexion scientifique. Nous appellerons cet acte rhétorique *annonce de la problématique*. La phrase suivante en est un exemple. L'annonce du problème de recherche dans la réflexion scientifique est véhiculée par la partie en italique :

(1)-CC *La recherche a mis au jour la réalité d'un effet-maître (Bressoux, 1994) mais n'est pas encore en mesure de l'expliquer* de façon précise.

Lors des travaux antérieurs (A. Aït Saïdi et Á. Sándor à paraître), nous avons défini les expressions véhiculant les actes de *résumé* et de *changement cognitif* - comme les parties en italique des phrases (5)-R et (1)-A en tant que des *formules de discours*³⁵. Les formules de discours sont plus complètes à un niveau conceptuel que les marqueurs de métadiscours (par ex. K. Hyland 2007), les collocations ou unités phraséologiques (par ex. M. Pecman 2005), ou encore les marqueurs de discours (D. Siepmann 2007, F. Boch et al. 2007). Nous consacrons le chapitre 2 à expliquer et exemplifier les FD, de manière à montrer en quoi elles se rapprochent et diffèrent des marqueurs de discours tels qu'ils sont définis par (M. Pecman 2005, K. Hyland, 2007, D. Siepmann 2007 et F. Boch, 2007).

2.3.5.1. Description des formules de discours

Dans notre article paru en 2015 (A. Aït Saïdi et Á. Sándor), nous considérons comme FD, toutes les associations lexicales traduisant un problème de recherche dont les éléments constituants détiennent une fonction conceptuelle donnée (PUBLICATION, OPERATION MENTALE, OBSTACLE/CONTRASTE, etc.).

Ces formules qui se trouvent en dépendance syntaxique communiquent le statut des assertions scientifiques.

Considérons les deux phrases (5)-R et (6). La phrase (5) contient une formule de discours *résumé* et la phrase (6) associe *résumé* et *changement cognitif*:

(5)-R *Dans la deuxième partie, nous présenterons, sur la base d'une enquête et d'observations menées récemment auprès d'enseignants en formation, le positionnement « classique » et spontané relativement à la question de l'erreur.*

³⁵ Rappelons que les formules de discours sont des suites de mots ou d'expressions syntaxiquement cohérentes, mais pas nécessairement contiguës, aux structures syntaxiques et composantes lexicales variées.

(6)-R *Nous entendons dans ce texte analyser ce débat*, dans un premier temps en présentant les protagonistes, les agendas qu'ils épousent et le contexte de concurrence entre les groupes en présence pour le contrôle de la politique éducative américaine.

Notre démarche, pour modéliser les FD consiste à reconnaître des patrons sous-jacents communs aux expressions de *résumé* - comme (5)-R, (6)-R, (7)-R - d'une part, et *annonce de la problématique* - comme (1)-CC, (2)-CC, d'autre part.

(1)-CC *La recherche a mis au jour ... mais n'est pas ... l'expliquer*

(2)-CC *.....ce cadre théorique, nous proposons notre problématique en termes 'étude des questions*

(5)-R *Le but de cet article est de montrer*

(6)-R *Dans la deuxième partie, nous présenterons*

(7)-R *Nous entendons dans ce texte analyser*

Dans le modèle, les mots et les expressions dans les FD sont des réalisations des concepts constituants³⁶. Par exemple,

(1)-CC *recherche* correspond au concept *PORTEE*

(1)-CC *a mis au jour, mais, n'est pas, débat* correspondent au concept *OBSTACLE/CONTRASTE*

(5)-R *but, article, partie, texte* correspondent au concept *PUBLICATION*

(6)-R *nous, nos, ce* correspondent au concept *DEICTIQUE*

(7)-R *montrer, présenterons, analyser, a mis au jour, expliquer* correspondent au concept *OPERATION_MENTALE*

En dehors de ces concepts constituants, nous en avons identifié deux autres – *POSSIBILITE* et *IMPORTANCE* - que nous allons présenter dans la section 4.

Les FD sont modélisées comme les réalisations des différentes configurations des concepts constituants. Ainsi, la formule générale du *résumé* est construite dans (5)-R, (6)-R et (7)-R à travers les différentes configurations des concepts constituants *PUBLICATION*, *DEICTIQUE* et *OPERATION_MENTALE*, alors que celle d'*annonce*

³⁶ Nous revenons plus en détails dans la section 3.6. qui définit et répertorie ces constituants.

du problème de recherche dans (1)-CC et (2)-CC est construite à travers les différentes configurations des concepts constituants OPERATION_MENTALE, PORTEE et OBSTACLE/CONTRASTE.

(2) CC ...*ce cadre théorique, nous proposons,notre problématique.....l'étude des questions....*

RELATION_SYNTAXIQUE(cadre théorique[PORTEE],ce[DEICTIC])

RELATION_SYNTAXIQUE(nous[DEICT],proposer[OP_MENT])

RELATION_SYNTAXIQUE(notre[DEICT],problématique[CONTRAST])

RELATION_SYNTAXIQUE(étude[DEICT],questions[CONTRAST])

En plus du fait qu'elles sont composées de configurations des concepts constituants, l'autre caractéristique des FD est le fait qu'elles sont des expressions syntaxiquement cohérentes, c'est-à-dire que les mots et expressions qui les réalisent sont dans des relations syntaxiques les uns avec les autres comme nous pouvons le voir à travers l'exemple ci-dessous :

2.3.5.2. Lexique des concepts constituants

Des travaux récents portent sur l'analyse du lexique et de la phraséologie des écrits scientifiques - ou de spécialisation - en français, et en particulier concernant la *langue scientifique générale* (LSG) (M. Pecman 2004) qui est caractérisée par un *lexique transdisciplinaire-transversal* (A.Tutin 2010). Ce lexique de genre est perçu comme commun à la communauté scientifique et transcendant les disciplines, car il renvoie aux concepts mis en œuvre dans l'activité scientifique au-delà de la terminologie du domaine. Les études menées dans ce cadre ont abouti à des classifications du lexique transdisciplinaire (M. Pecman 2007, A. Tutin 2007) certes différentes de la nôtre, mais partageant avec elle aussi de nombreux aspects, car le lexique des concepts constituants appartient à ce lexique transdisciplinaire.

L'approche de M. Pecman (2007) est similaire à la nôtre en ce qu'elle établit une catégorisation des « notions clés » du LSG qui sont réalisées par des mots de nature différente mais similaires aux concepts constituants : les notions « se réfèrent à une action, un objet, une entité, une abstraction, une qualité, etc. de nature purement

scientifique. » (M. Pecman, 2007, 89). Une autre similitude entre les catégories de M. Pecman et notre système est qu'un même mot peut appartenir à plusieurs catégories.

Toutes les notions de M. Pecman ne sont pas couvertes par notre système, notamment ce qu'elle appelle la « sphère de l'universalité » (voir partie 1.4). La catégorisation de A. Tutin (2007) concerne les noms, et toutes les catégories sont couvertes par un ou plusieurs concepts constituants. Nous retrouvons également dans la catégorisation de D. Siepmann (2007) des éléments lexicaux couverts par nos concepts constituants.

LEXIQUE DES CONCEPTS CONSTITUANTS	EXEMPLES
<ul style="list-style-type: none"> - Marques linguistiques qui renvoient à un processus de l'activité scientifique, - Noms d'objet construit par l'activité scientifique, 	<ul style="list-style-type: none"> - noms : <i>question, absurdité, anomalie, débat, doute, trouble</i> - verbes : <i>contester, rejeter, clarifier, repenser</i> - adjectifs : <i>énigmatique, insuffisant, erroné, contradictoire</i> - conjonctions de coordination, adverbes : <i>mais, cependant, à l'opposé</i> <hr/> <ul style="list-style-type: none"> - noms : <i>réflexion, raisonnement, affirmation, réponse, révision, argument</i> - verbes : <i>analyser, étudier, vérifier, chercher, cerner, découvrir, relever</i> <hr/> <ul style="list-style-type: none"> - <i>analyse, travail, projet, recherche, approche, modèle, système, théorie, idée, données, résultat, point, question, hypothèse</i>
<p>Marqueurs linguistiques qui sont des supports de la rédaction scientifique,</p>	<ul style="list-style-type: none"> - noms : <i>article, auteur, documentation, Figure</i> - verbes : <i>présenter, exposer, résumer</i>
<p>Marqueurs d'atténuation</p>	<ul style="list-style-type: none"> - adjectifs modaux : <i>possible</i> - verbes : <i>permettre, pouvoir, supposer</i> - nom : <i>possibilité</i>
<p>Marqueurs personnels</p>	<ul style="list-style-type: none"> - pronoms personnels de 1ère personne : <i>je, nous</i>

Figure 27 : Illustration du lexique des concepts constituants

Nous revenons en détails dans les parties 3.6 et 3.7 qui s'intitulent *Hiérarchie des concepts constituants*, sur la catégorisations et classification du lexique illustré dans la figure 57 et sur notre choix des traits conceptuels attribués et leurs correspondances en s'inspirant des systèmes de A. Tutin (2007), M. Pecman (2007) et D. Siepmann (2007).

Conclusion

Tout au long de ce chapitre, nous avons pu mettre en évidence en quoi les formules discours traduisant un problème de recherche, véhiculent des contenus plus larges que les collocations, que les marqueurs de discours tels qu'ils sont notamment définis par D. Siepmann (2007) ou encore que les marqueurs du métadiscours tels qu'ils ont été conçus par K. Hyland (2005).

A la question de savoir si les associations lexicales donnant lieu aux concepts généraux de *changement cognitif* et de *résumé* étudiées dans ce présent travail sont des collocations, la réponse pourrait être que :

- ces associations lexicales véhiculent des contenus plus larges que les collocations ;
- ces collocations ne constituent pas à elles seules nos FD et ne peuvent être isolées ;
- elles sont en dépendance syntaxique avec d'autres éléments de notre lexique à traits spécifiques propres à cette étude.

A la question de savoir en quoi les FD diffèrent des marqueurs de discours polylexicaux, on peut également apporter les éléments de réponse suivants.

Les marqueurs polylexicaux peuvent être décrits comme étant des collocations ou des expressions figées [D. Siepmann (2005, 49)]. Or, les FD, telles que nous les définissons,

- sont des cooccurrences régulières ;
- le critère de figement n'est pas en ce qui les concerne un paramètre définitoire ;

Il reste à préciser en quoi les FD se rapprochent des marqueurs du métadiscours.

Si l'on accepte de s'inscrire dans l'approche large du métadiscours, on constate que plusieurs marqueurs du métadiscours appartiennent à notre lexique (marqueurs de logique-transition, d'atténuation, de cadre, personnel, attitudes, etc.). De plus, les associations lexicales telles qu'elles sont décrites dans cette présente étude relèvent bien d'un méta-niveau.

Toutefois,

- ces segments isolés ne correspondent pas forcément à nos FD,
- toutes les associations du corpus ne sont pas toutes des marqueurs du métadiscours,

- les segments constituant nos associations lexicales doivent d'être en dépendance syntaxique avec d'autres éléments.

Les FD représentent bien un objet original, qui mérite un traitement spécifique d'où notre intérêt pour ce sujet.

Chapitre 3 : Méthodologie et analyse de corpus

Introduction

Dans cette partie, nous décrivons la méthode à laquelle nous avons fait appel pour l'implémentation de l'analyse des FD dans un système informatique pour leur détection. Ainsi, afin de développer notre système de détection automatique des FD véhiculant le problème de recherche, nous procédons à l'implémentation d'un module appelé « le concept-matching » qui repose sur un formalisme qui décrit les différentes FD au-dessus de la fonctionnalité de Xerox Incremental Parser (XIP). Cette action qui s'effectue au sein de ce parseur syntaxique robuste se fait « au-dessus » de l'analyse linguistique générale.

Afin de développer ce système de détection de ces phrases où apparaissent ces associations lexicales discursives, nous avons constitué un corpus de référence à partir de la base de données Scientext. Puis, nous avons procédé à une annotation manuelle des phrases véhiculant le problème de recherche. Ce corpus a contribué au développement des règles de grammaires formelles qui ont permis de mettre en place un système d'identification des phrases-clé.

3.1. Méthodologie et Concept-matching

Notre étude s'inscrit dans les domaines de TAL, de la linguistique de corpus et de la recherche documentaire. Dans le traitement automatique du langage (TAL), il devient primordial de faire appel à des outils qui permettent une exploitation de corpus plus ciblée et plus fine (Silberztein & A. Tutin, 2005). En effet, le TAL permet de travailler sur des volumes textuels importants et variés en effectuant des analyses fines et complexes avec des recherches à différents niveaux grammaticaux, conceptuel, syntaxiques et à travers l'application de grammaires locales qui permettent de cerner des phénomènes très précis.

Afin de développer notre système de détection des FD nous avons utilisé un corpus d'analyses tiré du projet Scientext³⁷.

³⁷ <http://scientext.msh-alpes.fr/scientext14/?corpus=sctexts&langCorpus=fr>

Ce corpus d'écrits scientifiques variés a été établi au sein du Laboratoire de Linguistique des Langues Etrangères et Maternelles Appliquée (LIDILEM) de l'Université de Grenoble 3 – STENDHAL.

Il a principalement porté sur l'étude linguistique du positionnement et du raisonnement à travers la phraséologie, les marques énonciatives et les marques syntaxiques liées à la causalité.

Ce projet réunit des corpus de plusieurs disciplines ; sciences humaines (linguistique, psychologie, sciences de l'éducation, traitement automatique des langues), sciences expérimentales (biologie, médecine) et les sciences appliquées (électronique et mécanique) et sous plusieurs types de documents ; Article, communication, HDR, thèse.

Nous avons travaillé plus précisément sur un sous-ensemble du corpus, composé d'articles appartenant au domaine des sciences de l'éducation et la psychologie. Il est constitué entièrement d'articles rédigés en français et publiés en France. Ainsi, ces écrits sont uniformes du point de vue du genre.

Le corpus comporte 51 articles³⁸ au format XML issus de la base de données Scientext et compte 395 864 mots. Le choix des domaines est motivé par ma contribution en qualité de stagiaire au projet European Educational Research Quality Indicators (EERQI). C'est un projet de recherche du 7e programme cadre pour la recherche en sciences socio-économiques et humaines (SSH) de l'Union Européenne. Il avait pour but de renforcer et d'améliorer la visibilité mondiale et la compétitivité de la recherche européenne en éducation. Mon stage a été réalisé au sein du centre de recherche européen XEROX.

Mon sujet de thèse ainsi que le choix du domaine s'inscrivent donc dans la continuité de mes travaux au sein de Xerox. Comme il a déjà été précisé dans les chapitres qui précèdent, cette étude se situe dans la perspective d'une analyse rhétorique, conceptuelle et syntaxique de la structuration discursive des FD traduisant un problème de recherche dans les écrits scientifiques. Plus spécifiquement, nous proposons une analyse conceptuelle des structures rhétoriques du discours.

³⁸ N'ayant pas assez d'articles en sciences de l'éducation sur la plate-forme Scientext, nous avons fait le choix d'intégrer les communications issues de cette même discipline.

Nous avons adopté une démarche inductive à partir d'observations empiriques qui nous permettent d'établir notre hypothèse et de construire un modèle scientifique de formulation des problèmes de recherche.

Notre postulat de départ est que certaines associations spécifiques – que nous appelons *formules de discours* – jouent le rôle de signaux alertant le lecteur sur le fait que l'auteur scientifique est en train de formuler un problème de recherche.

Nous avons donc commencé par constituer un corpus de référence constitué de 25 articles de recherche en science de l'éducation et en psychologie de l'éducation extrait du corpus Scientext.

Le corpus de référence constitue le contexte global de l'analyse. Il a le statut de référentiel représentatif, car c'est par rapport à lui que va se construire l'interprétation des résultats (F. Rastier et B. Pincemin, 1999). Ce corpus est conçu pour fournir une information en profondeur sur une langue. Il doit être suffisamment grand pour représenter toutes les variétés pertinentes de cette langue et son vocabulaire caractéristique, de manière à pouvoir servir de base à des grammaires, des dictionnaires, des thésaurus, et autres matériels de référence fiables [(J.M. Sinclair (1996)³⁹]. Notons que ce corpus de référence est très spécialisé,

Nous avons, par la suite, appliqué notre système d'identification de FD sur 51 textes extraits de la base de données SCIENTEXT.

3.1.1. Le Concept-matching

Le Concept-matching (Á. Sándor, 2007) est une méthode d'analyse qui consiste à implémenter des règles de grammaire de dépendances dans un formalisme de TAL. Pour les besoins de notre étude, nous avons implémenté cette méthode dans le Xerox Incremental Parser (XIP) dans le but de mettre en évidence les stratégies rhétoriques véhiculées dans les écrits scientifiques.

³⁹ J. M. Sinclair (1996) considère que “*A reference corpus is one that is designed to provide comprehensive information about a language. It aims to be large enough to represent all the relevant varieties of the language, and the characteristic vocabulary, so that it can be used as a basis for reliable grammars, dictionaries, thesauri and other language reference materials*”.

Le Concept-matching est défini dans les travaux de Á. Sándor (2007) comme une approche qui permet de modéliser les fonctions rhétoriques à partir de combinaisons des concepts constituants syntaxiquement reliés.

La notion de concept constituant représente un sous-ensemble des concepts généraux associés aux associations lexicales qui composent le métadiscours. Il est important de souligner que chaque réalisation de concept constituant est associée à une fonction donnée. Afin d'illustrer cette notion de Concept-Matching, nous reprendrons dans ce qui suit, pour un exemple développé par Á. Sándor (2007).

Le concept constituant de *contraste* qui est une fonction discursive qui peut être un adverbe (*paradoxalement*) ou un nom (*contradiction*).

D'autre part, la fonction rhétorique que ces concepts constituants accomplissent au sein des phrases est plus évidente à l'aide d'autres dépendances syntaxiques comme nous pouvons le voir à travers la figure 28 qui synthétise l'application du concept-matching à la rhétorique du métadiscours dans les résumés de biologie.

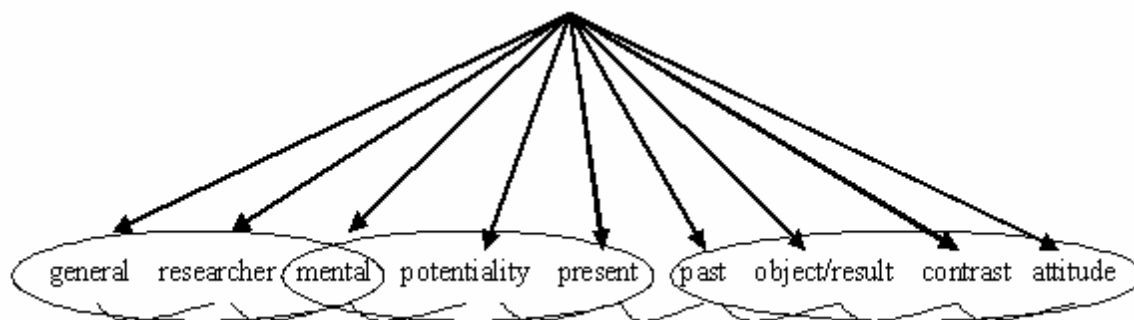


Figure 28 : Modèle rhétorique du métadiscours dans les résumés de biologie (Á. Sándor 2007)

Á. Sándor (2007) a identifié neuf concepts constituants qui sont totalement impliqués dans la rhétorique propre au métadiscours. Le même concept constituant peut-être impliqué dans plusieurs fonctions entre autres « Connaissances de base », qui concerne la contribution de l'auteur. Les courbes font référence à des liens entre les sous-ensembles de concepts-constituants utilisés ici symboliquement pour indiquer l'état de la présence de liens syntaxiques entre deux concepts constitutifs d'une part, et le fait que les expressions sont réalisées par des sous-ensembles de concepts constitutifs syntaxiquement liés, d'autre part.

F. Lisacek et al. 2005 (cité in Á. Sándor, 2007, 6) ont développé à l'aide de la méthode du concept-matching un moyen de détecter les marques du métadiscours qui véhicule la stratégie rhétorique de l'auteur dans la littérature scientifique. Leur méthode allie une approche conceptuelle à l'analyse syntaxique.

Ce modèle; dans lequel le présent travail s'inscrit; propose donc un formalisme qui décrit différentes FD – comme celles du *résumé* et *changement cognitif*– pour les besoins de cette étude, comme des cooccurrences dans des dépendances syntaxiques des concepts qui les constituent. Nous illustrons la démarche mise en œuvre dans la suite (voir figure 29).

Le concept-matching est donc basé sur une approche qui consiste à associer l'analyse syntaxique et l'analyse rhétorique à travers la conceptualisation des lexies et associations lexicales. Cette méthode consiste à établir des règles qui associent les sous-ensembles de dépendances syntaxiques de concepts aux FD comme nous pouvons le voir à travers la figure 29.

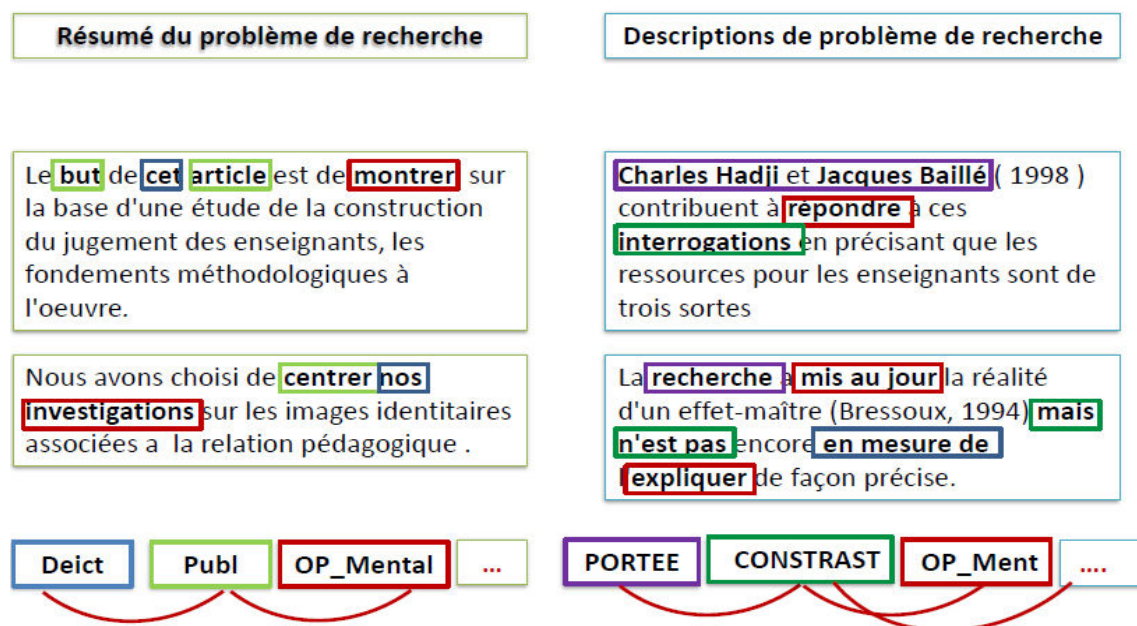


Figure 29 : Analyse syntaxique à travers la conceptualisation des lexies et associations lexicales.

Cela permet d'obtenir des sous-ensembles de séquences lexicales auxquels des fonctions rhétoriques sont associées, entre autres pour les besoins de cette étude ; PUBLICATION, OP_MENTALE, PORTEE, CONTRAST.

Cette méthode a été utilisée avec succès dans des applications de fouille de texte (Lisacek et al. 2005, Á. Sándor et al. 2006, Á. Sándor, 2009, Á. Sándor et Vorndran, 2010), ce qui explique notre choix.

Dans l'article de S. Sándor et Vorndran (2010), les auteurs proposent d'extraire les messages saillants contenus dans les publications de sciences sociales à l'aide de ce système de règles de cooccurrence « Concept-matching » qui les détecte et ce dans le but d'améliorer la pertinence de la recherche de document. Le problème abordé relève du fait que ces messages pertinents sont souvent des arguments et non des faits ce qui rend les technologies traditionnelles d'extraction d'informations inaptes à les saisir.

3.2. Développement de l'analyseur

Bien que le modèle conceptuel des FD soit indépendant des langues, le développement des systèmes pour les langues individuelles demande des traitements spécifiques. Il faut procéder à la compilation des mots et expressions qui peuvent réaliser les concepts constituants, et à l'établissement des règles des cooccurrences de ces mots et expressions pour la reconnaissance des deux FD.

L'analyseur syntaxique Xerox Incremental Parser (XIP) (S. Aït-Mokhtar et al. 2002) permet l'intégration du modèle de FD pour leur reconnaissance automatique. XIP exécute une chaîne de traitements consécutifs, à partir de la segmentation et l'analyse morphologique jusqu'à l'extraction des dépendances syntaxiques entre les mots et expressions. Le formalisme permet d'enrichir le lexique par des traits, et construire des règles de reconnaissance qui les utilisent. Ainsi, les dépendances syntaxiques extraites se prêtent à la construction des règles conceptuelles, qui permettent la reconnaissance des FD.

Notre tâche a consisté à développer des règles de grammaire de dépendances qui mettraient en évidence les associations lexicales, que nous qualifions de FD dans les articles de recherche. Nous avons donc mis en place un fichier de règles basées sur les dépendances syntaxiques des segments au sein d'une phrase en prenant donc en considération le contexte dans lequel ils apparaissent.

Notre modèle recourt à des grammaires de dépendances car la condition pour que les fonctions de *résumé et de changement cognitif* soient réalisées est conditionnée par la relation syntaxique des concepts constituants (voir 29).

Notre approche encode des règles grammaticales à l'aide de Xerox Incremental Parser (Ait et al., 2002). Des combinaisons de concepts que nous convenons d'appeler *concepts constituants* sont alors construites à l'aide du système de règles de cooccurrence appelé "concept-matching" (Á. Sándor 2007). C'est-à-dire que des règles de grammaire sont développées à partir de sous-concepts précis. Dans la section suivante, nous proposons une description plus détaillée des concepts constituants, ainsi que du lexique des FD.

Afin de modéliser ces FD, nous avons adopté une démarche qui consiste à reconnaître des patrons sous-jacents communs aux expressions de *résumé* et de *changement cognitif* (voir figure 31).

Ces patrons sont modélisés grâce à des descripteurs de natures différentes :

- Un trait qui catégorise le statut textuel ;
- Un trait qui explicite le statut linguistique de l'unité lexicale considérée ;
- Un trait qui réfère à l'étendue de l'opération mentale ;
- Un trait qui regroupe les éléments lexicaux qui représentent une négation, une contradiction, ou encore un changement accentué.

Ces patrons présentent donc des combinaisons des traits pertinents concernant les mots et les expressions constituant les FD étudiées. Voici quelques exemples de mots et leurs traits conceptuels :

4. *but, article, partie, texte* contiennent le trait conceptuel PUBLICATION
5. *nous, nos, ce* contiennent le trait conceptuel DEICTIQUE
6. *montrer, présenterons, analyser, a mis au jour, expliquer, explorée* contiennent le trait conceptuel OPERATION_MENTALE
7. *recherche* contient le trait conceptuel PORTEE
8. *a mis au jour, mais, n'est pas, peu* contiennent le trait conceptuel CONTRASTE

Par la suite, nous appelons ces traits pertinents dans la constitution des formules des discours des *concepts constituants*. Soulignons que nous préférons la notion de « trait conceptuel » à celle de « trait sémantique », que nous avons utilisée au départ, car nous nous situons dans une démarche plus cognitive. La notion de « concept » est donc utilisée

en référence aux fonctions de *résumé* et de *changement cognitif* selon la représentation de la figure 30.

La figure 30 représente l'analyse de XIP de la phrase « *Nous avons choisi de centrer nos investigations sur les images identitaires* » où les dépendances syntaxiques entre les mots sont extraites à la base d'arbres de « chunks » qui sont les constituants syntagmatiques des phrases.

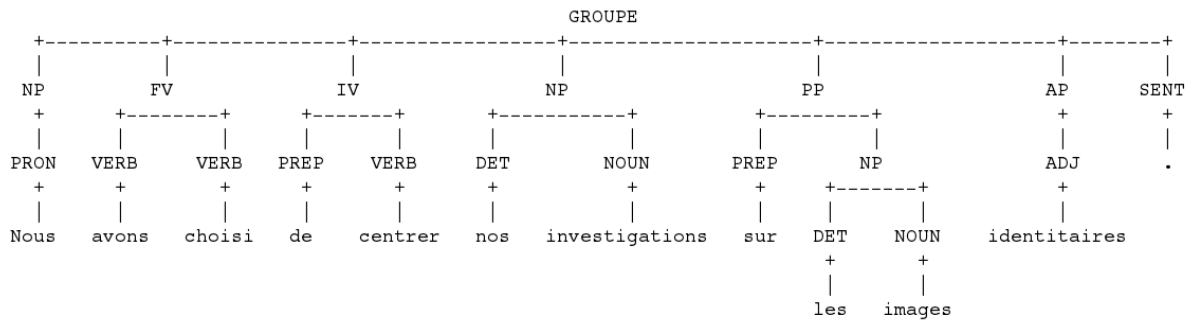


Figure 30 : Arbre de constituants

Nous obtenons la sortie de dépendances syntaxiques ci-dessous :

OBJ_KSDEP_SCOPE_MENTAL

(centrer_OP_MENTAL:+,investigations_PORTEE:+)

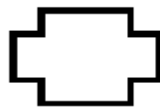
DETERM_KSDEP_RESUME_DEICT_SCOPE(investigations_PORTEE:+,nos_DEICT:+)

SUMMARY_KS_RESUME(Nous avons choisi de centrer_OP_MENTAL:+ nos_DEICT:+ investigations_PORTEE:+ sur les images identitaires .)

GROUPE{NP{Nous} FV{avons choisi} IV{de centrer_MENTAL:+} NP{nos_DEICT:+ investigations_SCOPE:+} PP{sur NP{les images}} AP{identitaires} .}

Nous représentons dans la figure 31 l'implémentation du module de Concept-matching au sein du parseur robuste.

- **Matching des patrons de concepts par des règles de co-occurrence**
- **Attribution des traits aux mots-clé, aux dépendances et aux phrases**



Analyse linguistique générale des textes :

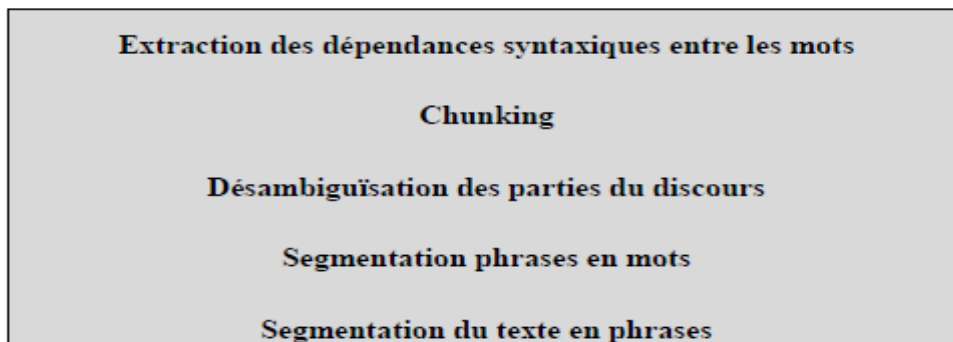


Figure 31 : Implémentation du Concept-matching au sein de l'analyseur automatique Xerox
Incremental Parser

Comme nous pouvons le voir à travers la figure 30, l'implémentation consiste, d'une part, à faire correspondre des patrons de concepts par des règles de cooccurrence et d'autre part, à l'attribution de traits aux mots-clés, dépendances et aux phrases exprimant le *résumé* et de *changement cognitif*. Cette intégration s'effectue au-dessus de l'analyse linguistique générale de XIP.

3.2.1. Démarche suivie

Afin de développer le système de détection des phrases où apparaissent les FD, nous avons procédé comme suit.

Tous les formats textes ne peuvent être analysés par XIP, ce qui nous contraint à extraire les contenus sous format txt ou xml et ainsi perdre des données importantes telles que les contenus des Figures, tableaux, etc.

Selon nos observations de corpus, ces données contenues au sein de ces derniers peuvent se révéler importantes et traduire un problème de recherche. Nous pouvons donc

supposer que cela minore la portée de l'expérimentation sur d'autres disciplines. Ce corpus a contribué au développement des règles de grammaire.

Puis, nous avons procédé à une annotation manuelle des phrases *résumé* et des phrases de *changement cognitif*. Cette annotation nous a permis de recenser un noyau de mots et expressions qui donnent lieu aux concepts constituants, ainsi que d'adapter des règles de cooccurrence à la grammaire de dépendances du français. Notre lexique comprend à l'heure actuelle 773 entrées et 19 règles de cooccurrences.

```
<ksref> Le <kswref> but </kswref> de <kswref> cet </kswref> <kswref>  
article </kswref> est de <kswref> montrer </kswref> , sur la base d'une étude  
de la construction du jugement des enseignants, les fondements  
méthodologiques à l'oeuvre. </ksref>
```

Figure 32 : Annotation sous XML du corpus de référence

Les balises <ksref>⁴⁰ et <kswref>⁴¹ sont utilisées pour informer XIP de ne procéder qu'à l'analyse des balises définies dans notre grammaire.

Puis, nous avons procédé à l'intégration du module du Concept-matching dans XIP. Nous reviendrons plus en détail dans ce qui suit sur cette étape et le fonctionnement de cette grammaire. Le module de Concept-matching prend en entrée l'analyse syntaxique qui est une fonctionnalité générale de XIP. La figure 32 illustre les dépendances syntaxiques comprenant les mots d'une formule de discours fournies par XIP :

⁴⁰ L'encodage fait référence à « KEY SENTENCE REFERENCE », c'est-à-dire à « phrase-clé de référence ».

⁴¹ L'encodage fait référence à « KEY SENTENCE WORD REFERENCE », c'est-à-dire à « mots contenus au sein des phrases-clés de référence ».

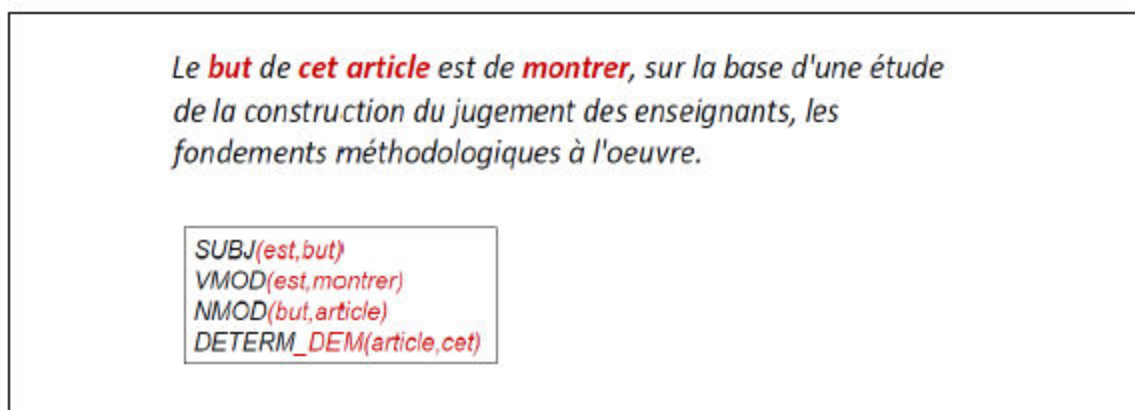


Figure 33 : Dépendances syntaxiques comprenant les mots d’une formule de discours fournies par XIP

L’intégration du module de Concept-matching dans XIP nous a permis d’avoir des sorties d’analyse avec des phrases dont les éléments lexicaux appartenant –au lexique recensé sont identifiés et surlignés comme dans l’exemple de la figure 32 extrait de notre corpus.

3.2.2. Actions pour développer cette grammaire

Afin de développer la grammaire, nous n’avons extrait que des textes qui ont pour domaine les sciences de l’éducation de la base de données Scientext. Ce corpus avait pour but de nous aider à développer nos règles de grammaire à travers l’observation des occurrences des lexies appartenant à notre lexique des écrits scientifiques et de tester par la suite la validité des règles sur d’autres corpus non consultés auparavant.

Afin de procéder à la détection des lexies qui sont représentatives du lexique scientifique nous nous assurons que ces dernières se trouvent en dépendance syntaxique avec d’autres lexies au sein d’une même phrase comme nous pouvons le voir à travers l’exemple qui suit :

(2)-CC ...ce cadre théorique, nous proposons, ...notre problématique....l’étude des questions....

⁴²DEPENDANCE_SYNTAXIQUE (cadre théorique, ce)

DEPENDANCE_SYNTAXIQUE (nous, proposer)

DEPENDANCE_SYNTAXIQUE (notre, problématique)

DEPENDANCE_SYNTAXIQUE (étude, questions)

⁴² Voir figure 34 pour mieux comprendre le fonctionnement de ces dépendances.

Une fois cette condition remplie, nous avons procédé à la constitution une liste de mots et d'associations lexicales qui composent nos concepts constituants. Puis, nous avons établi les règles de grammaire qui permettent d'attribuer des traits conceptuels aux lexies représentant le lexique scientifique. Nous obtenons ainsi une liste de mots et d'expressions, appartenant à ces concepts constituants, auxquels sont attribués les traits définis (voir figure 57). Voici la sortie d'analyse syntaxique enrichie par les traits conceptuels.

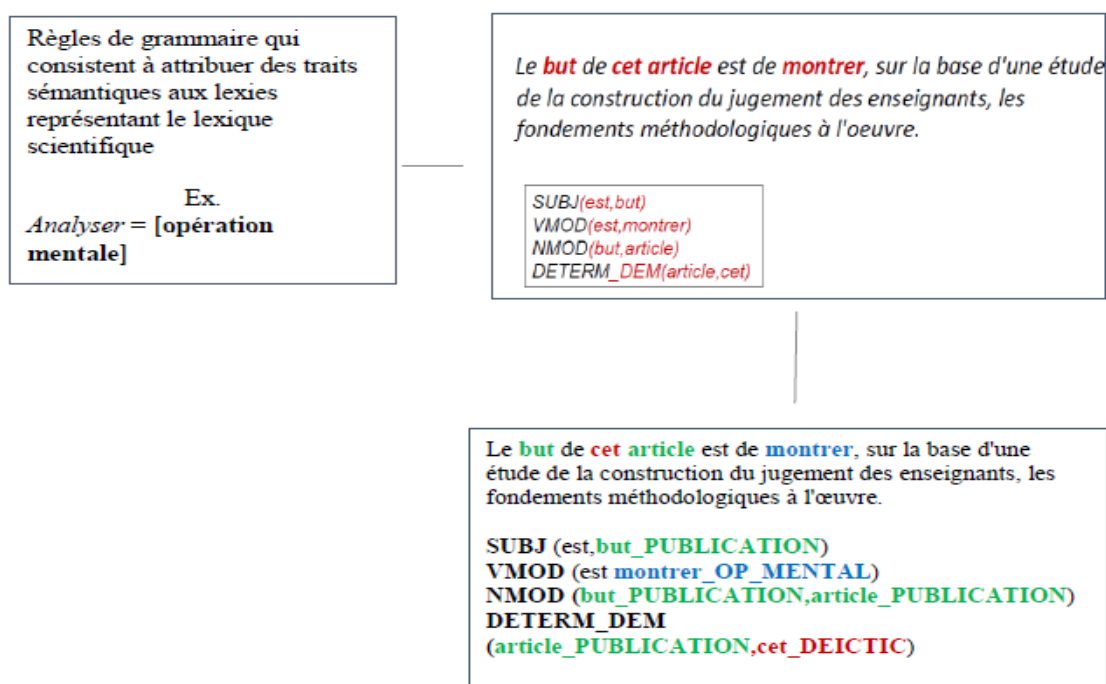


Figure 34 : Développement du lexique des concepts constituants

Enfin, on procède à l'analyse de notre corpus annoté sous XIP. Une fois cette analyse syntaxique réalisée, on visualise la sortie. Nous observons la structure syntaxique de chaque phrase et l'occurrence et la dépendance des lexies représentant nos concepts spécifiques dans leur co-texte (ces dépendances sont à la base du concept général de *résumé* et de *changement cognitif*).

A partir de là, on établit des règles de grammaire dans un fichier donné. De nouveau, ces règles sont appliquées sur notre corpus afin de tester leur validité. Cette étape de l'analyse nous permet d'ajuster nos règles pour minimiser le bruit.

3.2.3. Méthodologie d'intégration dans XIP et applications

Comme souligné précédemment, le développement de l'analyseur consiste à appliquer la méthodologie de « Concept-matching » (Á. Sándor 2007) : il s'agit d'une part, d'associer des traits aux listes des mots et des expressions préalablement compilées ce qui donne naissance aux concepts-constituants, d'autre part, de sélectionner parmi toutes les dépendances syntaxiques extraites celles qui sont des FD, ou en constituent des parties.

Les traits attribués au lexique des écrits scientifiques forment les concepts-constituants contribuant à construire notre concept général de *résumé* et de *changement cognitif*. Ci-dessous des exemples des traits attribués à ce lexique :

- *article, partie, texte, auteur, présenter, résumer* → PUBLICATION
- *nous, nos, ce* → DEICTIQUE
- *montrer, présenterons, analyser, a mis au jour, expliquer, étudier* → OPERATION_MENTALE
- *recherche, résultat, travail, analyse, projet, approche* → PORTEE
- *a mis au jour, mais, n'est pas, peu, question, anomalie, rejeter, énigmatique* → CONTRASTE

Ces traits sont attribués entre autres à plusieurs marques du métadiscours ainsi qu'au lexique transdisciplinaire, de la langue générale, abstrait non spécialisé. En effet, parmi les lexies et associations lexicales constituant notre lexique des écrits scientifique, nous retrouvons :

- Le lexique propre aux écrits scientifiques et qui renvoie aux procédures de l'activité scientifiques (résultat, observation, évaluation, etc.) → lexique transdisciplinaire.
- Le lexique de base, qui n'est propre à aucune discipline ni genre en particulier. Il est peu spécialisé et intègre des lexies qui ont une fonction grammaticale. (*Je, nous, ce, mais, etc.*) → lexique de la langue commune/générale.
- Le lexique qui relève d'un méta-niveau et n'est pas associé à un genre ou discipline en particulier (*question, contester, etc.*) → lexique abstrait non spécialisé.
- Les marques métatextuelles (*nous nous intéressons ici*) . → marque du métadiscours
- Les marques de prise de position (*rejeter/contester l'hypothèse, etc.*) → marques du métadiscours.

Les exemples suivants montrent la sortie de l'analyse syntaxique exécutée par XIP et les règles de cooccurrences qui permettent la détection de (5)-R et de (1)-CC :

(5)-R *Le but de cet article est de montrer*

DEPENDANCE_SYNTAXIQUE(but[PUBLICATION], article[PUBLICATION])

DEPENDANCE_SYNTAXIQUE (but [PUBLICATION], montrer [OP_MENT])

DEPENDANCE_SYNTAXIQUE (article [PUBLICATION], ce [DEICTIQUE])

RESUME = (PUBLICATION, PUBLICATION) +

(PUBLICATION, OP_MENT) +

(PUBLICATION, DEICTIQUE)

(1)-CC *La recherche a mis au jour ... mais n'est pas ... l'expliquer*

DEPENDANCE_SYNTAXIQUE (mis au jour[OP_MENT],recherche[PORTEE])

DEPENDANCE_SYNTAXIQUE (n'est

pas[COTRASTE],expliquer[OP_MENT])

CHANGEMENT COGNITIF = (OP_MENT, PORTEE) +

(COTRASTE,OP_MENT)

Comme nous pouvons donc l'observer dans l'exemple R, identifiés après l'intégration du module de Concept-matching, les occurrences de lexies associées aux traits DEICTIQUE PUBLICATION, OPÉRATON_MENTALE / PORTÉE sous réserve d'être en dépendance syntaxique, donnent lieu aux phrases-clé qui représentent le concept de *résumé*.

Dans l'exemple R, l'occurrence des lexies représentées par les traits CONTRASTE + OPÉRATON_MENTALE/ PORTÉE / PUBLICATION avec les mêmes conditions que précédemment donne lieu à notre concept général *changement cognitif* et permet l'identification des phrases qui le représentent.

3.3. Approche lexicale adoptée pour l'analyse des associations donnant lieu aux concepts constituants


Comme nous l'avons mentionné dans le chapitre 3, le développement des règles de grammaire Concept-matching nécessite une compilation du lexique des FD. Nous inscrivant dans la continuité des travaux de Á. Sándor (2007, 2008, 2010), nous avons enrichi le lexique des écrits scientifiques en fonction des éléments linguistiques de notre corpus et l'avons répertorié, d'une part, selon la catégorisation linguistique traditionnelle de chaque lexie, d'autre part, selon la catégorisation des associations lexicales donnant lieu aux FD. Nous avons attribué à chacun de ces éléments des traits spécifiques définis sur une base conceptuelle comme illustrés ci-dessous :

- *question, absurdité, anomalie, débat, doute, trouble*= CONTRAST/OBSTACLE
- *permettre, pouvoir, supposer*= POSSIBILITE
- *article, auteur, documentation, Figure*= PUBLICATION
- ...

Nous optons pour une approche lexicale qui consiste à croiser la dimension syntaxique et conceptuelle. L'idée de départ est que le sens de l'association lexicale se construit à travers les relations syntagmatiques des lexies, elles-mêmes sélectionnées par des choix conceptuels. Ces lexies se trouvent être en relation syntaxique avec d'autres éléments linguistiques du lexique scientifique. Pour rendre compte du sens de l'association lexicale, une analyse syntaxico-conceptuelle s'impose donc au préalable. Dans l'exemple ci-dessous les flèches représentent les relations syntaxiques entre les éléments lexicaux et les encadrés les traits attribués aux mots-clé.

(8)-CC Depuis Bourdieu [BOURDIEU & PASSERON 63], les *sociologues* français ont largement *analysé cette problématique* et ont contribué à systématiser l'emploi des catégories socioprofessionnelles dans les enquêtes statistiques.

(8)-CC *sociologues* [PUBLICATION]....., *analysé* [OP_MENTALE] *cette* [DEICTIC] *problématique* [CONTRATS] ...



Sur le plan syntaxique, dans un premier temps, cette démarche consiste à faire une description des segments selon l'approche linguistique traditionnelle fondée sur les catégories syntaxiques (noms, verbes, adjectifs, adverbes), puis des associations lexicales qui donnent lieu aux FD.

Enfin, elle consiste à caractériser ses éléments lexicaux variés par des traits tels attribués à un niveau conceptuel : *contraste, portée, mental, publication, possibilité*, etc., traits qu'il convient souvent, dans le système de préciser et d'affiner. Cette démarche est nécessaire car elle représente une partie importante de l'intégration de notre système au sein de l'analyse linguistique générale de XIP.

En effet, afin de procéder à la désambiguïsation des mots polysémiques, il est nécessaire en premier lieu de procéder à un étiquetage grammatical. Nous procédons donc dans ce qui suit à cette catégorisation des éléments lexicaux constituant en partie notre lexique des écrits scientifiques, en recourant aux catégories traditionnelles (*noms, verbes, adjectifs, adverbes, etc.*).

3.4. Classification grammaticale des éléments lexicaux

Afin de décrire notre lexique scientifique, des listes de mots par partie du discours de nature différente (adjectif, nom, verbe, adverbes, prépositions, conjonction) communes aux disciplines étudiées ont été établies. Des listes des associations lexicales ont également été établies et implémentées dans notre système (collocations, unités phraséologiques, etc.). En effet, nous avons recensé toutes les associations lexicales auxquelles sont associées des traits spécifiques et les avons intégrées au-dessus de la fonctionnalité XIP. Bien que la classification traditionnelle des éléments lexicaux de notre corpus (noms, verbes, adverbes, etc.) soit discutable, toutefois, elle constitue un point de départ stable pour notre étude qui cherche à explorer la nature et le fonctionnement des éléments des FD qui se trouvent en relation syntaxique.

Notons que pour les besoins de cette étude la catégorie *préposition* ne sera pas abordée car les prépositions n'ont pas été recensées dans notre lexique si ce n'est dans le cas d'association lexicales ce qui n'est pas très représentatif.

Les prépositions ne sont pas des éléments centraux des FD. Elles n'ont qu'un statut secondaire dans la construction de nos FD.

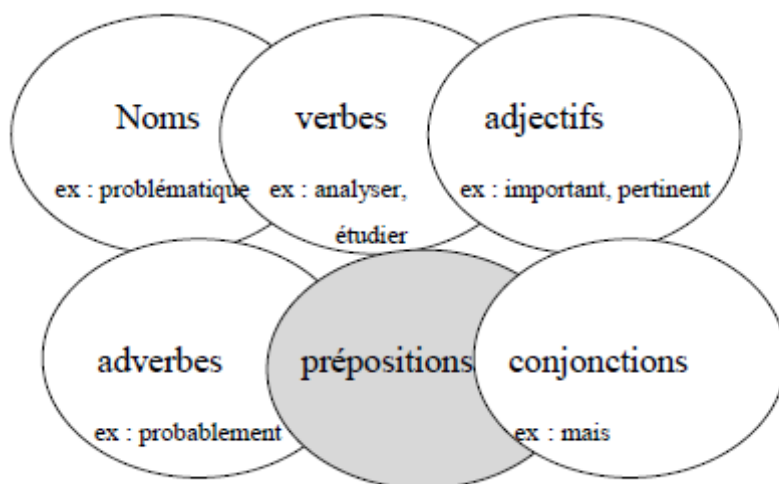


Figure 36 : Classification traditionnelle fondée sur les catégories syntaxiques d'éléments lexicaux de nature différente (M. Pecman, 2004)

En effet, afin de décrire au mieux notre lexique des écrits scientifiques, nous avons débuté par une classification linguistique traditionnelle, en nous inspirant des observations des modèles établis dans différents travaux de recherche, notamment ceux de A.Tutin et G. Bloquet (2004), K. Hyland (2005), A.Tutin (2007) et M. Pecman (2007) A.Tutin (2010).

- *La liste des adjectifs*

Pour les besoins de notre étude, nous ne recensons que les adjectifs dit *évaluatifs*. Cette classification s'inspire des travaux de Kerbrat-Orrechioni (1980) repris par A. Tutin (2008).

La catégorisation de C. Kerbrat-Orecchioni (1980) qui s'intéresse aux adjectifs subjectifs relève du cadre de l'énonciation. D'un point de vue général, la production langagière est basée sur deux paramètres qui sont l'interprétation et l'énonciation. Ils répondent à l'observation selon laquelle « les propriétés pragmatiques de l'énonciation se trouveraient être sémantiquement marquées » (Brockway, 1982 ; cité par R. Ghiglione, 1993). Voici donc la classification proposée par C. Kerbrat-Orecchioni (1980).

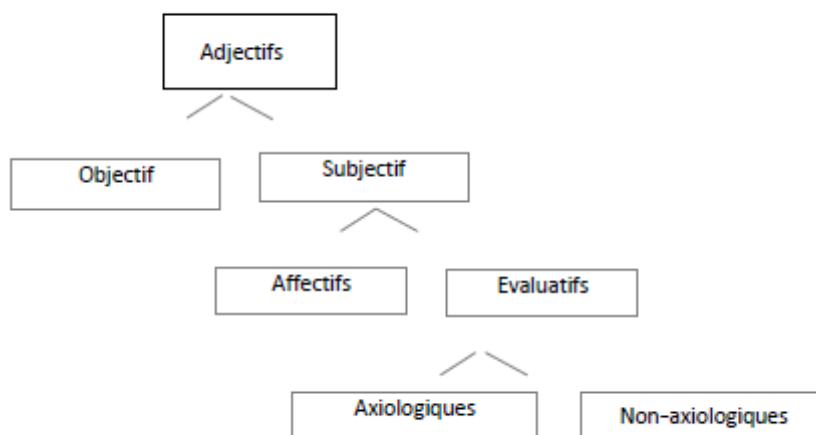


Figure 37: Classification des adjectifs subjectifs par C. Kerbrat-Orecchioni (1980)

Intéressons-nous de plus près à ces adjectifs évaluatifs. Ils sont considérés comme mettant en jeu la comparaison, le jugement (par rapport à une norme, même implicite), le jugement de valeur (axiologiques), et comportent aussi certains marqueurs de modalité.

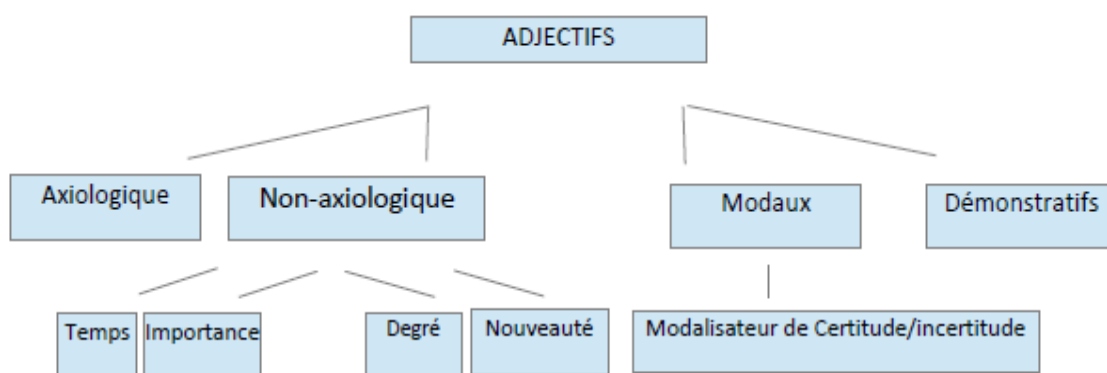


Figure 38 : Classification des adjectifs des écrits scientifiques basée sur les travaux de C. Kerbrat-Orecchioni (1980) & A.Tutin (2008)

Nous relevons des occurrences telles que *cette étude* pour la rubrique de démonstratifs ou encore des associations lexicales comme *il est possible de comprendre ...* pour celle de modalisateur de certitude/incertitude.

Concernant les démonstratifs, parmi les 129 occurrences dont le démonstratif *ce* est associé à des noms tels que *étude, travail, article, recherche*, etc. et se trouvent en

relation syntaxique avec d'autres éléments du lexique scientifique. Nous identifions 99 phrases-clé donnant lieu aux FD de *résumé* et de *changement cognitif* (voir annexe 1).

- *La liste des noms*

La classification linguistique des noms du lexique scientifique relevés de notre corpus s'inspire largement des travaux de N. Flaux et Van de Velde (2002) basés sur les propriétés syntaxiques, qui présentent des critères explicites. La classification qu'ils ont établie est large. Toutefois, elle concerne les noms mettant en jeu des agents humains.

A. Tutin et G. Bloquet (2004) qui s'inspirent également de ces travaux distinguent différentes classes linguistiques des noms du lexique scientifique (Schéma 1). Ils prennent l'exemple des noms répertoriés comme *procès*. Ces derniers sont des noms d'action qui se combinent avec le support *faire* et des variantes aspectuelles marquant une étape de l'action (*faire une étude, démarrer/entamer une étude, clore/achever une étude*).

A l'aide de ces précédents travaux, nous tentons d'établir un schéma qui reprend les différentes catégorisations de noms appartenant à notre lexique des écrits scientifiques.

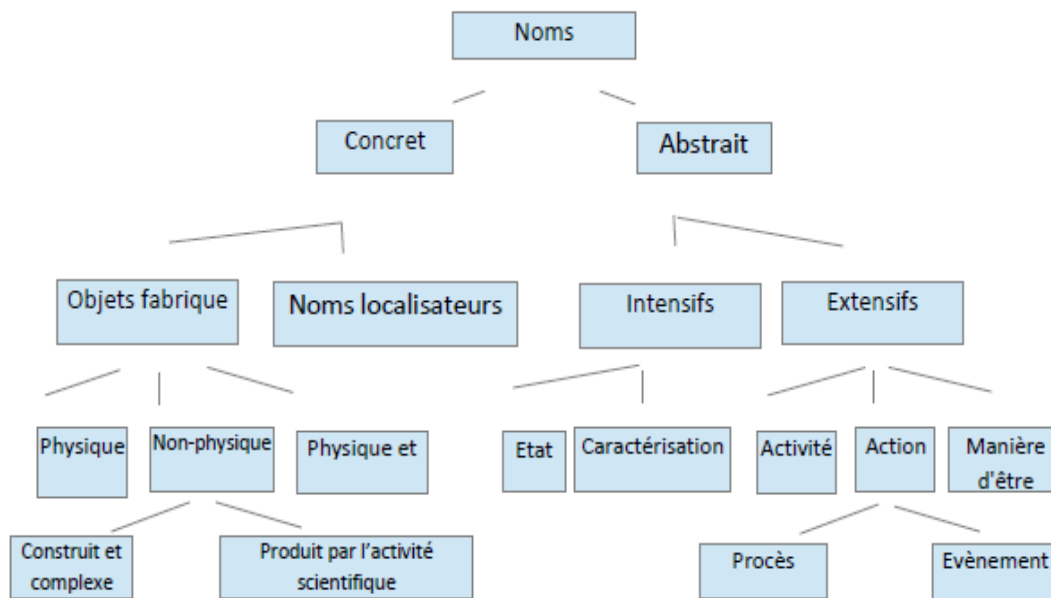


Figure 39 : Classification des noms transdisciplinaire inspirée des travaux de A. Tutin et G. Bloquet, 2004

Parmi les exemples de noms concrets qui appartiennent à la catégorie de *localisateurs*, nous avons des exemples d'occurrence comme *base*, *point*, etc. Nous relevons 151 occurrences, dont 12 qui se trouvent en relation syntaxique avec d'autres éléments de notre lexique scientifique comme illustré dans la figure 40 :

Contexte gauche	Occurrence	Contexte droit
PCDAI. Il nous paraît important d'approfondir ce	Point	dans une dernière partie en partant d'un éclairage
les limites que nous avons rencontrés sont autant de	Points	critiques révélateurs de modes de fonctionnement culturels et organisationnels
Nous avons noté aussi les limites du	Point	de vue de l'interfaçage dans un contexte institutionnel
nous ne ferons pour le moment qu'évoquer les	Points	qui suivent, réservant une réflexion plus approfondie lors
mettent souvent en place des actions proches des 6	Points	liés à un effet maître positif.
plus que ceux de " L " (1	Point	d'écart environ).
Sur ce	Point	, on peut remarquer que l'enseignante "
Ce	Point	est d'importance car un certain nombre de travaux ont
Les résultats présentés ci-dessous sont apportés pour servir de	Base	à l'élaboration d'instruments de collecte de données
L'ensemble de ces résultats vise à former la	Base	d'une recherche qui montrerait en quoi le TEC
Ce constat est la	Base	du présent texte qui présente les résultats d'une
Ce cadre constitue une	Bases	à partir desquelles l'analyse des

Figure 40: Occurrence des noms localisateurs identifiés dans les phrases-clé

Au sein de la même catégorie concernant les *objets fabriqués physiques et non physiques* nous retrouvons des illustrations comme *texte* ou encore *chapitre* (voir annexe 1). En effet, sur 129 occurrences, 99 se trouvent en relation syntaxique avec d'autres éléments de notre lexique scientifique et donnent lieu aux concepts de *résumé* et de *changement cognitif*.

Concernant les *objets fabriqués non-physiques*, nous relevons deux catégories ; la catégorie de noms construits et complexes, avec des exemples tels que *idées*, *système*, etc. et celle de produit par l'activité scientifique, comme dans le cas de *résultat*, etc. Sur les 285 occurrences du corpus, 17 se trouvent en relation syntaxique avec les lexies des écrits scientifiques recensées et sont identifiées dans les phrases-clé représentant nos concepts généraux. Celles-ci sont illustrées dans la figure qui suit.

Contexte gauche	Occurrence	Contexte droit
Comme on le voit, l'interprétation statistique des	résultats	soulève elle aussi de nombreuses interrogations.
Nous proposons les	résultats	d'une recherche relative à la question de la
A partir des	résultats	de l'enquête, nous nous interrogerons ici sur
On retrouve ici l'	idée	que l'éducation ne se limite pas à fournir
Les	résultats	de la recherche ont notamment mis en évidence :
Les	résultats	de cette première recherche nous apportent des indications importantes
Ces observations nous ont	idée	sur les compétences des élèves à

permis d'avoir une		utiliser un tableur
Nous empruntons ces	idées	, qui ont été discutées par de nombreux auteurs
Ainsi, nous pouvons montrer que les deux	systèmes	ne sont pas aussi éloignés l'un de l'autre

Figure 41 : Occurrence de noms construits et complexes, et ceux produit par l'activité scientifique

Au sein de la catégorie des *noms abstraits*, nous retrouvons celle des *noms intensifs* avec des occurrences comme *connaissance* dans la catégorie de « nom d'état » et *caractéristique* pour celle de « nom de caractérisation ».

Parmi les occurrences relevées dans cette dernière catégorie, nous trouvons des noms qualifiés de noms d'activité tels que *recherche*, d'autres exprimant un processus comme *démarche* et appartenant à la classe des noms d'action comme dans le cas des procès (ex. *étude, projet, etc.*) et celui des événements (ex. *choix*). Afin de traiter la polysémie, une lexie peut se voir attribuer plusieurs traits conceptuels comme c'est notamment le cas pour *analyse* qui en fonction de l'occurrence au sein d'une phrase peut être associée à *_OP MENTALE* ou *_PORTEE*.

Bien que cette classification soit large et privilégie des noms mettant en jeu des agents humains, elle permet de décrire, en grande partie, les noms composant notre lexique scientifique. C'est ce qui explique que nous nous en sommes inspirée afin d'attribuer des traits spécifiques aux noms appartenant à notre lexique des écrits scientifique.

- *La liste des verbes*

Les verbes retenus dans cette étude sont ceux qui s'inscrivent dans le fonctionnement métadiscursif (*nous présentons dans cette étude...*), ceux qui renvoient à une narration scientifique (*nous analysons les données...*), ou encore les verbes du positionnement qui renvoient notamment au questionnement de l'auteur (*nous nous interrogeons...*, etc.).

A travers nos observations empiriques et en nous inspirant de travaux antérieurs (K. Hyland, 2005, A.Tutin, 2010), nous avons établi différentes classes des verbes appartenant à notre lexique des écrits scientifiques.

Notons toutefois, que nous ne développerons dans le schéma qui suit que les catégories utiles pour le traitement des verbes des écrits scientifiques.

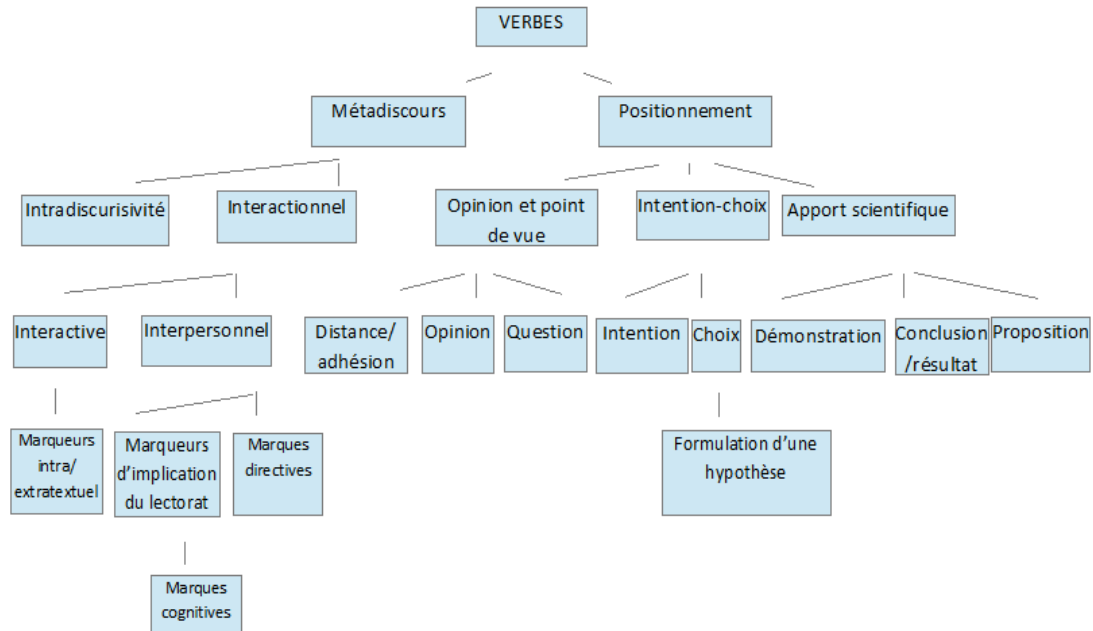


Figure 42 : Classification des verbes du lexique scientifique inspirée des travaux de K. Hyland (2005) & A.Tutin (2010)

Les verbes du métadiscours interactionnel extraits de notre corpus se divisent en deux catégories : interactive et interpersonnelle ; nous y retrouvons les marqueurs intra/intertextuels notamment comme dans l'exemple *nous l'avons noté plus haut* ainsi que des marqueurs d'implication du lectorat (ex. *notons que l'auteur*, etc.) et des marques directives comme dans les exemples suivant : *rappelons que, il est nécessaire de souligner que, etc.* Nous avons plus exactement 164 occurrences de ces verbes appartenant au métadiscours interactionnel. En revanche, il y a seulement 25 occurrences qui se trouvent en relation syntaxique avec d'autres éléments de notre lexique scientifique comme nous pouvons le voir à travers la figure suivant (Voir toutes les occurrences en annexe 1.2.).

Contexte gauche	Occurrence	Contexte droit
Nous nous contenterons ici de	souligner	qu'elle va exactement à l'encontre de l'idée
Ainsi avons –nous	noté	ci-dessus l'existence d'une dis-chronie , l'initiative
Nous avons	noté	aussi les limites du point de vue de l'
	Rappelons	que nous cherchons à montrer que les professeurs de
Il est intéressant de	noter	que la question de la problématique se pose pour
Nous avons	rappelé	ci-dessus que notre recherche s'inscrit dans un contexte

Figure 43 : Exemples de verbes du métadiscours interactionnel en relation syntaxique avec d'autres éléments du lexique scientifique.

Dans la rubrique de verbes qualifiés de *positionnement*, nous retrouvons des verbes de point de vue qui se divisent en trois catégories : les verbes dit de question (ex. *s'interroger*), d'adhésion/distante et enfin d'opinion comme dans le cas de *considérer que*. Dans la figure qui suit, nous retrouvons les occurrences des verbes de positionnement dits aussi de point de vue identifiés dans les phrases-clé (Voir toutes les occurrences en annexe 1.3).

Contexte gauche	Occurrence	Contexte droit
compte pour bien comprendre que l'analyse qui suit	Suppose	des conditions d'intervention gratifiantes pour ces professionnels,
Dans cette perspective, nous	considérons	d'abord les aspects situer, planifier et réaliser.

Incidemment J. Vignet	Considère	que le stress pose aussi " la question (
Il convient de s'	Interroger	: toute grille évaluative ressortit-elle du principe de
Une telle hypothèse m'amène à	Considérer	le terrain comme un système de signes

Figure 44: Exemples de verbes de point de vue qui se trouvent en relation syntaxique avec d'autres éléments du lexique scientifique

Sur les 332 occurrences de notre corpus, seules 28 se trouvent être en relation syntaxique avec d'autres éléments de notre lexique scientifique et sont donc identifiées dans des phrases représentant notre concept général de *changement cognitif* et/ou *de résumé*.

Concernant les occurrences des verbes d'apport scientifique (ex. *rechercher*), représentés par les verbes de démonstration, proposition et conclusion/résultat, nous retrouvons des exemples dans le corpus tels que *démontrer*, *dégager*, *conclure*, *proposer*.

Notons que sur les 159 occurrences relevées, seuls 25 verbes d'apports scientifiques sont identifiés dans les phrases représentant un *changement cognitif* et/ou un *résumé*. La figure suivante fait apparaître ces verbes au sein de leur contexte (Voir toutes les occurrences en annexe 1.4).

Contexte gauche	Occurrence	Contexte droit
Cette contribution introductive au symposium	proposé	sur le thème " Dispositifs de formation numériques et
Nous	proposons	les résultats d'une recherche relative à la question
Cette contribution se	propose	de rendre compte d'une recherche doctorale en cours
présentés dans cette contribution ne permettent certes pas de	conclure	à l'existence d'effets d'une raison numérique

Dans le cadre de cet article, nous	proposons	de cadrer l'analyse autour des décisions de justice
La problématique	proposée	ci-dessus est contextualisée à la formation professionnelle universitaire dont
Dans un premier temps, il s'agira de	dégager	la place qu'occupent les langues dans les programmes
2000) ou canadiennes de Mayhew (2000)	démontrent	pourtant que le sur risque des conducteurs novices décroît
Nous	proposons	que ces modes peuvent être mis en évidence par
Cet article	propose	que le modèle de compétences soit promu comme objectif

Figure 45: Occurrence des verbes d'apports scientifique en relation syntaxique avec d'autres éléments lexicaux du lexique scientifique

Nous remarquons une correspondance sémantique importante entre la détection du problème de recherche par XIP et la valeur des verbes relevés. En effet, la grande majorité des verbes employés dans les FD étudiées sont des verbes d'apport scientifique (*illustrer, dissuader, tenter, proposer, mettre en évidence, montrer, contribuer, présenter, confirmer, répondre, démontrer*).

Les verbes d'opinion sont beaucoup moins employés et les verbes d'intention ont un rapport particulier avec les hypothèses. La formulation d'hypothèses correspondrait bien avec l'introduction du problème de recherche dans le texte.

Nous illustrons dans ce qui suit les occurrences d'une partie des verbes d'opinion de notre corpus sans que cette liste soit exhaustive.

- Verbe d'apport scientifique *illustrer*

Nous retrouvons 16 occurrences dont 6 se trouvent être en relation syntaxique avec d'autres éléments de notre lexique scientifique et qui donne lieu à nos concepts généraux de *changement cognitif* et *résumé* comme observé dans la figure suivante.

Contexte gauche	Occurrence	Contexte droit
Dans la figure suivante nous avons	illustré	plus de détails concernant les enseignants interviewés.
Afin d'	illustrer	la problématique de la participation progressive de l'apprenant
(Rabardel, 1995) scientifiques sont donc ici	illustrés	dans leur fonction d'instruments pour comprendre, raisonner
menée a permis de mettre en évidence et d'	illustrer	ces phases ainsi que les processus qui les sous-
D'après les résultats	illustrés	ici, nous pouvons constater que la progression entre
Ces analyses	illustrent	les écarts de performance, dès la maternelle,

Figure 46 : Occurrence du verbe **illustrer** en relation syntaxique avec d'autres éléments lexicaux du lexique scientifique

- Verbe d'apport scientifique **tenter** :

Nous avons 42 occurrences du verbe **tenter** dans notre corpus, dont 14 sont en association lexicale avec d'autres lexies recensées dans notre système comme nous pouvons le voir à travers la figure 47 (Voir toutes les occurrences en annexe 1.5).

Contexte gauche	Occurrence	Contexte droit
Cet article	tente	de donner quelques éléments de réponses concernant les usages
C'est pourquoi nous	tenterons	d'aborder la question de

		la place de l'
que les travaux de Françoise Waquet, que nous	tenterons	de retourner à la source de la culture scolaire
Ces travaux	tenteront	d'articuler à la fois des investigations théoriques relevant
Pour	tenter	de répondre à cette question, nous avons mené
En résumé, cette recherche	tente	de dégager les critères qui font passer un événement
Au cours de cette recherche, nous avons	tenté	de caractériser le savoir analyser chez une vingtaine de

Figure 47: Occurrence du verbe **tenter** en relation syntaxique avec d'autres éléments lexicaux du lexique scientifique

- Verbe d'apport scientifique *mettre en évidence*

Nous relevons 46 occurrences du verbe *mettre en évidence* dont 18 apparaissent dans des phrases-clé comme illustré dans la figure 48 (Voir toutes les occurrences en annexe 1.5).

Contexte gauche	Occurrence	Contexte droit
Les résultats de la recherche ont notamment	mis en évidence	
nous cherchons à	mettre en évidence	les différents rôles qu'ils endossent , ainsi que
Il est intéressant de	mettre en	les verbes qu'il utilise à ce moment .

évidence		
Nous nous proposons de	mettre en évidence	de quelle manière les tâches et les rôles sont
Dans cette étude de cas, nous avons	mis en évidence	différents modes d'organisation de l'activité commune entre
Nous proposons que ces modes peuvent-être	mis en évidence	par l'analyse des interactions dans lesquelles est engagé
Cette approche tend ainsi à	mettre en évidence	des constructions d'usages différentes révélées par les significations
) ... , approches à partir desquelles on peut	mettre en évidence	un certain nombre de points essentiels :

Figure 48: Occurrence du verbe **mettre en évidence** en relation syntaxique avec d'autres éléments lexicaux du lexique scientifique

Contexte gauche	Occurrence	Contexte droit
données et d'analyse d'une recherche visant à	répondre	à des questions comme les suivantes :
Ainsi, afin de	répondre	à la problématique posée, nous chercherons à identifier
Pour tenter de	répondre	à cette question, nous avons mené une recherche
but d'une analyse globale de contenu est de	répondre	aux questions suivantes : existe-t-il une ou des
en vue de sélectionner parmi les documents ceux qui	répondent	exactement à la question traitée, demande la mise
Nos hypothèses nous conduisent à	répondant	à plusieurs exigences :

- Verbe d'apport scientifique *répondre*

Parmi les 77 occurrences du verbe *répondre*, nous en avons 6 qui apparaissent en relation synta

xique Figure 49 : Occurrence du verbe **répondre** en relation syntaxique avec d'autres éléments lexicaux du lexique scientifique

d'autres éléments de notre lexique des écrits scientifiques et qui sont donc identifiées dans les phrases-clé. La figure suivante illustre ces occurrences.

Enfin parmi les verbes d'intention ou de choix, nous relevons les occurrences suivantes : *choisir, écarter, souhaiter, etc.* Les verbes dit de formulation d'hypothèse sont des verbes de choix (nous faisons l'hypothèse que, etc.).

Comme indiqué précédemment, les verbes retenus sont tous ceux qui s'inscrivent dans le positionnement scientifique ou dans le métadiscours de l'auteur.

- *La liste des adverbes*

Pour les besoins de notre étude, nous recensons les catégories d'adverbes répartis sémantiquement comme suit :

- ceux qui expriment la *manière* dont le but est de montrer l'acte discursif réalisé soit dans l'environnement immédiat ou dans la progression du discours (N. Toumi 2009), l'affirmation, le *doute* ou la *négation*,

- ceux qui renforcent notamment le *jugement* (*central, fondamental, etc.*), les adverbes *d'illocution* (*pour résumer*). À ces groupes il conviendra d'ajouter les *adverbes employés comme mots interrogatifs*, enfin, les *adverbes modaux*.

Notons que pour la catégorie d'adverbes *métatextuels*, nous ne retenons que ceux introduisant une conclusion (résumer) et qui, associés à d'autres éléments linguistiques, construisent le concept de *résumé*. En effet, nous ne retenons que la catégorie métatextuelle *rhétorique* qui traite des éléments qui permettent au lecteur potentiel d'apercevoir les actes rhétoriques de l'auteur dans son processus d'argumentation (T.Dahl, 2004, 1811-1812).

La figure 50 reprend la classification de ces adverbes.

Comme nous pouvons le constater à travers la figure suivante en associant les travaux de T.Dahl (2004) et ceux de N. Toumi (2009), les adverbes propres au lexique scientifique peuvent être modaux, interrogatif, metatextuels, d’illocution, de jugement ou de manière (Affirmation, doute, négation).

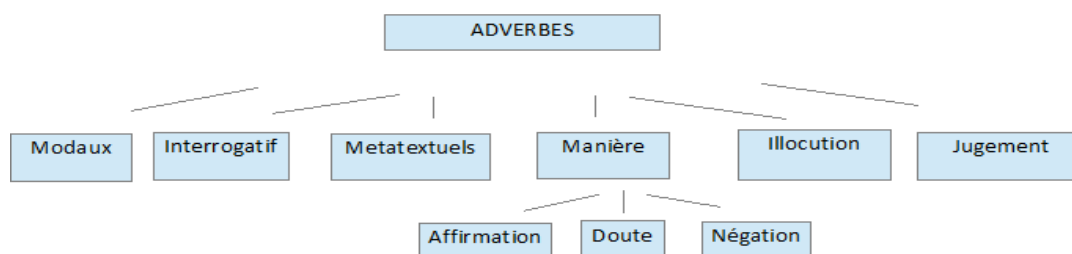


Figure 50 : Classification des adverbes du lexique scientifique inspirée des travaux de T.Dahl (2004) & N. Toumi (2009)

Parmi les 36 occurrences d’adverbes, dits de manière, qui se trouvent en relation syntaxique avec d’autres éléments linguistique du lexique scientifique, nous retrouvons les phrases illustrées dans la figure qui suit :

Contexte gauche	Occurrence	Contexte droit
Notons	Toutefois	que par son effet de filtre, le diplôme
Or, il serait	sans doute	plus opportun de rechercher la meilleure " proximité "
Des travaux sont	sans doute	encore nécessaires afin de rendre ces échelles de niveaux
Signalons	toutefois	que quelques garçons désirent travailler dans l’aéronautique ou
Cela en même temps contribuerait	certainement	, non sans heurts, à faire sortir l'
Si	toutefois	cette requête avait pour objectif de cerner la représentation

une société conquise par l'information, on peut	toutefois	émettre l'hypothèse que l'initiation aux outils et
de repérer des ensembles d'événements co- occurents ;	toutefois	, nous nous intéressons à l'ordre d'apparition de

Figure 51: Occurrences des adverbes de manière en relation syntaxique avec d'autres éléments du lexique scientifique

Les phrases illustrées dans la figure 51 dont les adverbes se trouvent en relation syntaxique avec d'autres éléments de notre lexique des écrits scientifiques sont identifiées comme phrases-clé.

Dans la rubrique des adverbes qui renforcent le jugement et ceux d'illocution, nous retrouvons plus de 64 occurrences dont seulement 8 se trouvent en relation syntaxique avec d'autres éléments du lexique scientifique donnant lieu à nos concepts généraux de *résumé* et de *changement cognitif* comme nous pouvons le voir à travers les exemples du Figure qui suit :

Contexte gauche	Occurrence	Contexte droit
	pour résumer	succinctement la problématique, le mémoire est soumis à
Notons enfin que le rôle moins	Central	de la tutrice en 2006 n'a pas freiné
c'est la tutrice qui a joué un rôle	Central	sur la plate-forme en 2005 ; mais cette dernière
C'est la thèse	Centrale	de L'éducation du genre humain de Lessing (
et que la question de l'équité est devenue	Centrale	pour penser les pratiques, libérée de tout

		instrumentalisme
La question de l'herméneutique est devenue	Centrale	en pédagogie dès lors que celle de la justice
Or, ceci est une dimension	fondamentale	de l'approche par les compétences qui est généralement
La question	fondamentale	n'est plus de savoir quels sont les déterminismes

Figure 52 : Occurrences des adverbes d'illocution et de jugement dans les phrases-clé

3.5. Structure syntaxique des associations lexicales (AL)

Comme nous avons pu le souligner dans le chapitre 2, les FD telles que nous les définissons dans ce travail, véhiculent des contenus plus larges que les collocations et/ou d'autres types d'unités phraséologiques. Toutefois, celles-ci apparaissent dans notre corpus, et se trouvent être en relation syntaxique avec d'autres éléments appartenant au lexique scientifique.

Nous allons donc procéder à la classification syntaxique de cet autre type d'association lexicale qui donne lieu aux FD étudiées.

La démarche consiste à identifier les différents moules syntaxiques qui permettent de former les AL considérées comme des constructions complexes qui se trouvent en relations syntaxiques les unes aux autres.

Au vu des phénomènes lexicaux récurrents, nous avons pu dégager les moules représentés dans la figure suivante :

STRUCTURE	EXEMPLES
v.+adv.	<i>Décrire personnellement, etc.</i>
v.+pré +n.	<i>Mettre l'accent, mettre en avant, mettre en perspective, mettre en jeu, analyse des résultats, etc.</i>
n.+adj.	<i>Cadre explicatif, cadre expérimental, modèle général, etc.</i>
n.+pré +n.	<i>Cadre d'analyse, cadre d'étude, champs de recherche, champ d'étude</i>
n.+adj.	<i>Point important, résultat important, concept fondamental, idées centrale, etc.</i>
adj.+n.	<i>Nombreux objets, nombreux textes, grand nombre, etc.</i>

Figure 53: Illustrations de la classification des AL selon le critère syntaxique

Nous pouvons donc observer que les AL apparaissent dans des constructions syntaxiques classiques et qu'elles occupent différentes fonctions au sein de la phrase. En effet, comme nous pouvons le voir à travers la figure qui suit, ces associations lexicales peuvent se distinguer par leurs différentes catégories grammaticales.

ASSOCIATION LEXICALE	VALEUR
<i>Mettre l'accent</i>	<i>Verbale</i>
<i>Mise au point</i>	<i>Nominale</i>
<i>Nombreux textes</i>	<i>Adjectivale</i>
<i>Bien que</i>	<i>Conjonctivale</i>
<i>Malgré</i>	<i>Prépositionnelle</i>

Figure 54: Classification des AL selon le critère catégoriel

3.6. Contenu conceptuel

Au départ, nous avons pensé adopter une approche sémantique afin de décrire et classer nos constructions phraséologiques. Cependant, nous nous apercevons au fur et à mesure de nos lectures, de nos analyses et de la diversité de mots de nature différente (adjectif, nom, verbe, adverbes, prépositions, conjonction) appartenant au lexique qu'il serait plus approprié de les classer selon leur contenu conceptuel. C'est pourquoi, nous tentons d'organiser dans ce qui suit les constructions phraséologiques selon ce contenu.

Nous avons alors recours à une approche notionnelle à travers une démarche onomasiologique. En effet, nous partons du concept pour aller vers son expression linguistique.

NOTIONS	EXEMPLES
OBSTACLE/CONTRASTE	<ul style="list-style-type: none"> - <i>question, absurdité, anomalie, débat, doute, trouble</i> - <i>contester, rejeter, clarifier, repenser</i> - <i>énigmatique, insuffisant, erroné, contradictoire</i> - <i>mais, cependant, à l'opposé</i>
OPERATION_MENTALEE	<ul style="list-style-type: none"> - <i>réflexion, raisonnement, affirmation, réponse, révision, argument</i> - <i>analyser, étudier, vérifier, chercher, cerner, découvrir, relever</i>
PORTEE	<ul style="list-style-type: none"> - <i>analyse, travail, projet, recherche, approche, modèle, système, théorie, idée, données, résultat, point, question, hypothèse</i>
PUBLICATION	<ul style="list-style-type: none"> - <i>article, auteur, documentation, Figure</i> - <i>présenter, exposer, résumer</i>
POSSIBILITE	<ul style="list-style-type: none"> - <i>possible</i> - <i>permettre, pouvoir, supposer</i> - <i>possibilité</i>
DEICTIQUE	<ul style="list-style-type: none"> - <i>je, nous</i> - <i>ce</i> - <i>suivant, présent</i>
IMPORTANCE	<ul style="list-style-type: none"> - <i>important, principal, intéressant, pertinent, grand, nombre</i> - <i>pertinence, importance</i>

Figure 55 : Classification des FD selon le contenu conceptuel

Inspirée des travaux de M. Pecman (2007), l'analyse de nos associations lexicales permet de décrire en partie quelques notions listées dans la figure 57 que nous convenons d'appeler *concepts constitutants* ; elle permet de dégager 3 sphères conceptuelles.

Rappelons que la quatrième sphère qui est celle de « l'Universalité » n'est pas prise en considération au sein de cette étude.

En effet, *celle-ci réunit les concepts qui font partie de l'univers conceptuel commun à une communauté linguistique plus large que celle représentée par la LSG [(concepts tels que |TEMPORALITY|, |QUANTITY|, |MANNER|, |RARITY|, etc.) (M. Pecman, 2007, 293)]* et qui ne sont pas importants et donc pas pris en considération pour les besoins de cette présente étude.

3.6.1. La sphère de la scientificité

Au sein de cette sphère, nous retrouvons tous les concepts constituants relevant du monde des connaissances scientifiques. Cette catégorie fait référence à une entité, objet, action, abstraction ou encore qualité (M. Pecman 2004, 332).

Parmi les concepts les plus représentatifs extraits de notre corpus se trouvent entre autres *contraste, opération mentale, portée*. Les éléments lexicaux représentés à travers ces différents concepts sont, toutefois, hétérogènes. En effet, au sein d'une seule et même catégorie, nous pouvons être amenés à retrouver des noms, verbes, adjectifs, etc.

3.6.1.1. Expression du contraste

L'expression du contraste regroupe tous les éléments lexicaux qui représentent le questionnement initial du chercheur. Les éléments lexicaux qui constituent cette catégorie sont divers et variés. En effet, nous retrouvons des conjonctions de coordination ou encore des adverbes (*mais, par conséquent, pourtant, néanmoins, cependant, à l'opposé, etc.*), des marqueurs d'hypothèses et de restriction qui peuvent être des noms ou des verbes (*problématique, interrogation, rejeter, réfuter*), adjectifs (*erronés, énigmatique, contradictoire, etc.*). La classification des marqueurs d'hypothèse et de restriction s'inspire des travaux de D. Siepmann (2007).

(4)- CC Depuis Bourdieu [BOURDIEU & PASSERON 63], les *sociologues* français ont largement *analysé cette problématique* et ont contribué à systématiser l'emploi des catégories socioprofessionnelles dans les enquêtes statistiques.

(4)-CC DEPENDANCE_SYNTAXIQUE(analyser[OP_MENT],sociologue[PORTEE])
 DEPENDANCE_SYNTAXIQUE(analyser[MENT],*problématique*[CONTRAST])
 DEPENDANCE_SYNTAXIQUE(*problématique*[CONTRAST], ce[DEICT])

Dans l'exemple ci-dessus, la phrase véhicule le concept général de *changement cognitif*. En effet, la lexie *problématique* est associée au trait CONTRASTE car celle-ci représente un problème, un contraste au niveau de la réflexion scientifique. En effet, *problématique* est un marqueur exprimant un problème considéré comme marqueurs de restriction et d'hypothèse par D. Siepmann (2007).

De plus, elle trouve se être en relation syntaxique avec d'autres éléments lexicaux associés à des traits spécifiques sur lesquels nous revenons plus en détail dans les parties suivantes.

3.6.1.2. Expression d'une opération mentale

Ce sont tous les éléments lexicaux qui renvoient à une activité mentale ou une opération cognitive quelle que soit la catégorie syntaxique (ex. *réflexion, raisonnement, analyser, vérifier*).

(7)-R Transparaît ici, d'une certaine manière, le difficile problème de l'articulation et de la complémentarité entre les considérations discursives externes à dominante théorique et l'exercice de la pratique inscrite plus ou moins fortement dans une théorisation ou réflexion en action

(7)-R DEPENDANCE_SYNTAXIQUE[(**inspirer**[OP_MENT],auteur[PUBLICATION])
 DEPENDANCE_SYNTAXIQUE[(**texte**[PUBLICATION],avancer[PORTEE])
 DEPENDANCE_SYNTAXIQUE[(**réflexion**[OP_MENT],menées[PORTEE])

Dans cet exemple, le verbe *inspirer* et le nom *réflexion* sont associés au trait *opération mentale* car ces derniers renvoient à un processus mental. De plus, ils se trouvent être en relation syntaxique avec d'autres éléments lexicaux représentant notre lexique des écrits

scientifiques auxquels sont attribués d'autres traits dont nous explicitons la nature dans ce qui suit.

3.6.1.3. Expression de la portée

Cette catégorisation qui s'inspire des travaux de A. Tutin (2007) est principalement représentée par trois catégories de noms : les noms de processus de l'activité scientifique, les noms d'objet construit par l'activité scientifique ainsi que les noms observables de l'activité scientifique (*analyse, travail, projet, recherche, approche, modèle, système, théorie, idée, données, résultat, point, question, hypothèse, démonstration, etc.*). Les verbes qui découlent ou qui sont associés à ces noms, peuvent se voir également attribuer ce trait conceptuel comme c'est notamment le cas de *démontrer, classer, observer, prouver, résulter, valider, vérifier, etc.*

(9)-R Pour documenter nos propos, outre une présentation qualitative des points centraux des modèles pédagogiques véhiculés par ces deux programmes nous proposons quelques statistiques descriptives à propos des programmes.'

(9)- R DEPENDANCE_SYNTAXIQUE(proposer[OP_MENT],nous[DEICTIC])

DEPENDANCE_SYNTAXIQUE(présentation[PUBLICATION],**point[PORTEE]**)

DEPENDANCE_SYNTAXIQUE(point[PORTEE], central[IMPORTANCE])

Comme nous pouvons le voir à travers l'exemple ci-dessus, la lexie *point* est représentée par le trait *portée*. En effet, cet élément lexical observable par l'activité scientifique est considéré comme étant un nom examiné par l'activité scientifique (A.Tutin, 2007). C'est pourquoi, nous lui avons attribué ce trait. La relation syntaxique de cet élément avec d'autres lexies auxquels des traits spécifiques sont attribués donne lieu à concept général de *résumé*.

3.6.2. La sphère de la modalité

En nous inspirant toujours du système de M. Pecman (2007), nous nous intéressons à la modalité qui permet à l'auteur d'exprimer son avis ou encore d'émettre un jugement à

l'égard de son travail. Deux catégories se dégagent au sein de cette sphère ; la modalité épistémique (*probabilité, possibilité*) et la modalité déontique (*nécessité*). Notons que pour les besoins de notre étude, nous ne nous intéressons qu'aux éléments linguistiques représentés par la modalité épistémique notamment à la *possibilité*. Rappelons que la modalité épistémique également appelée modalité de connaissance (P. Larreya, 2004), consiste à attribuer à une proposition une valeur de vérité (VRAI, FAUX, PROBABLE, POSSIBLE).

3.6.2.1. Expression de la possibilité

Cette catégorie regroupe les adjectifs, noms et verbes modaux comme nous pouvons le voir à travers la figure qui suit :

NOTIONS	Noms	Verbes	Adjectifs
POSSIBILITE	<i>possibilité</i>	<i>permettre, pouvoir, supposer</i>	<i>possible</i>

Figure 56: Expression de la possibilité

L'exemple ci-dessous illustre l'occurrence du verbe modal *permettre* auquel nous avons attribué le trait *possibilité* qui se trouve être en relation syntaxique avec d'autres éléments lexicaux de notre lexique scientifique et qui est donc identifiée comme une phrase clé.

(10)-P⁴³ La recherche en éducation comparative doit nous permettre de passer d'un stade simplement impressionniste à une description systématique, et au développement de théories explicatives des similitudes et différences entre les systèmes.

(10)-P

DEPENDANCE_SYNTAXIQUE(**permettre**[**POSSIBILITE**],recherche[PORTEE])

DEPENDANCE_SYNTAXIQUE(**permettre**[**POSSIBILITE**],nous[DEICTIC])

DEPENDANCE_SYNTAXIQUE(développement[OP_MENTALE],théorie[PORTEE])

⁴³ « P » fait référence au concept constituant « possibilité ».

3.6.3. La sphère de la discursivité

« La sphère de la discursivité englobe tout ce qui repose sur des impératifs purement argumentatifs, tels que la présentation du sujet d'étude (cf. |PRESENTATION|), le positionnement de l'auteur (cf. |POSITIONING|), l'introduction des exemples (cf. |EXAMPLE|), l'indication d'une source externe (cf. |QUOTATION|), l'établissement d'une transition (cf. |TRANSITION|), etc. » (M. Pecman, 2007).

En nous basant sur la description précédente, nous avons recensé dans notre corpus, sous l'étiquette de *sphère de discursivité* toutes les marques linguistiques qui signalent notamment les exemples, les citations, la présence de l'auteur. En effet, cette sphère englobe tous les éléments linguistiques permettant une progression logique du discours et qui illustrent les stratégies utilisées par l'auteur pour structurer le contenu de son discours en fonction des objectifs communicatifs qu'il poursuit (M. Pecman, 2007, 409).

Dans le cadre de cette étude, nous nous intéressons particulièrement aux déictiques, dont les pronoms personnels font référence à l'auteur, ou aux adjectifs démonstratifs. L'intérêt porté aux déictiques résulte de nos observations empiriques et par le fait que ces éléments lexicaux soient en dépendance syntaxique le plus souvent avec les segments répertoriés sous l'étiquette PUBLICATION.

(10)-R Le présent article retrace cette évolution méthodologique, tout en commentant les principaux résultats des recherches.

(10)-R DEPENDANCE_SYNTAXIQUE(article[PUBLICATION], présent [DEICTIC])

DEPENDANCE_SYNTAXIQUE (résultat[PORTEE],recherche[PORTEE])

DEPENDANCE_SYNTAXIQUE (principaux [IMPORTANCE],résultat [PORTEE])

Nous avons recensé plusieurs catégories de déictique dans notre lexique des écrits scientifique. En effet, parmi les marques discursives auxquelles le traits *Deictic* a été attribué, nous retrouvons l'exemple d'adjectif comme *présent* qui dans l'exemple ci-dessus en étant en relation syntaxique avec d'autres éléments lexicaux donne lieu au concept général de *résumé* et représente donc une phrase clé.

3.7. Hiérarchie des concepts constituants

Dans la figure 57, nous présentons les éléments constituants auxquels nous avons attribué des traits, définis à un niveau conceptuel du discours ainsi que leurs correspondances dans les différentes classifications passées en revue [A.Tutin (2007), M. Pecman (2007) et D. Siepmann (2007)].

CONCEPTS CONSTITUANTS	EXEMPLES	CORRESPONDANCES
OBSTACLE/CONTRASTE	<ul style="list-style-type: none"> - noms exprimant une problématique: <i>question, absurdité, anomalie, débat, doute, trouble</i> - verbes : <i>contester, rejeter, clarifier, repenser</i> - adjectifs : <i>énigmatique, insuffisant, erroné, contradictoire</i> - conjonctions de coordination, adverbes : <i>mais, cependant, à l'opposé</i> 	<p>noms de processus de l'activité scientifique,</p> <p>noms d'objet construit par l'activité scientifique,</p> <p>noms d'observables de l'activité scientifique (A. Tutin, 2007)</p>
OPERATION_MENTALEE	<ul style="list-style-type: none"> - noms : <i>réflexion, raisonnement, affirmation, réponse, révision, argument</i> - verbes : <i>analyser, étudier, vérifier, chercher, cerner, découvrir, relever</i> 	<p>sphère de scientificité (M. Pecman 2007)</p> <p>marqueurs de restriction et d'hypothèse (D. Siepmann, 2007)</p>
PORTEE	<ul style="list-style-type: none"> - noms : <i>analyse, travail, projet, recherche, approche, modèle, système, théorie, idée, données, résultat, point, question, hypothèse</i> 	

	<i>résultat, point, question, hypothèse</i>	
PUBLICATION	noms : <i>article, auteur, documentation, Figure</i> • verbes : <i>présenter, exposer, résumer</i>	noms de supports de la rédaction scientifique, noms d'acteurs de l'activité scientifique (Tutin 2007)
POSSIBILITE	- adjectifs modaux : <i>possible</i> - verbes : <i>permettre, pouvoir, supposer</i> - nom : <i>possibilité</i>	sphère de la modalité (Pecman 2007)
DEICTIQUE	- pronoms personnels de 1ère personne : <i>je, nous</i> - adjectifs démonstratifs : <i>ce</i> - adjectifs : <i>suivant, présent</i>	sphère de la discursivité (Pecman 2007)
IMPORTANCE	- adjectifs : <i>important, principal, intéressant, pertinent, grand, nombre</i> - noms : <i>pertinence, importance</i> - verbes : <i>intéresser, souligner</i>	noms de caractérisation (Tutin 2007)

Figure 57: Concepts constituants et correspondance avec les catégories d'autres classements

En dehors des différences des portées des catégories dans les trois systèmes, on observe que contrairement aux systèmes de M. Pecman et de A.Tutin, notre liste contient des mots fonctionnels et des adjectifs subjectifs (C. Kerbrat-Orecchioni 1980).

En effet, ces catégories de mots n'appartiennent pas au lexique transdisciplinaire, mais au langage général.

C'est justement l'inclusion de ces mots dans le cadre de notre analyse qui distingue le concept de formule de discours des collocations ou unités phraséologiques, car rappelons que le lexique spécifique constituant ces FD, n'est pas nécessairement formé de collocations, les éléments figurant dans les formules n'étant pas obligatoirement en relation de dépendance syntaxique.

En plus des suites de mots répertoriés dans notre lexique (voir annexe 2), rappelons que les FD sont également des suites d'expressions syntaxiquement cohérentes, mais qui ne se suivent pas nécessairement, aux structures syntaxiques et aux composantes lexicales variées comme nous pouvons le voir à travers la figure 58. L'analyse des articles en sciences de l'éducation a donc permis d'identifier plusieurs types de stratégies discursives à travers ces associations d'expressions lexicales qui expriment des *changements cognitifs* et le *résumé* de l'auteur.

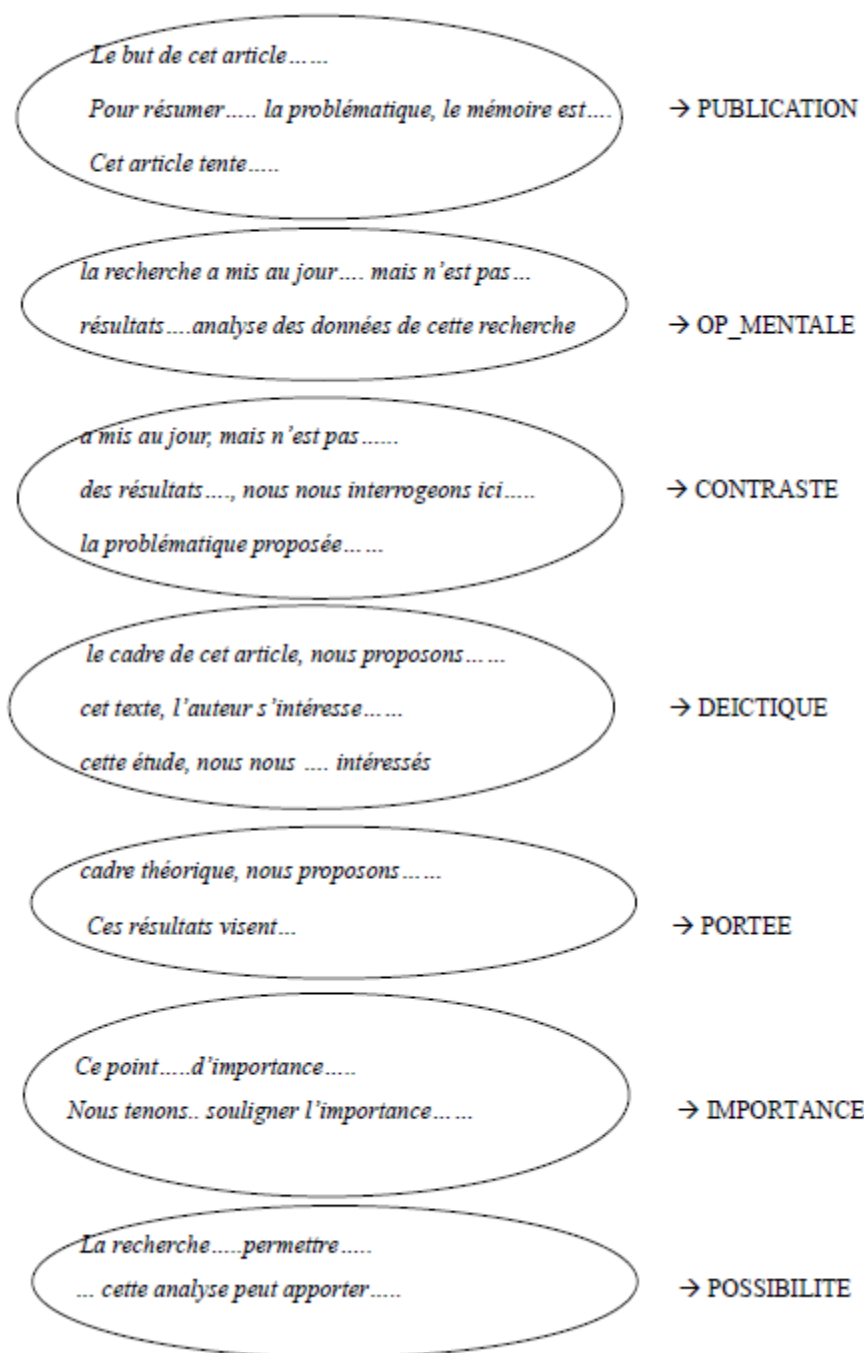


Figure 58 : Formules de discours exprimant diverses stratégies discursives qui véhiculent des changements cognitifs

La figure 58 illustre la typologie de signal du *changement cognitif* et de *résumé* exprimé par l'auteur des articles de recherches.

En plus de l'analyse lexicale puis conceptuelle des concepts constituants donnant lieu au FD traduisant un problème de recherche dans l'article de sciences de l'éducation, nous tentons également de fournir quelques éléments de caractérisation linguistique à ces associations lexicales s'appuyant sur une analyse morpho-syntaxiques.

En regard des caractéristiques supposées de ces associations lexicales particulières, nous décrivons la structure générale de la phrase, les temps verbaux, les marques de personne et de non-personne, tout en nous appuyons sur l'analyse antérieure du lexique spécifique à l'activité de recherche.

Pour les besoins de cette caractérisation linguistique, nous prenons en compte des aspects hétérogènes à savoir ;

- les analyses textuelles relevant de la structuration de la phrase,
- les temps verbaux relevant de son rapport au lecteur
- les marques de personne et de non-personne relevant de l'inscription de son auteur

Nous ne nous intéressons pas ici aux aspects des caractéristiques marquées en termes d'usages disciplinaires, mais plutôt à des usages liés au statut de l'auteur.

3.8. Structuration des phrases traduisant un problème de recherche

Nous observons une prédominance de la *phrase déclarative* (par opposition aux phrases interrogatives et impératives). Ces phrases, les plus répandues, sont le plus souvent sous forme active, et parfois passive. Nous aborderons les formes de discours indirect employées par l'auteur dans la partie consacrée aux marques de personne et de non-personne relevant de l'inscription de son auteur. Les autres phrases sont écartées car les FD n'ont été identifiées qu'au sein de ces structures.

3.8.1. Structure de phrase déclarative

Nous remarquerons que la forme la plus répandue à laquelle l'auteur recourt pour formuler son problème de recherche est la *phrase déclarative* comme nous pouvons le voir à travers les exemples extraits de notre corpus.

(11) R- *L'étude présentée dans cet article rend compte* du rôle des caractéristiques perçues des stimuli, que ce soit dans le cadre d'un jugement de similarité plus classique ou dans celui du raisonnement analogique.

(12) R- *Ce résultat nous permet de conclure* à une intégration des unités lexicales des deux langues dans la mémoire bilingue

(13) CC- Or, dans *cette première expérience, nous répliquons un résultat déjà observé* (Martinot et al., 2002) en montrant que les femmes en comparaison ascendante avec un membre de l'exogroupe (un homme) présentent une plus faible estime de soi que celles en comparaison ascendante avec un membre de leur groupe (une femme).

(14) CC- Saisis dans leur ensemble, *nos résultats viennent confirmer* d'autres résultats issus de *recherches récentes* sur la comparaison sociale.

(15) CC- D'autres *recherches*, au contraire, *ne confirment pas* cette dichotomie

(16) CC- *Mais*, contrairement à notre attente, la seconde *comparaison* planifiée *montre que cette différence* n'est pas statistiquement significative, $F(1,44) = 1,09, p = .30$.

(17) CC- *Ce résultat* particulier est imputable au fait que les durées des derniers essais d'habituation obtenues pour ce groupe, sont particulièrement brèves (4,2 s) et donc plus sensibles, sans *doute*, à une augmentation des durées qui, *nous le rappelons*, est générale pour les tests de reconnaissance quelle que soit la condition.

(18) R- Les *présentes expériences* **apportent** une pierre supplémentaire à l'édifice des *travaux portant* sur la ressemblance : contre les conceptions classiques objectivistes, ce travail s'inscrit dans la perspective constructiviste qu'on trouve par exemple chez Goldstone (1995).

Nous observons, au sein de ces phrases, une dimension argumentative importante avec l'usage de traits tels que la négation et les connecteurs.

De façon générale, l'emploi de forme active pour véhiculer un problème de recherche est plus répandu que l'usage passif. Cela rejoint les travaux précédents qui ont porté sur la question ; Givon (1979) avait obtenu pour résultat 4% et 18% de passifs selon le type de texte étudié, C.L. Barber (1962), quant à lui, a trouvé un usage plus important dans les

textes scientifiques avec plus de 30%, ce qui vient corroborer les constatations de S. Carter-Thomas dans des travaux antérieurs [cité in S. Carter-Thomas, 2009]. Toutefois, il n'est pas conseillé de recourir à l'emploi du passif dans les manuels de grammaire malgré son importance dans la structuration informationnelle. Car cette forme est souvent considérée comme lourde et complexe (Day 1995 cité in S. Carter-Thomas, 2009).

3.8.2. Structure de phrase interrogative

L'emploi de structures interrogatives est assez rare. En effet, nous avons constaté que les phrases véhiculant un problème de recherche sont parfois suivies de *phrases interrogatives* comme nous pouvons l'observer dans l'exemple suivant.

(19) CC- À ce titre, *on peut formuler* une *question centrale* de la manière suivante : Qu'est-ce qui permet à l'enseignant d'exercer une influence sur les élèves dans une classe ?

Nous avons là, un emploi occasionnel, de la *phrase interrogative* car rappelons que nous ne nous intéressons pas au contenu asserté des questions principales auxquelles le chercheur tente de répondre, mais aux indices qui annoncent que celui-ci est en train de formuler son problème de recherche. Ce qui pourrait justifier cet emploi rare au sein des phrases-clés.

3.8.3. Structure de phrase impérative

Nous avons également relevé un emploi occasionnel de *phrases impératives* qui permettent à l'auteur d'établir des liens avec son lecteur potentiel.

(20) CC- *On reviendra*, dans la conclusion, sur *cette difficile question* des effets affectifs du redoublement ; **rappelons** dès à présent qu'il s'agit d'une composante psychologique plus difficilement appréhendable par des tests que les acquis cognitifs.

L'usage de la forme impérative reste rare au sein de notre corpus. Pourtant, cet emploi peut permettre aux auteurs d'attirer l'attention des lecteurs sur la formulation à venir du problème de recherche. Des études portées sur cette question, soutiennent l'idée selon laquelle l'emploi des phrases impératives de façon générale dans l'écrit scientifique, est

beaucoup plus répandu en anglais (S. Carter-Thomas, 2009, 199) qu'en français. C'est peut-être ce qui justifie cette absence de forme impérative.

3.9. Temps verbaux

Dans ce qui suit, nous tentons de fournir un aperçu rapide et partiel des temps verbaux utilisés au sein des phrases-clé. Effectuer une analyse approfondie sur ce sujet requerrait une application de nos grammaires formelles sur un corpus plus important et une étude spécifique qui dépasse notre cadre de recherche.

3.9.1. Présent de l'indicatif

Conformément aux caractéristiques attendues du genre de l'article : c'est le présent de l'indicatif qui domine largement le corpus comme nous pouvons notamment le voir à travers les exemples illustrés dans la partie concernant la structure déclarative des phrases (3.9.1.) ainsi que les exemples traités dans le développement de l'analyseur au sein de ce même chapitre à savoir :

(2)- CC En *nous inspirant* de ce *cadre théorique*, nous **proposons** de préciser *notre problématique* en termes de gestion des distances aux différents savoirs pour l'étude des questions économiques et sociales

(3)- CC *Le problème* **est** que *cette interprétation* de la transposition didactique *mène à penser* que la transformation *opère* sur un objet pour produire un nouvel objet.

Cette observation vient confirmer celle constatée dans les travaux de F. Rinck (2006), à savoir une prédominance du présent de l'indicatif considérée comme caractéristique centrale relativement stable du genre de l'article. Ce n'est donc pas un indice particulièrement intéressant pour identifier les FD, puisque c'est un trait qui concerne le genre dans son ensemble.

Nous avons également observé que, l'emploi de ce présent prend souvent la valeur de vérité générale notamment en caractérisant des procès d'ordre scientifique au sein de notre corpus [verbes d'apport scientifiques (ex. *rechercher*), représentés par les verbes de démonstration, proposition et conclusion/résultat tels que *démontrer, dégager, conclure, proposer*].

3.9.2. Le passé composé

Dans les exemples extraits de notre corpus, le passé composé est employé en alternant avec le présent de vérité générale. Cet emploi caractérise des procès d'ordre scientifique (*analyser, rechercher, etc.*) ou langagiers (*parler*) comme nous pouvons le voir à travers les exemples qui suivent.

(4)- CC Depuis Bourdieu [BOURDIEU & PASSERON 63], les *sociologues* français **ont** largement **analysé** cette *problématique* et ont contribué à systématiser l'emploi des catégories socioprofessionnelles dans les enquêtes statistiques.

(21) CC& R- *Ici*, nous **n'avons pas parlé** d'expertise, mais de spécialité car le domaine d'expertise n'était pas l'objet de l'étude ; cependant, ce que montre la situation C ainsi que la comparaison des situations, c'est que la spécialité est une attitude contingente, toujours susceptible d'être renversée par une autre attitude

Cet usage sert, d'une part, à la mise en scène à l'auteur dans sa démarche scientifique et des sujets qui ont été soumis à l'expérimentation, d'autre part, il est employé pour rendre compte de la méthodologie de la recherche dans les articles scientifiques (F. Rinck, 2006, 203).

3.9.3. Le futur

Nous constatons dans l'exemple qui suit un emploi du futur associé à l'impératif.

Nous rejoignons F. Rinck (2006, 204) pour dire que

« le futur et l'impératif s'inscrivent en effet dans la majorité de leurs emplois dans des Formes de métadiscours, que l'on peut définir en référence aux travaux de J. Rey Debove (1978) comme une forme spécifique de métalangage où les autonymes désignent le texte lui-même et les opérations accomplies par l'auteur dans sa construction »

(22) CC- *On reviendra*, dans la conclusion, sur *cette difficile question* des effets affectifs du redoublement ; **rappelons** dès à présent qu'il s'agit d'une composante psychologique plus difficilement appréhendable par des tests que les acquis cognitifs.

L'emploi de l'impératif dans le cas de figure de l'exemple précédent est utilisé pour introduire une visée temporelle de futur. En effet, dans l'exemple (20)-CC, cet usage temporel s'est vu associer une valeur injonctive. Toutefois, nous ne pouvons généraliser cet emploi sur l'ensemble de la discipline car il n'est pas répandu, et cette observation est réduite à un seul emploi bien que des travaux antérieurs portant sur l'article de recherche confirment son caractère productif (voir en particulier F. Rinck, 2006).

3.9.4. L'imparfait

L'usage de l'imparfait dans notre corpus est corrélé au présent. Nous constatons que les auteurs en font usage dans les articles de sciences de l'éducation sans que celui-ci ne soit très répandu. On peut penser que cet emploi marque l'effacement de l'auteur face à son énoncé. Dans l'exemple suivant, nous observons un emploi à valeur objectivante :

(23) R- L'ambition de *cette étude était* de *mettre en évidence* l'effcience d'un contexte d'anticipations sémantiques de nature axiologique.

(24) CC& R- *Ici*, nous n'avons pas parlé d'expertise, mais de spécialité car le domaine d'expertise n'*était* pas *l'objet de l'étude* ; *cependant*, ce que *montre* la situation C ainsi que la *comparaison* des situations, c'est que la spécialité est une attitude contingente, toujours susceptible d'être renversée par une autre attitude.

Comme nous avons pu le voir tout au long de cette partie portée sur les temps verbaux, la valeur temporelle du présent de l'indicatif est la plus répandue et utilisée en terme d'opération rhétorique de l'auteur.

3.10. Les marques de personne et de non-personne relevant de l'inscription de son auteur

Dans une démarche énonciative, nous nous intéressons ici aux personnes linguistiques telles qu'elles se manifestent dans l'usage des pronoms personnels et des déterminants auxquels nous avons attribué le trait conceptuel de *Deictic*.

Deux groupes de variables ont été identifiés comme nous pouvons le voir à travers les exemples (4)-CC et (21) CC& R :

- Les pronoms personnels sujets avec pour pronom la 2^{ème} personne au pluriel: nous ;

- Les déterminants possessifs, 2^{ème} catégorie au pluriel : *notre* ;

(4)-CC Concernant les aspects pragmatiques du récit, les *résultats obtenus confirment* partiellement *notre* hypothèse de départ.

(21) CC& R- *Ici, nous* n'avons pas parlé d'expertise, mais de spécialité car le domaine d'expertise n'était pas l'objet de l'étude ; cependant, ce que montre la situation C ainsi que la comparaison des situations, c'est que la spécialité est une attitude contingente, toujours susceptible d'être renversée par une autre attitude.

(25)-R Dans *cet article, nous* avons *présenté* les diverses *études menées* en français pour *tester* dans cette langue et dans cette culture, la généralisabilité du second postulat de la théorie de la catégorisation à fondement émotionnel

(26)-CC& R Dans *cette catégorie, nous* *distinguerons* les extensions temporelles vers le passé ou le futur, l'investissement affectif lié à l'histoire, les indices de possession, avec l'utilisation des adjectifs possessifs.

Notons que le pronom *nous* est associé au pronom *on* et aux *il* impersonnel. Toutefois, l'absence du *je* peut laisser supposer qu'il est également associé au pôle formalisation dans certaines occurrences de phrases-clé.

F. Rinck (2006, 308) soutient l'idée selon laquelle ces

« marques personnelles sont utilisées dans le même sens pour définir l'objet traité dans le texte. L'énoncé peut avoir pour but de spécifier des exclusions ou inversement, de circonscrire l'objet en termes positif »

comme nous pouvons le voir à travers les deux exemples suivants :

(27)- CC *Nous* ne *discuterons* pas *ici* de la validité de *ces propos*.

(28)- R *Notre* *objectif* était de *tester* la validité de *ce modèle* en fonction d'un autre contenu ; nous pensons l'avoir atteint

Certaines occurrences du pronom *nous*, peuvent-être associées à une valeur potentiellement collective et donc interprétées comme une représentation qui associe l'auteur et le lecteur. C'est notamment le cas des exemples qui suivent

(29)- D'après les *résultats illustrés ici*, **nous** pouvons constater que la progression entre les deux échelles est parallèle et équilibrée.

Au sein de cet exemple, le lecteur se trouve mis en scène dans la réalisation du procès sous forme de modélisation « *pouvoir* ».

Dans sa démarche énonciative, l'auteur emploie parfois des formes de discours indirect et se joint à son lecteur en se faisant passer pour un lecteur parmi d'autres tout en étant capable de décoder l'interaction du texte et du lectorat comme illustré dans les deux exemples suivants ;

(30)-CC *Cependant*, **il faut bien admettre** qu'un tel *travail* sera particulièrement *difficile* à réaliser car les *tests* d'intelligence usuels présentent un pouvoir discriminant limité pour les sujets retardés et ne *permettent* pas, le plus souvent, d'obtenir des mesures d'efficacité intellectuelle en deçà du retard mental modéré.

(31)- CC S'il apparaît *clair*, après cette première expérience, que la reconnaissance visuelle de mots n'est pas dirigée par la langue mais par des données formelles amenant à des compétitions lexicales entre des candidats des deux langues, **il faut se poser la question** du moment où l'information concernant la langue des mots devient active.

Dans ces exemples, l'auteur de l'article s'immisce dans le texte et décode ce que celui-ci dit à son lecteur. Dans ces structures, l'auteur va jusqu'à adopter momentanément le point de vue de l'auteur étudié pour mieux faire parler le texte (F. Rinck, 2006, 312).

C'est le seul emploi de la 3^{ème} personne linguistique, *il*, qualifiée de **non-personne** correspondant au délocuté par Benveniste (1966) au sein des phrases-clé. Cet usage correspond à l'effacement énonciatif de l'auteur caractéristique de l'écrit scientifique.

Ces analyses d'emploi du pronom personnel *nous* dans les phrases exprimant le problème de recherche dans les articles de sciences de l'éducation, nous conduisent à relever deux emplois de ce pronom : emploi comme marque personnelle qui renvoie à l'auteur, et inclusion du lecteur dans l'emploi du *nous*.

Conclusion

Tout au long de ce chapitre, nous avons décrit la méthode à laquelle nous avons fait appel afin d'identifier les FD, signaux susceptibles d'annoncer la formulation du problème de recherche dans l'écrit scientifique. En effet, nous nous étions fixés pour tâche de définir les conditions nécessaires à la formalisation d'un système d'identification de ces associations lexicales, qui permettent aux auteurs scientifiques de mettre en évidence leur question de recherche ; pour atteindre cet objectif, il s'agissait de leur attribuer les fonctions rhétoriques de *résumé* et de *changement cognitif* selon le cas de figure si l'auteur résume l'article entier ou bien une partie de l'article ou si celui-ci signale le problème traité au sein de l'article par l'expression d'un changement cognitif .

Pour ce faire, nous avons eu recours à un modèle conceptuel, dit « concept-matching » intégré à la fonctionnalité XIP qui nous a permis d'identifier les FD mobilisées dans des articles en sciences de l'éducation en français.

A l'aide du formalisme XIP, nous avons pu enrichir le lexique par les traits conceptuels, et construire des règles de reconnaissance qui les utilisent. Ainsi, les dépendances syntaxiques extraites se prêtent à la construction des règles de cooccurrence, qui permettent la reconnaissance de nos FD.

Les conjonctions typiques de concepts constituants liés par des dépendances syntaxiques pour la réalisation des deux FD signalant que les problèmes scientifiques sont présentées.

Ces concepts constituants se rangent entre autres au sein de trois des quatre grandes sphères conceptuelles définies par Pecman (2005) ; la sphère de la scientificité, de la discursivité et de la modalité.

Ce système de traitement du langage naturel est basé sur un modèle d'analyse de discours, inspiré entre autres des travaux de M. Pecman (2004, K. Hyland 2005, Rick, 2006, A.Tutin, 2007b-2010). Le modèle fournit une représentation d'éléments discursifs qui accomplissent des fonctions rhétoriques telles que *changement cognitif et résumé* en intégrant des analyses conceptuelles et syntaxiques.

Nos analyses nous ont permis de faire le constat selon lequel les FD ont un rôle significatif pour signaler les contenus relevant du problème de recherche dans les articles qui ont pour champs disciplinaire les sciences de l'éducation.

Afin de rendre compte des caractéristiques linguistiques des phrases-clé où figurent les associations lexicales donnant lieu aux FD, nous nous sommes intéressés à la structure de la phrase, aux temps verbaux employés par les auteurs et à l'inscription de l'auteur dans son texte. Toutefois, ces analyses requièrent des travaux supplémentaires sur un corpus plus important afin qu'il soit possible de généraliser nos observations sur les phrases-clé traitées dans cette étude.

Chapitre 4 : Limites & évaluation du système

Dans cette partie, nous tenterons, d'une part, de soulever les problèmes rencontrés avec notre système. Nous aborderons également la question de savoir dans quelle mesure ces problèmes peuvent être résolus en partie 5.2. D'autre part, nous procéderons à l'évaluation du système à travers l'analyse d'une dizaine d'articles transmis à des évaluateurs d'écrits scientifiques qui avaient pour consigne de souligner manuellement les phrases qui traduisent le problème de recherche. L'objectif de cette démarche est de vérifier que notre système est pertinent dans l'aide à la lecture experte en comparant les phrases identifiées manuellement à celles identifiées par notre système.

Nous avons également fait le choix de revenir sur une étude antérieure réalisée avec Á. Sándor (à paraître)⁴⁴ où nous testons la pertinence de notre système en le comparant avec des résumés d'articles rédigés par des experts en sciences de l'éducation toujours dans le cadre du Projet européen EERQI. Nous avons tenté de voir dans quelle mesure les résumés étaient comparables à une ou plusieurs phrases donnée(s) figurant dans l'article qui avait pour fonction *résumé*.

4.1. Evaluation du système

Afin de procéder à l'évaluation de notre système de détection de phrases véhiculant le problème de recherche traité dans l'article, nous avons transmis 15 articles à des enseignants de l'Université de STENDHAL- Grenoble 3 qui font partie du comité de relecture notamment pour la revue Linguistique et didactique des langues (LIDIL)⁴⁵ évaluateurs d'écrits scientifiques et avons donné pour consigne d'identifier manuellement le problème traité dans chaque article de recherche.

Parallèlement, nous avons transmis un seul article à ce même public en lui donnant la même consigne, à savoir surligner les phrases où l'auteur formule son problème de recherche.

⁴⁴ A paraître dans l'ouvrage « Stéréotypie et figement » qui sera publié dans les Presses Universitaires de Rennes.

⁴⁵ <http://lidil.revues.org/>

Le but étant d'estimer le pourcentage où la même phrase-clé comportant des FD est surlignée par plusieurs experts.

Rappelons que l'objectif de cette démarche est de tester la pertinence de notre système dans l'aide à la lecture experte.

L'auteur de l'article scientifique renvoie au problème de recherche en déclarant ses buts, problèmes spécifiques et conclusions. Comme nous l'avons soutenu, les FD de *résumé* et *changement cognitif* véhiculent certains de ces contenus au sein des articles qui expriment le problème de recherche véhiculant la problématique scientifique en général.

Pour rendre compte de l'usage de ces FD signalant les problèmes de recherche dans les articles de sciences de l'éducation, nous avons mené une expérience dans la continuité du cadre du projet EERQI, parallèlement à notre collaboration avec le projet de recherche SCIENTEXT.

Malgré nos différentes relances aidées de mon directeur de thèse, nous n'avons pas réussi à avoir le retour escompté de la part des experts. C'est pourquoi, nous avons choisi d'intégrer une étude antérieure (voir parti 4.2 Comparaison avec une étude antérieure) afin de renforcer l'évaluation de notre système.

En effet, parmi les 10 différents articles adressés aux 75 enseignants et relecteurs, nous n'avons reçu que 5 retours d'articles avec 25 phrases surlignées.

Quant au retour par rapport à l'identification du problème de recherche au sein d'un même article, 7 experts nous ont répondu avec 49 phrases surlignées.

Nous procéderons donc dans ce qui suit à l'analyse de 12 articles dont 7 en commun avec un ensemble de 74 phrases identifiées par les experts comme exprimant un problème de recherche.

Afin d'évaluer la pertinence de ce dernier, nous nous sommes arrêtée sur chacune des phrases surlignées par les experts parallèlement au sein du même et dans les différents articles recueillis et nous avons tenté de voir ;

- si celles-ci étaient comparables aux phrases identifiées par XIP,
- si elles comportaient des FD,
- quelle était la précision de XIP,
- nombre moyen de phrases détectées par article à savoir le nombre moyen de la même phrase identifiée par plusieurs experts.

Les résultats de ces analyses sont retranscrits dans la figure 59.

	Parmi les 7 articles similaires	Parmi les 5 articles différents						
Pourcentage de phrases surlignées par les experts comparables aux phrases détectées par XIP	30.61%	41.37%						
Pourcentage des phrases détectées par XIP comportant les concepts-constituants dans les articles	44.89%	79.31%						
Précision estimée de XIP	95.91%	93.10%						
Nombre moyen de phrases par article détectée par XIP	15	12						
Nombre de phrases comportant des formules de discours selon la fréquence d'utilisation de la même phrase par plusieurs experts	Nombre de fois où la même phrase d'un article est surlignée par plusieurs experts							
	5	3	1	2	1	1	1	1

Figure 59 : Evaluation du système

Notre expérience indique également dans quelle mesure les phrases détectées par XIP peuvent être considérées comme pertinentes dans la compréhension de l'essentiel de l'article.

Comme nous pouvons le constater à travers la figure 59, l'analyse des résultats montre que plus de 30% de phrases surlignées par les experts parmi les 7 articles similaires et plus de 41% parmi les articles différents sont comparables aux phrases détectées par XIP. Nous tenons à signaler que dans le champ du TAL, un tel résultat n'est pas mauvais, contrairement à ce qu'un non spécialiste du champ pourrait croire.

L'illustration suivante [(32)- CC&R] est une phrase surlignée par les experts (voir annexe 4.2) et identifiée comme véhiculant un problème de recherche que nous retrouvons dans l'exemple suivant :

(32)- CC&R Néanmoins, le caractère portatif et les limitations de l'outil posent des *défis* aux niveaux institutionnel et pédagogique que *nous présentons* dans *cette étude*.

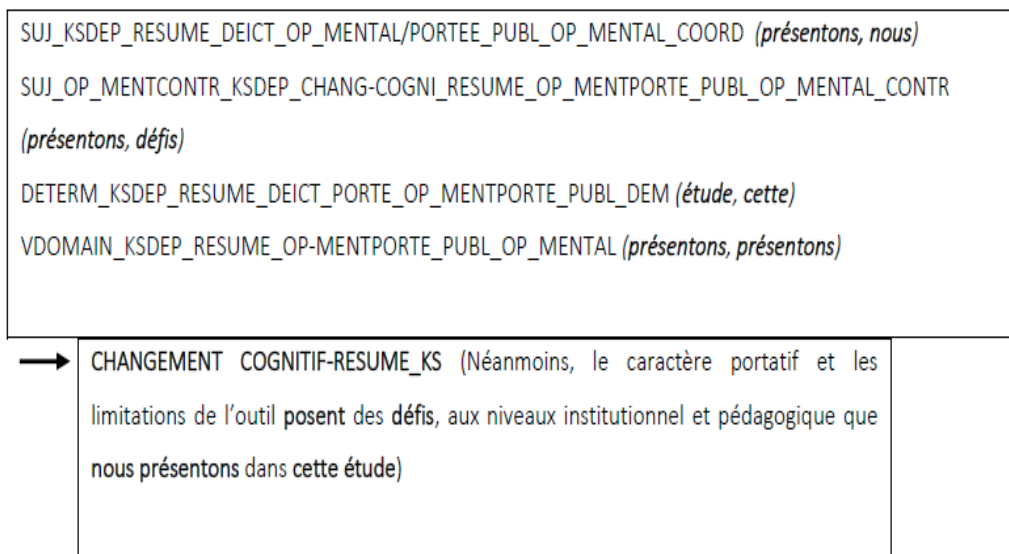


Figure 60 : Extraction d'un exemple de phrase surlignée par les experts et identifiée par XIP

L'exemple (32)-CC&R est une illustration d'une phrase surlignée par les experts et identifiée par XIP.

Ces résultats impliquent que ce sont des phrases associées au concept de *changement cognitif* et de *résumé*, où figurent ces formules en particulier, qui sont détectées. Ceci indique que les FD ont bien un rôle significatif pour signaler des contenus relevant du problème de recherche dans les articles

Nous avons également trouvé que plus de 44 % de ces phrases dans la première catégorie et plus de 79% dans la seconde, comportaient les concepts constituants identifiés comme appartenant au lexique scientifique, sans pour autant donner lieu aux concepts généraux *changement cognitif* et de *résumé*. Cette non-identification par XIP de ces phrases comme traduisant un problème de recherche, est due au fait que la condition de dépendance syntaxique n'est pas remplie.

En ce qui concerne la performance de XIP, une moyenne de 35% entre les deux identifications des phrases contenant des FD ont été détectées automatiquement. Le problème d'analyse liée à XIP concerne le fait que les dépendances syntaxiques des associations lexicales ne sont pas détectées. L'analyse détaillée des phrases non détectées

fait partie des travaux à venir. En outre, nous n'avons relevé que très peu de phrases détectées par XIP qui ne contiennent pas des FD.

Afin d'estimer la performance de XIP, nous avons choisi de rendre compte de la précision de cet outil. Rappelons que la précision montre quel est le pourcentage de tous les résultats que la machine considère comme juste par les évaluateurs humains. Le calcul de la précision est établi par le rapport nombre des phrases correctes trouvées par la machine divisé par le nombre de toutes les phrases trouvées par la machine.

Comme nous pouvons le constater à travers les résultats obtenus, nous avons un pourcentage de plus de 95% pour la première catégorie et plus de 93 pour la seconde en terme de précision ce qui rend compte de la performance XIP.

Enfin, concernant le nombre de phrases comportant des FD selon la fréquence d'utilisation de la même phrase par plusieurs experts, nous remarquerons que les résultats sont mitigés. Nous retrouvons uniquement 2 mêmes phrases surlignées 5 et 3 fois par les experts.

Nous pouvons donc conclure que ces phrases signalent bien la formulation d'un problème de recherche par les auteurs au sein des articles étudiés et que les FD sont des moyens linguistiques privilégiés pour véhiculer le problème de recherche.

Ce système de reconnaissance des FD véhiculant les problèmes scientifiques appliqué dans le cadre du projet EERQI, peut donc être proposé comme aide à la lecture experte comme illustré dans la figure 62. Parfois, nous pouvons rencontrer des phrases où les deux concepts sont exprimés comme c'est notamment le cas des phrases figurant dans la figure 61 :

CC & R- Toutefois, mon propos sera moins ici de discuter les résultats de l'étude et ses implications théoriques que d'explicitier la méthodologie à l'œuvre.

CC & R- Le but de cet article est de montrer le rôle que peuvent jouer la formalisation et la modélisation dans l'étude des pratiques pédagogiques, d'en montrer les apports et les limites.

Figure 61 : Illustration de phrases exprimant le concept de *résumé* et de *changement cognitif* au sein de la même phrase

Formalisation et modélisation dans les sciences sociales: une étude de la construction du jugement des enseignants

Pascal Bressoux

Changement cognitif

(Description du problème de recherche)

Le **but** de **cet article** est de **montrer**, sur la base d'une **étude** de la construction du jugement des enseignants, les fondements méthodologiques à l'œuvre. **On se centrera** moins sur les **résultats** de l'**étude** et leurs implications théoriques que sur l'explicitation de l'articulation entre la problématique de recherche et la méthodologie. **En se fondant au préalable sur des réflexions épistémologiques générales et sur des réflexions épistémologiques propres aux sciences sociales, on essaiera de montrer** l'apport, les **difficultés** et les **limites** de la mise en œuvre d'un raisonnement de type expérimental dans des études réalisées en milieu naturel. **On tentera de montrer** le rôle d'une modélisation formelle pour l'étude des phénomènes éducatifs, en particulier dans son lien avec l'empirie : la distinction entre faits virtuels et faits actuels, de même que la prise en compte du probable, nous aideront dans cette voie.

Résumé du problème de recherche

Mots-clés : modélisation, formalisation, faits virtuels, faits actuels, jugement, probabilité.

Résumé + Changement cognitif

INTRODUCTION

Les pratiques pédagogiques sont diverses et variées : modes de gestion et d'organisation de la classe, interactions verbales, utilisation du temps scolaire, découpage des activités, modalités d'évaluation, etc. Ces pratiques, qui semblent si familières à tous ceux qui ont fréquenté l'École, ne se laissent pourtant pas appréhender facilement et elles semblent rétives à une approche globale et unitaire ; les instruments, les modes d'approche, les méthodologies différents d'une étude à l'autre. C'est sans doute que les pratiques pédagogiques ne sont pas *en elles-mêmes* un objet de recherche ; leur constitution en tant qu'objet scientifique suppose une structuration qui impose de réaliser certains choix, d'opérer des simplifications, de renoncer à une appréhension globale et entachée.

Cet article traite d'une pratique pédagogique particulière : le jugement que les enseignants portent sur la valeur scolaire de leurs élèves. On cherche à montrer grâce à une étude

empirique que, loin d'être le simple reflet d'une stricte valeur scolaire qui serait fondée sur les performances effectives des élèves, le jugement de l'enseignant définit une valeur sociale, en ce sens qu'il se fonde en partie sur des normes sociales générales telles que la norme d'internatité.

Toutefois, **mon propos** sera moins *ici* de discuter les **résultats** de l'**étude** et ses implications théoriques que d'explicitier la **méthodologie** à l'œuvre. L'**étude** obéit à un raisonnement de type expérimental qui **visé** à **établir** des liens de causalité entre certaines variables et le jugement des enseignants.

Pour autant, l'étude n'est pas une expérimentation réalisée en laboratoire. Elle a été opérée en milieu naturel, ce qui empêche un contrôle strict de l'ensemble des éléments potentiellement influents. Prenant acte de cette difficulté, il s'agira d'établir un modèle formel susceptible de mettre en évidence les facteurs qui influent sur la construction du jugement des enseignants. **Le but** de **cet article** est de **montrer** le rôle que peuvent jouer la formalisation et la modélisation dans l'étude des pratiques pédagogiques, d'en **montrer** les **apports** et les **limites**.

Formules de discours scientifique

Figure 62 : Extrait d'un article analysé : les formules de discours sont surlignées, en italiques et en gras, les phrases *résumé* sont surlignées en

La figure 62 illustre un article où les FD sont en gras, italiques et soulignées, les phrases *résumé* en gris clair et les phrases exprimant un *changement cognitif* en gris foncé. Ce système est adapté des travaux d'Á. Sándor et al. (2009) dans le cadre du même projet EERQI.

4.2. Problèmes rencontrés

Au fur et à mesure de la mise en place de notre système, nous avons constaté que celui-ci présentait des limites.

Parmi les problèmes que nous avons rencontrés, nous mentionnerons en premier lieu le problème de *coréférence*.

C'est le cas de l'exemple (33)-CC avec une phrase pertinente qui n'a pas été détectée, où le pronom personnel *elle* fait référence à « étude » figurant dans la phrase précédente.

(33)-CC [L'étude tente d'une part de caractériser ces images identitaires en utilisant une méthodologie d'analyse inductive des données brutes.] D'autre part, *elle vise à apporter des éléments* de compréhension des modifications de ces images identitaires au cours du temps.

En effet, cette phrase représente bien le problème de recherche traité. En revanche, notre système ne la détecte pas, car, il aurait fallu que le pronom *elle* soit associé au trait PORTEE. Or, si ce trait était attribué à cette lexie, tous les pronoms du corpus seraient associés à ce trait ce qui fausserait la pertinence de notre système et générerait beaucoup de bruit.

Nous avons tenté de traiter ce problème. Toutefois, sa résolution reste un projet très ambitieux qui pourrait être un sujet en soi pour des travaux à venir. Nous tentons de proposer dans le chapitre 5, une solution pour pallier ce problème et assurer la pertinence de notre système.

Un autre problème relève du fait que dans certaines disciplines comme la psychologie de l'éducation qui relève du domaine des sciences de l'éducation, le sujet de l'article relève de la résolution d'un problème. Dans ce cas, notre système prélève des phrases qui traitent d'un problème mais qui n'est pas à un méta-niveau comme nous pouvons le voir à travers l'exemple (34). Or, rappelons-le, les FD telles que nous les avons définies relève d'un méta-niveau et sont dissociées du sujet traité dans l'écrit scientifique. En effet, le sujet traité dans l'article dont est extraite cette illustration traite du travail d'un chercheur et des problèmes que celui-ci rencontre.

(34)- Le *chercheur* se donne pour *but* de *connaître* les pratiques des enseignants dans leurs réalisations concrètes *mais* l'accès aux terrains d'enquête comporte nombre de difficultés.

Dans ce cas de figure, il nous est difficile de faire en sorte de corriger ce problème. En effet, si nous dissociions le trait PORTEE de la lexie *chercheur*, cela s'appliquera pour toutes les autres occurrences de notre corpus ce qui induira à la non-identification de phrases pertinentes. Nous revenons dans la partie 5.2 sur la solution éventuelle de ce problème.

Nous avons rencontré d'autres problèmes d'ordre technique. La base de données Scientext recense un corpus sous différents formats ; Pdf, txt, etc. Or tous les formats textes ne peuvent être analysés par XIP ce qui nous oblige à extraire les contenus sous format txt/xml et ainsi à perdre des données importantes entre autres : tableau, schémas, figure, etc. Mais encore, peu de formats de citations sont pris en compte jusqu'à présent.

Certaines phrases d'articles de recherche apparaissent comme véhiculant des messages importants. Cependant, ces dernières ne sont pas détectées. Ci-dessous un exemple (35)-I⁴⁶ de phrase que nous pourrions étiqueter sous la dénomination « importance », et qui n'est pas identifié.

(35)-I *L'importance* du regard porté par *cette discipline* universitaire réside dans une *analyse* plurielle (d'où l'utilisation du pluriel du mot « science ») qui dépasse la simple juxtaposition de plusieurs lectures disciplinaires et qui intègre la *mise en relation* des regards.

En effet, bien que cette phrase n'ait pas été identifiée car elle ne véhicule ni de *changement cognitif* ni de *résumé*, elle se trouve être importante car elle vient mettre l'accent et compléter la formulation des problèmes de recherche.

4.3. Evaluation du système

4.3.1. Comparaison avec une étude antérieure

Dans une recherche antérieure menée avec Á. Sándor (à paraître), dans le cadre du projet EERQI, nous avons demandé à des experts en sciences de l'éducation de résumer brièvement les buts, les problèmes et les conclusions des articles. Nous nous sommes arrêtées sur chacune des phrases rédigées par les experts et nous avons tenté de voir si elles étaient comparables à une ou plusieurs phrases donnée(s) figurant dans l'article. Dans les cas où nous avons trouvé des phrases comparables, nous nous sommes

⁴⁶ « I » correspond au concept constituant « importance ».

demandés si ces phrases comportaient une ou plusieurs formule(s) de discours ; et si oui, si elles étaient détectées ou non par XIP. Ainsi, notre expérience indique également dans quelle mesure les phrases détectées par XIP peuvent être considérées comme pertinentes dans la compréhension de l'essentiel de l'article. Les résultats d'une telle expérience ne peuvent apporter que des indications approximatives, et non pas des chiffres exacts, pour plusieurs raisons. Premièrement, il peut y avoir des différences considérables de style parmi les auteurs, et dans un petit échantillon, nous ne pouvons pas assurer une représentativité équitable des styles. Il peut en outre y avoir également des différences de style indépendamment des styles personnels, suivant le genre de l'article (théorique, expérimental, revue de la littérature, etc.), aussi bien que la sous discipline (psychologie, histoire, philosophie, etc.). Enfin, il peut y avoir des différences parmi les experts : les différents experts peuvent s'inspirer à des degrés divers des formulations originales de l'article pour rédiger leur résumé, et les résumés peuvent également être plus ou moins détaillés. L'analyse des résultats montre que la moitié des phrases des résumés experts sont comparables à une ou plusieurs phrases des articles. Ces phrases des articles signalent donc la problématique dans les articles. Nous avons également trouvé que 40% de ces phrases comportaient des FD. Ceci indique que ces éléments sont des moyens linguistiques privilégiés pour véhiculer la problématique. En ce qui concerne la performance de XIP, un peu plus de la moitié des phrases contenant des FD ont été détectées automatiquement. Les problèmes d'analyse liés à XIP sont de deux types : soit les mots ou expressions dans la formule ne font pas partie du lexique composant les concepts constituants, soit les dépendances syntaxiques ne sont pas détectées par XIP. L'analyse détaillée des phrases non détectées, fait partie des travaux à venir. En outre, nous n'avons relevé que très peu de phrases détectées par XIP qui ne contiennent pas des FD.

Chapitre 5 : Conclusion générale et perspectives

5.1. Conclusion

Le présent travail exploite les ressources textuelles et lexicales constituées pour « l'étude linguistique du positionnement et du raisonnement » au sein du projet Scientext, pour le développement d'un système d'analyse automatique qui détecte les passages annonçant la formulation du problème de recherche dans les publications en science de l'éducation. Il a été effectué au sein du projet européen Educational Research Quality Indicators.

L'auteur de l'article scientifique renvoie à son problème de recherche en déclarant ses buts, problèmes spécifiques et conclusions. Comme nous l'avons soutenu, les FD de *résumé* et *changement cognitif* véhiculent certains de ces contenus au sein des articles.

A travers la méthodologie du *concept-matching*, nous avons tenté tout au long de ce travail d'identifier les phrases signalant des problèmes de recherche dans les articles de sciences de l'éducation, à partir de l'identification de « formules de discours » stéréotypées.

Nous avons défini les FD comme des suites de mots ou d'expressions syntaxiquement cohérentes, mais pas nécessairement contigües, correspondant à des structures syntaxiques et composantes lexicales variées.

Ces FD sont des associations lexicales dont les lexèmes peuvent appartenir à différents lexiques de l'écrit en général ; transversal, abstrait non spécialisé, celui de la langue générale-commune, et enfin le lexique des activités intellectuelles.

Elles diffèrent parfois des expressions appelées des « marqueurs de métadiscours » (K. Hyland 2007) car toutes nos associations lexicales ne sont pas répertoriées parmi les marqueurs de métadiscours, des collocations ou unités phraséologiques (M. Pecman 2005, A. Tutin 2007b), ainsi que des marqueurs de discours (D. Siepmann 2007, F. Boch et al. 2007) à cause de leur transparence et de leur non-figement, bien qu'elles ne constituent pas pour autant des propositions autonomes comme nous avons pu le voir au sein des chapitres 2 et 3.

Les FD telles que nous les avons définies dans la présente étude indiquent les fonctions rhétoriques tels que le *changement cognitif* ou encore le *résumé*. Rappelons que nous avons déterminé ces fonctions à un niveau conceptuel du discours.

Dans le cadre de l'analyse de « concept-matching », nous décrivons ces FD comme la modélisation de concepts constituants liés à l'argumentation académique. Les traits pertinents sont définis à un niveau conceptuel associés. Nous retrouvons entre autres : DEICTIC (*Je, nous, notre*), CONTRAST (*question, débat, doute, contester, rejeter, cependant, à l'opposé*), OPERATION MENTALE (*étudier, analyser, réflexion, raisonnement*), PORTEE (*travail, analyse, idée*), PUBLICATION (*texte, article, étude*), POSSIBILITE (*possible, peut*), IMPORTANCE (*primordial, important*).

Les FD sont donc caractérisées à différents niveaux :

- au plan rhétorique, il s'agit des éléments signalant au lecteur le fait qu'un problème de recherche va être mentionné dans l'énoncé
- au plan conceptuel, il s'agit des combinaisons des concepts constituants.
- au plan syntaxique, cette structure conceptuelle s'incarne elle-même dans une structure linguistique syntaxiquement liée au moment de l'actualisation en discours

Nous avons décrit le développement du système adopté et avons présenté une comparaison entre les phrases détectées automatiquement et celles identifiées manuellement par des enseignants et relecteurs. Le but est d'estimer la pertinence de notre système. Les résultats (voir partie 4.1.) montrent que les FD jouent un rôle important dans l'identification du problème de recherche formulé par les auteurs.

Nous n'avons pas pu présenter les règles informatiques implémentées sous XIP permettant l'identification des FD, exprimant le problème de recherche, pour des questions de propriétés intellectuelles et de confidentialité auprès de XRCE.

Nous n'avons pas pu, non plus, présenter en annexes les différentes ressources linguistiques et techniques mises en place à travers notamment la compilation de notre lexique scientifique, des associations lexicales et autres éléments lexicaux et techniques auxquels nous avons fait appel lors de nos analyses de corpus pour les mêmes raisons.

5.2. Perspective

Dans cette section, nous décrivons nos perspectives à court terme, puis à plus long terme. Nos perspectives à court terme auront pour objet de répondre à chacune des questions suivantes :

- a) Comment résoudre le problème de coréférence ?
- b) Comment résoudre le problème de confusion de méta-niveau si le sujet de l'étude traite de la résolution d'un problème ?
- c) Comment élargir la conceptualisation de notre lexique des écrits scientifiques afin de couvrir toutes les phrases susceptibles d'apporter un support à celles véhiculant un problème de recherche ?

Nos perspectives, à plus long terme, portent sur l'application de notre approche à des corpus d'autres langues ainsi que la mise en place d'une plate-forme qui permettrait de traiter les textes scientifiques, en obtenant en sortie, les phrases véhiculant le problème de recherche, surlignées par exemple sans passer par l'analyseur XIP qui requiert un droit d'accès et l'obligation de détenir une licence. Il serait également intéressant d'envisager d'appliquer le modèle du concept-matching pour la détection d'autres fonctions rhétoriques, et également d'autres éléments du discours.

5.2.1. Perspectives à court terme

Il est primordial pour tout auteur de respecter les règles de la cohérence, notamment celles liées à la répétition. C'est pourquoi, il se voit contraint de faire appel à l'anaphorisation dont l'emploi est considéré comme un facteur indispensable à la construction d'un texte quel qu'il soit. L'emploi de ces anaphores réduit graduellement la performance de notre système car il se trouve que plusieurs phrases portant sur le problème de recherche ne peuvent être identifiées à cause de cet emploi anaphorique, faisant référence à notre lexique répertorié comme spécifique aux écrits scientifiques.

Nous pensons que ce problème peut être éventuellement résolu en intégrant la détection d'anaphore dans l'analyse syntaxique. Toutefois, nous sommes toujours en pleine réflexion afin de voir comment procéder pour éviter que l'emploi anaphorique soit généralisé à tous les pronoms, ce qui aurait pour effet de réduire la performance de notre système d'identification de phrases-clé.

Concernant le problème de confusion de métaniveau, il reste propre à la discipline des sciences de l'éducation et peut-être également au domaine de la sociologie et des sciences du langage. En effet, il apparaît peu probable qu'en sciences appliquées ou expérimentales, on traite de sujet portant sur *le chercheur*, *la recherche* ou encore sur la

formulation de *l'hypothèse* par exemple. Retravailler le corpus des sciences de l'éducation nous apparaît donc comme la solution afin de résoudre ce problème. Toutefois, à l'heure actuelle, nous continuons à nous interroger sur la méthode conceptuelle la plus pertinente pour améliorer la performance de notre système à ce sujet. Nous n'avons donc pas trouvé de solution fiable pour le résoudre. Nous continuons à tenter de clarifier et d'approfondir la notion de FD et de résoudre les problèmes rencontrés de façon à renforcer notre système et à pouvoir généraliser cette méthode sur d'autres genres d'écrits scientifiques.

Quant à la question de savoir si nous devons élargir la conceptualisation de notre lexique des écrits scientifiques, afin de couvrir toutes les phrases susceptibles d'apporter un support à celles véhiculant un problème de recherche, nous tendons à répondre de manière positive. En effet, nous avons observé qu'il y avait des phrases dont les noms, verbes ou adjectifs, que nous pouvons qualifier de caractérisation, qui sont répertoriés dans notre lexique des écrits scientifiques, et qui sont en relation syntaxique avec d'autres éléments linguistiques de notre lexique scientifique véhiculant des contenus importants susceptibles d'introduire dans ce qui suit un problème de recherche.

5.2.2. Perspectives à long terme

La perspective sur du long terme serait dans un premier temps une adaptation du système de détection des FD à d'autres langues. Rappelons que cette méthode de concept-matching à des niveaux rhétoriques différents a été adoptée dans les nombreux travaux de Á. Sándor (2007, 2009, 2010) notamment sur un corpus de biologie sur des articles rédigés en anglais (Á. Sándor, 2007). Nous l'avons également mise en œuvre dans une étude antérieure avec Á. Sándor (2015) où nous avons utilisé les mêmes mesures que celles qui ont été détaillées dans la partie 4.2 pour des résumés d'articles en anglais. Le corpus de 189 résumés de 44 articles de sciences de l'éducation a également été compilé par l'équipe d'EERQI. Cette comparaison montre que les résumés anglais synthétisent moins les articles que les résumés français, ce qui montre que les résumés d'experts anglais sont davantage basés sur les phrases des articles. Cela peut indiquer des différences soit dans les manières de résumer, soit dans le style des articles dans les deux langues. La couverture des phrases comparables comportant des FD est également plus grande en anglais. Ceci signifie que le rôle des FD est plus important dans les articles

anglais que dans les articles français. On voit également que la couverture de XIP est plus grande en anglais qu'en français.

Ceci est probablement prioritairement dû au plus long développement de la grammaire utilisée par XIP en anglais qu'en français, car il faut rappeler que l'étude n'a été réalisée que sur 25 articles en français ce qui n'a pas permis de développer une grammaire étendue contrairement à l'anglais. Cette meilleure couverture résulte aussi dans le plus grand nombre de phrases détectées par article en moyenne en anglais par rapport au français. La précision de XIP est bonne dans les deux langues.

Dans notre approche, la partie linguistique est liée essentiellement à trois ressources : un analyseur syntaxique, une ressource terminologique et un corpus de textes en langage naturel. L'adaptation à une autre langue dépend donc de la disponibilité de ces trois ressources.

Il faudrait voir s'il est possible d'indiquer pour chaque terme défini, la langue à laquelle celui-ci appartient. Cela permettrait de définir pour une même classe conceptuelle les déclinaisons dans différentes langues de son terme préféré et de ses termes alternatifs.

Le passage à une autre langue ne nécessiterait donc pas la création d'une nouvelle terminologie mais uniquement la traduction des termes déjà existants. En outre, les corpus utilisés ne nécessiteraient aucune pré-annotation.

Il serait également intéressant pour des travaux futurs d'appliquer le modèle du concept-matching pour la détection d'autres fonctions rhétoriques, et également d'autres éléments du discours comme par exemple les expressions qui indiquent la causalité. En ce qui concerne la mise en œuvre, nous devons travailler sur des méthodes générales pour construire facilement notre système de concept-matching.

Concernant la mise en place d'une plate-forme qui permettrait d'insérer les écrits scientifiques et d'avoir en sorties les phrases véhiculant le problème de recherche, il est possible d'implémenter le module de concept-matching dans n'importe quel système de traitement du langage naturel qui permet de requérir à l'analyse de dépendances. Car l'usage de XIP n'est pas accessible à tous et requiert un droit d'utilisation à travers l'achat de la licence.

Références

- Adam, J.M. (1999). *Linguistique textuelle. Des genres de discours aux textes*. Paris: Nathan Université.
- Adel, A. (2006). *Metadiscourse in L1 and L2 English*. Philadelphie: John Benjamins.
- Aït-Mokhtar, S., Chanod, J- ,et Roux, R. (2002). Robustness beyond shallowness: incremental dependency parsing. *Natural Language Engineering*, 8(2/3), 121-144.
- Aït Saïdi, S. et Sándor, Á. (2005). Analyse et détection de formules de discours signalant la problématique dans les genres scientifiques. *Stéréotypie et figement à l'origine du sens*, 205-215, Presses Universitaires du Mirail (PUM).
- Aktouf, O. (1987). *Méthodologie des sciences sociales et approche qualitative des organisations*. Les Presses Université du Québec : Québec
- Authier-Revuz, J. (1982). Hétérogénéité montrée et hétérogénéité constitutive. *DRLAV*, 26. Université Paris VIII, Saint-Denis.
- Bazerman, C. (1988). *Shaping written knowledge. The Genre and Activity of the Experimental Article in Science*. Madison: University of Wisconsin Press.
- Bartsch, S. (2004). *Structural and Functional Properties of Collocations in English. A Corpus Study of Lexical and Pragmatic Constraints on Lexical Cooccurrence*. Tübingen: Narr.
- Berthelot, J-M. (2003). *Figure du texte scientifique*. Paris : Presse Université de France.
- Bloquet, G, & Tutin, A. (2004, mai). Lexique transdisciplinaire et structure des articles scientifiques. *Sciences et écritures*, 13-14. Besançon.
- Boch, F., Grossmann, F. & Rinck, F. (2007). Conformément à nos attentes ... : les marqueurs de convergence/divergence dans l'article de linguistique. *Revue Française de Linguistique Appliquée*, 7 (2) ,109-122.
- Boch, F. (2010). Énonciation et rhétorique dans l'écrit scientifique. *Lidil. Revue de linguistique et de didactique des langues*, (41).

- Bordet, G. (2011). *Etude contrastive de résumés de thèse dans une perspective d'analyse de genre*, Thèse de doctorat, Université Paris 7-Paris Diderot.
- Bourdieu P. (1982). *Ce que parler veut dire. L'économie des échanges linguistiques*. Paris : Fayard.
- Bourigault, D. (2007). *Un analyseur syntaxique opérationnel : SYNTAX*, Thèse d'habilitation à diriger des recherches, Université Toulouse le Mirail.
- Bronckart, J. (1996). *Activité langagière, texte et discours*. Pour un interactionnisme socio-discursif. Paris : Neuchâtel.
- Bourigault, D., & Fabre, C. (2000). Approche linguistique pour l'analyse syntaxique de corpus. *Cahiers de Grammaire*, 25, 131-151
- Bunton, D. (1999). The use of higher level metatext in Ph.D. theses. *English for Specific Purposes*, 18, 41-56
- Chavez, I. (2008). *La démarcation dans les écrits scientifiques*, Mémoire de Master 2, Université Stendhal – Grenoble III.
- Cavalla, C. & Grossmann, F. (2005). Caractéristiques sémantiques de quelques 'Noms scientifiques' dans l'article de recherche en français, *Akademisk Prosa*, 2, 47-59, Romansk institutt, Universitetet i Bergen.
- Carter-Thomas, S. (2008). Teaching Coherence through Genre. In *l'écriture dans tous ses états : actes du colloque « De la France au Québec »*. Colloque de Poitiers, IUFM Poitou- Charentes, le 12-15 novembre 2008. En ligne : <http://hal.archives-ouvertes.fr/halshs-00385250>
- Carter-Thomas, S. (2009). *Texte et contexte : pour une approche fonctionnelle et empirique*, Dossier en vue de l'obtention de l'habilitation à diriger des recherches (HDR). Vol. Université Sorbonne Nouvelle – Paris 3.
- Cavalla, C. & Grossmann, G. (2005). Caractéristiques Sémantiques de Quelques 'Noms Scientifiques' Dans L'article de Recherche En Français. In *Akademisk Prosa*, edited by Torodd Kin, 47–59. Bergen: Universitetet i Bergen. En ligne: http://halshs.archives-ouvertes.fr/docs/00/39/42/08/PDF/Cavalla_Grossmann_Kiap2005.pdf.

- Campbell, N. (1975). The personae of scientific discourse. *Quarterly Journal of Speech*, 61, 391-405
- Ceballos-Escalera, J-G. (2012). La combinatoire collocationnelle dans les discours juridique : élément indispensable d'aide à la traduction. *Hal- Acrarchives Ouverts*, La traduction Juridique : Points de vue didactiques et linguistiques, 167-186.
- Crismore, A., Markkanen R., & Steffensen, M. (1993). Metadiscourse in persuasive writing: a study of texts written by American and Finnish university students. *Written Communication*, 10, 39-71.
- Crismore, A. (1989). Talking with Readers: Metadiscourse as Rhetorical Act. *Communication*, 10, 39-71.
- Cruse, David A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.
- Dahl, T. (2004). Textual metadiscourse in research articles: a marker of national culture or of academic discipline. *Journal of Pragmatics*, 36, 1807-1825.
- Delcambre, I. & Reuter, Y. (2002). Images du scripteur et rapports à l'écriture. *Pratiques*, 113-114, 7-28.
- Dionne, B. (1998). *Guide méthodologique pour les études et la recherche* (3^e éd). Montréal : Études vivantes, 190-196.
- Donahue, C. (2002). Effets de l'écrit sur la construction du sujet textuel à l'université. *Spirale*, 29, 75-99
- Drouin, P. (2007). Identification automatique du lexique scientifique transdisciplinaire. *Revue française de linguistique appliquée*, 12(2), 45-64.
- Flaux, N. & Van De Velde, D. (2002). *Les noms en français*. Paris : Ophrys
- Fløttum, K., Dahl, T. & Kinn, T. (2006). *Academic Voices. Across languages and disciplines*. Amsterdam/Philadelphia : John Benjamins.
- Fløttum, K. & Vold, E. T. (2010). L'éthos auto-attribué d'auteurs-doctorants dans le discours scientifique. *Lidil*, (41), 41-58.
- Fløttum, K. (2007). *Language and Discipline Perspectives on Academic Discourse*. Cambridge: Cambridge Scholars Publishing.
- Ghiglione, R. & Trognon, A. (1993). Ou va la Pragmatique ? De la Pragmatique à la Psychologie Sociale. *Presses Universitaires de Grenoble*.

- Gross, A. G. (1996). *The rhetoric of science*. Harvard University Press.
- Gross, A. G. (2006). *Starring the Text: The Place of Rhetoric in Science Studies*. Carbondale: Southern Illinois University Press.
- Grossmann, F. & Tutin, A. (2014). **L'écrit scientifique : du lexique au discours. Autour de Scientext**. Presses de l'Université de Rennes.
- Grossmann, F. & Wirth, F. (2007). Marking Evidentiality in Scientific Papers: The Case of Expectation Markers. In Fløttum K., *Language and Discipline Perspectives on Academic Discourse*. Cambridge: Cambridge Scholars Publishing, 2002-2018.
- Grossmann, F., Tutin, A, Falaise, & A, Kraif, O (2009). Autour du projet Scientext : étude des marques linguistiques du positionnement de l'auteur dans les écrits scientifiques. *Journées Linguistique de Corpus*. Lorient
- Hausmann, F. J. (1989). Le dictionnaire de collocations. In Hausmann F.J., Reichmann O., Wiegand H.E., Zgusta L. (eds), *Wörterbücher : ein internationales Handbuch zur Lexicographie. Dictionaries. Dictionnaires*. Berlin. New-York: De Gruyter. 1010-1019.
- Hausmann, F-J. & Blumenthal, P. (2006). Présentation : collocations, corpus, dictionnaires. *Langue française*, 150, 3-13.
- Halliday, M.A.K. (1994). *An Introduction to Functional Grammar* (2nd ed.). Londre: Edward Arnold. Cité in Hyland, K. (2005).
- Heid, U. (1992). « Décrire les collocations : deux approches lexicographiques et leur application dans un outil informatisé », Colloque international, Phraséologie et terminologie en traduction et en interprétation, *Terminologie et traduction*, 2/3, 523-548.
- Heslot, J. (1980). La formation des chercheurs à l'expression scientifique écrite. *Langage et Société*, supplément au n° 12, Paris, Maison des Sciences de l'Homme, 35-40.
- Hyland, K. (2002). Authority and invisibility: authorial identity in academic writing, *Journal of pragmatics*, 34, 1091-1112.
- Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, 25(2), 156-177.

- Hyland, K. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language writing*, 13, 133-151.
- Hyland, K. (2007). *Metadiscourse. Exploring Interaction in Writing*. London: Continuum.
- Ivanič, R. (1998). Writing and identity. *The discursal construction of identity in academic writing*. Amsterdam/Philadelphia : John Benjamins.
- Jakobson, R. (1980). *Langage enfantin et aphasie*. Paris : Flammarion
- Kerbrat-Orecchioni, C. (1980). *L'Énonciation. De la subjectivité dans le langage*. Paris : Armand Collin.
- Latour, B. & Fabbri, P. (1977, février). La rhétorique de la science [pouvoir et devoir dans un article de science exacte]. In *L'économie des biens symbolique : actes de la recherche en sciences sociales*. Vol. 13(81-95).
- Larreya, (2004). L'expression de la modalité en français et en anglais (domaine verbal) In *Revue belge de philologie et d'histoire*, 82 (3). *Langues et littératures modernes Moderne taal en litterkunde* (pp.733-762).
- Le, E (2004). Active participation within written argumentation: Metadiscourse and editorialist's authority. *Journal of Pragmatics* 36(4), 687-714.
- Le Guern, M. (2003). *Les deux logiques du langage*. Paris : Champion.
- Maingueneau, D. (1981). *Approche de l'énonciation linguistique française*. Classique Hachette : Paris.
- Maingueneau, D. (1998). *Analyser les textes de communication*. Paris: Dunod.
- Mauranen, A. (1992). *Cultural Differences in Accademic Rhetoric. A textlinguistic study*. Unpublished PhD thesis. Birmingham: Université de Birmingham
- Mauranen, A. (1993a). Contrastive ESP Rhetoric: Metatext in Finnish-English economics texts. *English for Specific purposes*, 12, 3-22.
- Mauranen, A. (1993b). *Cultural Differences in Academic Rhetoric: A Textlinguistic Study*. Frankfort am Main: PeterLang Publisher.
- Mel'čuk, I. (1998). Collocations and Lexical Functions. In A. Cowie (Ed.), *Phraseology. Theory, Analysis and Applications*, 23-53. Oxford: Clarendon Press.

- Mel'čuk, I. (1999). *Dictionnaire explicatif et combinatoire du français contemporain*. Canada: Les presses de l'Université de Montréal.
- Moreno, A.I. (1998). The explicit signalling of premise-conclusion sequences in research articles: a contrastive framework. *Text* 18, 544-585.
- Myers, G. (1990). *Writing Biology: Texts in the Social Construction of Scientific Knowledge*. Madison: University of Wisconsin Press.
- Née, E., Sitri, F. & Veniard, M. (2014). Pour une approche des routines discursives dans les écrits professionnels. *Actes du 4ème Congrès Mondial de Linguistique Française* (en cours de publication).
- Nucheze, V. (1998). Approche pragmatico-énonciative du discours de recherche (à l'usage des apprentis-chercheurs). *Lidil*, 17, 38-47
- Orliac B. (2004). *Automatisation du repérage et de l'encodage des collocations en langue de spécialité*. Thèse de doctorat, l'Université de Montréal, Montréal.
- Pecman, M. (2004). *Phraséologie contrastive anglais-français*. Thèse de doctorat, Université de Nice-Sophia Antipolis, Nice.
- Pecman. M. (2007). Approche onomasiologique de la langue scientifique générale. *Revue Française de la Linguistique Appliquée*, 7 (2), 79-96.
- Piot, P., Mann, J-M., Chin, J. & Quinn T. (1988). The International Epidemiology of AIDS. *Scientific American*, 259(4), 82-9.
- Peterlin, A.P. (2005). Text-organising metatext in research articles: an English-Slovene contrastive analysis. *English for Specific Purposes*, 24, 307-319.
- Phal, A. (1971). *Vocabulaire général d'orientation scientifique (VGOS) : part du lexique commun dans l'expression scientifique*. Paris : CREDIF.
- Poudat, C. (2006). *Etude contrastive de l'article scientifique de revue linguistique dans une perspective d'analyse des genres*. Thèse de doctorat, Université d'Orléans.
- Prelli, L. J. (1989). *A Rhetoric of Science : Inventing Scientific Discourse*. University of South Carolina Press.
- Rastier, F. & Denise, M. (2001). Genres et variations morphosyntaxiques. *Traitement Automatique des Langues*, 42(2), 547-577

- Rastier, F. & Pincemin, B. (1999). Sémantique de l'intertexte. *Cahiers de Praxématique*, Ioannis K ANELLOS (dir.), 33, 83-111.
- Reuter, Y. (2004). Analyser les problèmes de l'écriture de recherche en formation. *Pratiques*, (121/122), 9-27.
- Rinck, F. (2006). *L'article de recherche en Sciences du Langage et en Lettres, Figure de l'auteur et approche disciplinaire du genre*. Thèse de doctorat en Sciences du Langage, sous la direction de F. Boch et F. Grossmann, Université de Grenoble
- Rinck, F. (2006). Écrire au nom de la science et de sa discipline : les figures de l'auteur dans l'article en sciences humaines. *Sciences de la Société*, 67, 95
- Rinck, F. (2010 décembre). L'analyse linguistique des enjeux de connaissance dans le discours scientifique. *Revue d'anthropologie des connaissances*, 4 (3) , 427–450. doi:10.3917/rac.011.0427.
- Rinck, F. (2013 février). Pratiques d'écriture de l'article de recherche en sciences humaines : une approche en corpus. *Séminaires des Aspects Concerts de la thèse- EHESS*.
- Risselada, R. & Spooren, W. (1998). Introduction: Discourse Markers and Coherence Relations. *Journal of Pragmatics* 30 (2), 131-133.
- Reutner, Y. (2004). Analyser les problèmes de l'écriture de recherche en formation. *Pratiques*, 121/122, 9-27.
- Sándor, A. (2007). Modeling metadiscourse conveying the author's rhetorical strategy in biomedical research abstracts. *Revue Française de Linguistique Appliquée* 200(2) ,97-109.
- Sándor, Á., Kaplan, A. & Rondeau, G. (2006). Discourse and citation analysis with concept-matching. International Symposium : Discourse and document (ISDD). Caen, France. En ligne : http://www.unicaen.fr/services/puc/article.php3?id_article=686.
- Sándor, Á. & Vorndran, A. (2009). Detecting key sentences for automatic assistance in peer reviewing research articles in educational sciences. Aus: Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries, ACL-IJCNLP 2009, Suntec, Singapore, 7 August 2009 Singapore (2009), S. 36-44.

- Sándor, Á., De Liddo, A., & Buckingham Shum, S. (2012, in Press). Contested collective intelligence: Rationale, technologies, and a human-machine annotation study. *Computer Supported Cooperative Work*, 12(4-5), 417-448. Retrieved 12 June, 2012. En Ligne: <http://oro.open.ac.uk/31052>
- Siepmann, D. (2005). *Discourse markers across languages: a contrastive study of secondlevel discourse markers in native and non-native text with implications for general and pedagogic lexicography*. London and New York: Routledge.
- Siepmann, D. (2007). Les marques de discours polylexicaux en français. *Revue Française de Linguistique Appliquée*, 2, 123-136.
- Silverstein, M (1976). Hierarchy of Features and Ergativity, in R. M. W. Dixon (ed.), *Grammatical Categories in Australian Languages*, Australian Institute of Aboriginal Studies, Canberra, pp. 112–171.
- Sinclair, J. M (1996). *The search for units of meanings*. Textus: English Studies in Italy, 9, 75-106.
- Swales, J.M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Thibaudeau, V. (1997). *Logique et expression de la pensée*. Montréal: Gaëtan Morin.
- Teufel, S. (1999). *Argumentative Zoning: Information Extraction from Scientific Articles*. Thèse de doctorat, Université d'Edinburgh.
- Teufel, S. & Moens, M. (2002). Summarizing Scientific Articles -- Experiments with Relevance and Rhetorical Status. *In Computational Linguistics*, 28 (4), Dec. 2002.
- Tremlay, R-R. & Perrier, Y. (2006). *Savoir plus : outils et méthodes de travail intellectuel* (2ed). Les éditions de la Cheneliere inc.
- Toumi, N. (2009). A model for the investigation of reflexive metadiscourse in research articles. *Language Studies Working Papers*, 1, 64-73
- Tutin, A. (2007, a). Autour du lexique et de la phraséologie des écrits scientifiques. *Revue française de la linguistique appliquée*, XII/ 2, 5-13.
- Tutin, A. (2007, b). Modélisation linguistique et annotation des collocations : application au lexique transdisciplinaire des écrits scientifiques. In S. Koeva, D. Maurel, M.

- Silberztein (Eds). *Formaliser les langues avec l'ordinateur*. Besançon : Presses universitaires de Franche-Comté.
- Tutin, A. (2007c). Traitement sémantique par analyse distributionnelle des noms transdisciplinaires des écrits scientifiques, *TALN*, Toulouse, 5-8 juin 2007.
- 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).Vol. Université de Grenoble 3 – Stendhal.
- Tutin, A. & Grossmann, F. (2002). Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif. *Revue Française de Linguistique Appliquée, Lexique : recherches actuelles*, vol VII (7-25).
- Vande Kopple, W. J. (1985). Some explanatory discourse on metadiscourse. *College Composition and Communication*, 36, 82–93.
- Valero Garces, C. (1996). Contrastive ESP rhetoric: metatext in Spanish-English economics texts. *English for Specific Purposes*, 5,279-294.
- Williams, G, 2003. *Les collocations et l'école contextualiste britannique*, dans Grossmann (F.) et Tutin (A.) éd., *Les collocations : analyse et traitement*, coll. Travaux de recherches en linguistique appliquée. Paris : Éditions De Werelt.

ANNEXES

ANNEXE 1.

ANNEXE 1.1. EXEMPLES D'OCCURRENCES DE NOMS CONSIDERES COMME OBJETS FABRIQUES PHYSIQUE ET NON PHYSIQUE DONNENT LIEU AUX CONCEPTS DE RESUME ET DE CHANGEMENT COGNITIF

Contexte gauche	Occurrence	Contexte droit
théoriques qui ont permis de construire l'objet de	cette recherche	, mais il s'agit de choisir la méthode
, l'objet de recherche sur lequel nous proposons	ce texte	concerne la question de la signification du terme "
Les résultats de	Cette recherche	trouve en partie son origine dans le questionnement induit
	ces recherches	révèlent des différences d'un maître à l'autre
	Ce travail	permet aux enfants de questionner le sens global de
incite à nous interroger, dans le cadre de	Cet article	présente l'expérimentation d'un nouveau dispositif de FOADA
	cet article	sur la stratégie de recrutement des employeurs.
Nous nous sommes intéressées dans le cadre de	cet article	à la stratégie de recrutement des employeurs marquant ainsi
	Cette recherche	viserait l'étude des réactions

Dans	Ces travaux ce texte	émotives et affectives au sont menés en lien avec un programme de recherche , l'auteur s'intéresse à l'influence de
Pour clarifier les enjeux de	ce travail	, nous nous référons aux trois grandes sources formelles
Dans le cadre de	cet article	, nous avons sélectionné trois moments extraits de trois
Dans le cadre de	cet article	, nous proposons de cadrer l'analyse autour des
Dans	cette étude Cette recherche	, nous nous inspirons de l'approche de ' a été réduite et s'est essentiellement appuyée sur
Dans la phase exploratoire , dont nous présentons dans	cet article Cette étude	, nous allons analyser les programmes de cette filière ne fournit évidemment qu'une vue très partielle sur
	Ce travail	s'inscrit dans le projet " Tab " s'
présentons les résultats obtenus en s'intéressant plus dans	cet article	à la phase exploratoire .
Dans	cette étude	, nous nous sommes intéressés aux élèves en Terminale
	Cet article	tente de donner quelques

		éléments de réponses concernant les
Dans	cet article	, nous nous intéresserons à l'évolution de l'
	Cet article	interroge l'évolution de l'activité sur une plateforme
Dans	cette étude	de cas , nous avons mis en évidence différents
Avant de présenter les résultats de	cette recherche	, il est nécessaire de préciser la notion d'
	Ce travail	aide à mettre en regard des pratiques issues du
Nous décrivons dans	ce travail	l'élaboration et la validation de la version abrégée
Nous tenterons également, dans	ce travail	, de participer à la réalisation des deux objectifs
Dans	cette étude	, il ressort que la version abrégée de l'
L'une des implications de	cette étude	est de faire prendre conscience aux enseignants de la
	Ce travail	a été réalisé pour l'ensemble des formations de
A l'issue de	ce travail	sur le métier (3 domaines , 14 activités
	Cet article	propose que le modèle de compétences soit promu comme
	Cet article	propose que le modèle de

		compétences soit promu comme
Cependant,	ces études	tendent à déconsidérer la décision en tant qu'événement
	Cet article	porte donc sur la question du lien entre engagement
Les résultats de	cette recherche	nous conduisent à envisager le processus de décision en
. L'objectif de	cet article	est de contribuer à une meilleure compréhension de l'
	Cette recherche	, de nature exploratoire, est fondée sur une
. Nous nous sommes centrées dans le cadre de	cette recherche	, sur un contenu : les motifs d'agir
	Cette recherche	vise par une démarche compréhensive à rendre intelligibles les
Les enseignants interrogés lors de	cette recherche	ont tous une bonne connaissance des TICE et notamment
	Cette recherche	met en avant les enjeux inscrits dans les pratiques
L'hypothèse posée dans	cette recherche	est que cette notion repose sur une évidence non
contribution présente les résultats théoriques et conceptuels obtenus dans	cette recherche	de quelques années et les soumet à l'examen
Le corpus constitué pour	cette recherche	consiste en un ensemble de textes recueillis en philosophie

L'hypothèse de	cette recherche	est que cette analyse peut apporter un éclairage nouveau
(Mayen , 2004) dans lequel nous inscrivons	cette recherche	.
: dans le cadre d'un projet FSE ,	cette recherche	action s'appuie sur le travail de 2 chercheurs
du rapport à la communauté décrites au début de	cet article	, on remarque que les membres évoluent chacun à
Au début de	cet article	, nous posons l'hypothèse que pour chaque membre
l'argumentation et des dissonances entre les idées ,	cet article	propose de s'interroger sur le processus de socialisation
Le cadre théorique de	ce travail	s'inscrit donc dans un premier temps dans un
Paradoxalement ,	cette étude	se rapportant à l'oral utilisera pour méthodologie celle
choix culturels , justifiant ainsi la quête historique de	ce travail	de recherche visant le couple oralité et éducation ,
	Ces études	reposent sur un travail d'équipe pour préparer ,
Pour réaliser	ce travail	de conception la séance va se dérouler dans l'
Brandt-Pomares & amp ; Ranucci , 2005) ,	cette recherche	se propose d'étudier la question de l'enseignement
	Ce travail	se poursuit et ce , dans l'optique de
	Ces travaux	s'intéresseront de façon

		conjointe au repérage des processus
	Ces travaux	tenteront d'articuler à la fois des investigations théoriques
	Cet article	présente le contenu du programme , le cadre et
A travers	cette recherche	, nous avons mis en évidence plusieurs aspects touchant
auxquels nous nous sommes intéressés dans le cadre de	cette recherche	enseignement à l'ordre collégial .
	Cette recherche	menée auprès d'élèves ingénieurs à l'INT montre
	Cette recherche	s'inscrit dans le champ de l'intersubjectivité ,
Ce qui nous intéresse dans	ce travail	est la manière dont cette figure sociale est reçue
	Cette recherche	m'a amené à considérer les processus d'étiquetage
	Cet article	est issu d'une recherche sur les Pratiques Collectives
	Cette recherche	fait suite à l'appel à propositions " usages
Les résultats que nous avons présentés dans	cet article	ne sont qu'une facette d'un phénomène ,
Pour apprendre	cette recherche	d'informations -là il faut faire entrer ces pratiques
	Cette étude	s'inscrit dans une recherche plus large incluant des
compétences visées au	cet article	plus particulièrement les

B2i 1 , nous explorons dans		conditions et contraintes de l'enseignement
	Cet article	s'inscrit dans une recherche qui a pour objectif
En résumé ,	cette recherche	tente de dégager les critères qui font passer un
ou la colonne vertébrale de la mémoire autobiographique ,	cette recherche	proposa d'évaluer l'impact des événements dans la
	Cette recherche	autobiographique exploratoire a utilisé une méthodologie simple et rapide
pas que ce qui nous attirait , à travers	cette recherche	, n'était pas que les personnes âgées se
nous empêcher de les associer aux résultats principaux de	cette recherche	que nous ne souhaitions pas développer dans le cadre
	Cette recherche	porte sur l'observation du travail initiatique effectué au
	Ce travail	observe les pratiques de cette inter- unité .
	Cette recherche	étudie les modalités de l'acquisition de ce nouveau
A partir de là , l'objectif de	cet article	est de proposer un bilan des relations entre enseignement
Dans le contexte de	cette étude	, il semble influencer tout autant sur le renforcement
	Ces travaux	sont menés en lien avec un programme de recherche
a pas été développée dans	cet article	.

le cadre limité de		
Dans	ce texte	, l'auteur s'intéresse à l'influence de
Au cours de	cette recherche	, nous avons tenté de caractériser le savoir analyser
Le cadre conceptuel de	cette recherche	s'appuie sur le champ de la didactique clinique
C'est donc dans le cadre de	cette recherche	portant sur le " savoir analyser " que nous
	Cette étude	de cas rend compte à sa manière, des
À travers	cette étude	de cas, nous avons donc pu percevoir ce
	Cette recherche	pancanadienne, commencée en 1994, vise à constituer
Pour connaître l'ensemble des résultats de	cette étude	, le lecteur est invité à consulter d'autres
À partir de trois sources d'information,	cette étude	vise à rendre compte de l'état du développement

ANNEXE 1.2. EXEMPLES DE VERBES DE POINT DE VUE QUI SE TROUVENT EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS DU LEXIQUE SCIENTIFIQUE

Contexte gauche	Occurrence	Contexte droit
Nous nous contenterons ici de	souligner	qu'elle va exactement à l'encontre de l'idée
Ainsi avons –nous	noté	ci-dessus l'existence d'une dis-chronie , l'initiative
Nous avons	noté	aussi les limites du point de vue de l'
	Rappelons	que nous cherchons à montrer que les professeurs de
Il est intéressant de	noter	que la question de la problématique se pose pour
Nous avons	rappelé	ci-dessus que notre recherche s'inscrit dans un contexte
Nous	rappelons	que la question posée était : " Pourriez –vous
Il convient de	souligner	la contradiction entre les attitudes en réception et en
On peut donc	noter	plusieurs différences significatives entre ces deux groupes : la
Ce constat tend à	souligner	les difficultés inhérentes aux situations de plurilinguisme et incite
Notre questionnement est parti d'un constat	souligné	notamment par les rapports Joutard [JOUTARD 01]

Or, il est intéressant de	noter	que c'est justement les systèmes éducatifs nordiques ,
Comme le	rappelle	Jacques Lecomte (2004) dans un article sur
Il est intéressant de	boter	, à ce propos, que cette appréciation relativement
En effet, comme le	souligne	(Bouguin, 2003) , " la puissance
Il est intéressant de	noter	aussi que la boîte à crier est elle -
	rappelons	aussi que la problématique de l'étiquetage de l'
il est présent, le scolaire est beaucoup moins	rappelé	.
Un autre fait intéressant mérite d'être	souligné	.
événements ponctuel 3 et a minima stressant semblent donc	rappelés	.
Nous avons précédemment	souligné	(Lani-Bayle & amp ; Mallet , 2006)
2 Nous avons précédemment	souligné	(Lani-Bayle & amp ; Mallet , 2006)
Pour parvenir à ces objectifs , il faut	rappeler	que l'âge est un important constituant de l'
	Notons	que cette dimension n'est pas sans rappeler la
Nous tenons à	souligner	l'importance de l'écriture professionnelle en tant qu'

ANNEXE 1.3. EXEMPLES DE VERBES DU METADISOURS INTERACTIONNEL EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS DU LEXIQUE SCIENTIFIQUE.

Contexte gauche	Occurrence	Contexte droit
Nous nous contenterons ici de	souligner	qu'elle va exactement à l'encontre de l'idée
Ainsi avons –nous	noté	ci-dessus l'existence d'une dis-chronie , l'initiative
Nous avons	noté	aussi les limites du point de vue de l'
	Rappelons	que nous cherchons à montrer que les professeurs de
Il est intéressant de	noter	que la question de la problématique se pose pour
Nous avons	rappelé	ci-dessus que notre recherche s'inscrit dans un contexte
Nous	rappelons	que la question posée était : " Pourriez –vous
Il convient de	souligner	la contradiction entre les attitudes en réception et en
On peut donc	noter	plusieurs différences significatives entre ces deux groupes : la
Ce constat tend à	souligner	les difficultés inhérentes aux situations de plurilinguisme et incite
Notre questionnement est parti d'un constat	souligné	notamment par les rapports Joutard [JOUTARD 01]

Or, il est intéressant de	noter	que c'est justement les systèmes éducatifs nordiques ,
Comme le	rappelle	Jacques Lecomte (2004) dans un article sur
Il est intéressant de	boter	, à ce propos, que cette appréciation relativement
En effet , comme le	souligne	(Bouguin , 2003) , " la puissance
Il est intéressant de	noter	aussi que la boîte à crier est elle -
	rappelons	aussi que la problématique de l'étiquetage de l'
il est présent , le scolaire est beaucoup moins	rappelé	.
Un autre fait intéressant mérite d'être	souligné	.
événements ponctuel 3 et a minima stressant semblent donc	rappelés	.
Nous avons précédemment	souligné	(Lani-Bayle & amp ; Mallet , 2006)
2 Nous avons précédemment	souligné	(Lani-Bayle & amp ; Mallet , 2006)
Pour parvenir à ces objectifs , il faut	rappeler	que l'âge est un important constituant de l'
	Notons	que cette dimension n'est pas sans rappeler la
Nous tenons à	souligner	l'importance de l'écriture professionnelle en tant qu'

ANNEXE 1.4. OCCURRENCE DES VERBES D'APPORTS SCIENTIFIQUE EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS LEXICAUX DU LEXIQUE SCIENTIFIQUE

Contexte gauche	Occurrence	Contexte droit
Cette contribution introductive au symposium	proposé	sur le thème " Dispositifs de formation numériques et
Nous	proposons	les résultats d'une recherche relative à la question
La démarche que je	propose	repose notamment sur la définition donnée par Verón de
ajustement des signes , et dont l'approche sémiotique	propose	de rendre compte , se jouent des phénomènes d'
Nous nous	proposons	donc de considérer maintenant les points de ressemblances entre
Les réponses apportées , diffèrent considérablement et	proposent	une gamme d'interprétations très large .
Pourtant une des perspectives de recherche	proposée	comme suite à donner proposait d'explorer les répercussions
des perspectives de recherche proposée comme suite à donner	proposait	d'explorer les répercussions du travail en équipecycle sur
Cette contribution se	propose	de rendre compte d'une recherche doctorale en cours
présentés dans cette contribution ne permettent certes pas de	conclure	à l'existence d'effets d'une raison numérique
Dans le cadre de cet article , nous	proposons	de cadrer l'analyse autour des décisions de justice
La problématique	proposée	ci-dessus est contextualisée à la formation professionnelle universitaire dont
Cette communication	propose	de considérer l'arrivée des femmes dans l'enseignement
Dans un premier temps , il s'agira de	dégager	la place des langues dans les programmes féminins et
Dans un premier temps , il s'agira	dégager	la place qu'occupent les langues

de		dans les programmes
2000) ou canadiennes de Mayhew (2000)	démontrent	pourtant que le sur risque des conducteurs novices décroît
Nous nous	proposons	de mettre en évidence de quelle manière les tâches
Nous	proposons	que ces modes peuvent être mis en évidence par
Cet article	propose	que le modèle de compétences soit promu comme objectif
Dans un second temps , nous avons	proposé	aux intéressés de commenter la lecture soit plus d'
Ces résultats	démontrent	donc le parallélisme entre les deux échelles d'évaluation
Les propos échangés	démontrent	que les réunions inter degrés , annoncées comme lieux
En résumé , cette recherche tente de	dégager	les critères qui font passer un événement de minuscule
Pour	conclure	, nous pourrions donc méditer la citation suivante qui
de cette communication, nous nous attacherons également à	démontrer	que la prise en compte de la majorité des

Annexe 1.5. Occurrence du verbe tenter en relation syntaxique avec d'autres éléments lexicaux du lexique scientifique.

Contexte gauche	Occurrence	Contexte droit
Notre démarche se veut semi inductive 2 mais	tente	également de laisser l'horizon de la totalisation "
Cet article	tente	de donner quelques éléments de réponses concernant les usages
Nous	tenterons	également, dans ce travail, de participer à
recherches en sociologie des usages des médias qui ont	tenté	de modaliser cette relation, ont dans un premier
C'est pourquoi nous	tenterons	d'aborder la question de la place de l'
que les travaux de Françoise Waquet , que nous	tenterons	de retourner à la source de la culture scolaire
de compréhension que nous nous situons et que nous	tenterons	d'analyser ce paradoxe.
Ces travaux	tenteront	d'articuler à la fois des investigations théoriques relevant
Pour	tenter	de répondre à cette question, nous avons mené
Résumé : Pour	tenter	d'expliquer pourquoi les femmes ingénieurs accèdent moins souvent
Nous commençons par	tenter	de faire exprimer leurs besoins aux concepteurs et aux
En résumé, cette recherche	tente	de dégager les critères qui font passer un événement
Au cours de cette recherche, nous avons	tenté	de caractériser le savoir analyser chez une vingtaine de
Nous avons	tenté	de montrer un cas précis

où un enseignant se

ANNEXE 1.6. OCCURRENCE DU VERBE METTRE EN EVIDENCE EN RELATION SYNTAXIQUE AVEC D'AUTRES ELEMENTS LEXICAUX DU LEXIQUE SCIENTIFIQUE.

Contexte gauche	Occurrence	Contexte droit
Les résultats de la recherche ont notamment	mis en évidence	
, nous cherchons à	mettre en évidence	les différents rôles qu'ils endossent, ainsi que
Il est intéressant de	mettre en évidence	les verbes qu'il utilise à ce moment.
Nous nous proposons de	mettre en évidence	de quelle manière les tâches et les rôles sont
Dans cette étude de cas, nous avons	mis en évidence	différents modes d'organisation de l'activité commune entre
Nous proposons que ces modes peuvent être	mis en évidence	par l'analyse des interactions dans lesquelles est engagé
Cette approche tend ainsi à	mettre en évidence	des constructions d'usages différentes révélées par les significations
) ..., approches à partir desquelles on peut	mettre en évidence	un certain nombre de points essentiels : - L'
brève analyse que nous avons menée a permis de	mettre en évidence	et d'illustrer ces phases ainsi que les processus
Cette constatation	met en évidence	une des difficultés à prévoir au moment de généraliser
A travers cette recherche, nous avons	mis en évidence	plusieurs aspects touchant aux attentes et à l'évaluation

, nous avons pu	mettre en évidence	la relative proximité entre ces deux démarches évaluatives
L'objectif du paragraphe suivant, est de	mettre en évidence	empiriquement, l'hétérogénéité et la segmentation du marché
, de manière globale, puis nous tenterons de	mettre en évidence	les divers segments du marché du travail.
l'analyse des données, il s'agira de	mettre en évidence	la liaison formation / emploi en soulignant le rôle
du travail de recherche doctorale ; elle permet de	mettre en évidence	les zones de turbulence et de bafouillement dans lesquelles
de l'intérêt d'un ou plusieurs textes à	mettre en évidence	, de l'étayer en mettant en résonance divers
l'ensemble de cet entretien nous ont permis de	mettre en évidence	plusieurs points.

Annexe 1.7. Occurrence de noms construits et complexes, et ceux produit par l'activité scientifique.

Contexte gauche	Occurrence	Contexte droit
Les	résultats	doivent permettre de valider le modèle de compétence et
Comme on le voit, l'interprétation statistique des vérifications des hypothèses élaborées à partir de ces premiers	résultats	soulève elle aussi de nombreuses interrogations.
urgent de s'intéresser à ces questions que les	résultats	obtenus aux tests pourraient avoir une influence sur le
Nous proposons les	résultats	d'une recherche relative à la question de la
A partir des	résultats	de l'enquête, nous nous interrogerons ici sur
On retrouve ici l'	Idée	que l'éducation ne se limite pas à fournir
Les	résultats	de l'enquête confirment le rôle déterminant du diplôme
Cependant, plusieurs hypothèses peuvent expliquer un tel	résultat	.
L'ensemble de ces	résultats	vise à former la base d'une recherche qui
est la base du présent texte qui présente les	résultats	d'une analyse des données de cette recherche en
Les	résultats	de la recherche ont notamment mis en évidence :

Les	résultats	de cette première recherche nous apportent des indications importantes
Ces observations nous ont permis d'avoir une	Idée	sur les compétences des élèves à utiliser un tableur
Nous empruntons ces	idées	, qui ont été discutées par de nombreux auteurs
Ainsi, nous pouvons montrer que les deux	systèmes	ne sont pas aussi éloignés l'un de l'autre
Nous partageons l'	Idée	de nos collègues de l'ADMEE, qu'une

ANNEXE 2 : LISTE DE LEXIQUE

apparaître : [PUBLICATION].
apporter : [BUT].
apporter : [OP_MENTALE, PORTEE].
apport : [OP_MENTALE, PORTEE].
assumer : [OP_MENTALE].
auteur : [PUBLICATION].
cadre : [PORTEE ,nogrow].
ce : [DEICTIQUE].
chercher : [OP_MENTALE].
cibler : [OP_MENTALE].
éclaircir : [OP_MENTALE, CONTRAST].
éclairer : [OP_MENTALE].
clair : [OP_MENTALE].
clairement : [OP_MENTALE].
clarification : [OP_MENTALE, CONTRAST].
clarifier : [OP_MENTALE, CONTRAST].
classer : [PORTEE].
clé : [IMPORTANCE].
comparer : [CONTRAST].
confirmer : [PORTEE, posit].
démenti : [CONTRAST].
démentir : [CONTRAST].
démêler : [CONTRAST, OP_MENTALE].
démonstration : [PORTEE].
démontrer : [PORTEE].
devoir : [CONTRAST].
documentation : [PUBLICATION].
documentation : [PUBLICATION, PORTEE].
dogme : [PORTEE].
domaine : [PORTEE].
doute : [CONTRAST].

discuter : [CONTRAST].

distinguer : [CONTRAST].

déterminer : [OP_MENTALE].

émettre : [OP_MENTALE].

engager : [CONTRAST].

élément : [PORTEE].

ensemble : [PUBLICATION].

estimer: [OP_MENTALE].

entamer : [CONTRAST].

équipe : [PUBLICATION].

étonnamment : [attitude].

étudier: [PUBLICATION].

évaluer: [PORTEE].

évaluation: [PORTEE].

évidence: [PORTEE].

examen: [OP_MENTALE].

expérimentation: [PORTEE].

explication : [OP_MENTALE].

explication : [OP_MENTALE, PORTEE].

exposé: [PUBLICATION].

extrêmement: [attitude, nonot].

globalement: [PUBLICATION].

grand : [IMPORTANCE].

grand : [IMPORTANCE].

guère : [negative, CONTRAST].

heure : [timefact ,method].

identifier : [OP_MENTALE, PORTEE].

illustrer : [OP_MENTALE].

impossibilité : [POSSIBILITE, CONTRAST].

importance : [IMPORTANCE].

importance : [IMPORTANCE, PORTEE].

important : [IMPORTANCE].

important : [IMPORTANCE].

improbable : [CONSTRAST, negative].
indice: [PORTEE].
infirmier : [CONTRAST].
insuffisamment : [CONSTRAST].
irréfutable : [OP_MENTALE, CONSTRAST].

interroger : [CONSTRAST].

interrogation : [CONSTRAST].
manifestement : [PORTEE].
montrer : [PORTEE].
méthode : [PORTEE].
notion : [OP_MENTALE, PORTEE].
notre : [DEICTIQUE, possession].
nôtre : [DEICTIQUE, possession].
nous : [DEICTIQUE].
on : [DEICTIQUE, Ref_PUBLICATION].

opposer : [CONSTRAST].
nous : [DEICTIQUE, PORTEE].
observer : [PORTEE].
papier : [PUBLICATION].
paradoxalement : [CONSTRAST].
paradoxe : [CONSTRAST].
partie : [PUBLICATION].
penser : [OP_MENTALE].
penser : [OP_MENTALE].
pensée : [PORTEE].
pouvoir : [POSSIBILITE].
permettre : [POSSIBILITE].
perspective : [OP_MENTALE].
pertinence : [IMPORTANCE].
pertinent : [IMPORTANCE].
preuve : [PORTEE].
principe : [PORTEE].
principe : [PORTEE].

probablement : [POSSIBILITE].
problématique : [CONSTRAST].
prouver : [PORTEE].
question : [CONSTRAST, OP_MENTALE].
questionner : [CONSTRAST, OP_MENTALE].
raisonnement : [OP_MENTALE].
reget [CONSTRAST].
rejeter : [CONSTRAST].
rejet : [CONSTRAST].
réflexion : [OP_MENTALE, PORTEE].
résoudre : [change, CONSTRAST].
résoudre : [OP_MENTALE].
résulter : [OP_MENTALE, PORTEE].
suggestif : [OP_MENTALE].
suggestion : [OP_MENTALE].
suggérer : [OP_MENTALE].
supposer : [OP_MENTALE, POSSIBILITE].
supposer : [OP_MENTALE, POSSIBILITE].
supposition : [OP_MENTALE].
supposition : [POSSIBILITE, PORTEE].
suprême : [IMPORTANCE].
tableau : [PUBLICATION].
tester : [OP_MENTALE ,know ,PUBLICATION ,methodologie ,verification].
texte : [PUBLICATION].
théorie : [PORTEE ,widePORTEE ,nogrow].
thèse : [PORTEE ,widePORTEE ,nogrow].
trouvaille : [PORTEE ,PUBLICATION ,nothis].
valider : [PORTEE].
valide : [PORTEE].
validité : [PORTEE].
variable : [PORTEE].
vérifier : [PORTEE,demonstration].
visible : [attitude].

vrai : [OP_MENTALE].
expliquer : [PORTEE].
justifier : [PORTEE].
coupe : [reresent_graph].
courbe: [reresent_graph].
représentaion: [reresent_graph].
figure: [reresent_graph].
graphe: [reresent_graph].
tableau: [reresent_graph].
schéma: [reresent_graph].
analyse: [obj-const-activit_sc].
analyse : [OP_MENTALE, PORTEE].
application : [processus-activit_sc].
choix: [processus-activit_sc].
conception: [processus-activit_sc].
développement: [processus-activit_sc].
étude: [processus-activit_sc].
évaluation: [processus-activit_sc].
projet: [processus-activit_sc].
recherche: [processus-activit_sc].
test: [processus-activit_sc].
traitement: [processus-activit_sc].
travail: [processus-activit_sc].
utilisation : [processus-activit_sc].
application : [obj-const-activit_sc].
approche : [obj-const-activit_sc].
démarche : [obj-const-activit_sc].
définition : [obj-const-activit_sc].
idée : [obj-const-activit_sc].
chiffres: [reresent_graph].
donnée : [obs-activit_sc].
cas : [obs-activit_sc].
exemple : [obs-activit_sc].
observation : [obs-activit_sc].

visualisation : [obs-activit_sc].
facteur : [obs-activit_sc].
paramètre : [obs-activit_sc].
point : [obs-activit_sc].
question : [obs-activit_sc].
problème : [obs-activit_sc].
résultat : [PORTEE, obs-activit_sc].
méthode: [obj-const-activit_sc].
modèle : [obj-const-activit_sc].
solution : [obj-const-activit_sc].
hypothèse : [obj-const-activit_sc, CONTRAST].
système : [obj-const-activit_sc].
technique: [obj-const-activit_sc].
technologie: [obj-const-activit_sc].
théorie: [obj-const-activit_sc].
test: [obj-const-activit_sc].
article: [support-es].
chapitre: [support-es].
document: [support-es].
section: [support-es].
texte: [support-es].
but: [lien-logique].
conséquence: [lien-logique].
lien: [lien-logique].
rapport: [lien-logique].
relation: [lien-logique].
bibliographie: [support-es].
écrit : [support-es].
ceci : [corf].
cela : [corf].
interaction: [OP_MENTALE].
corrélation: [lien-logique].
valeur : [IMPORTANCE].
lien: [lien-logique].

nécessité: [IMPORTANCE].
concourir: [transparent].
tout: [corf].
notamment: [IMPORTANCE].
nécessité : [IMPORTANCE].
interaction: [lien-logique].
expérience : [obs-activit_sc].
possibilité: [POSSIBILITE].
impossibilité: [POSSIBILITE].
apporter: [transparent].
efficacité: [IMPORTANCE].
montage : [obs-activit_sc].
exister: [OP_MENTALE].
existence: [OP_MENTALE].
tendre : [transparent].
tenter: [transparent].
faire : [transparent].
technique : [obs-activit_sc].
imposer: [posit, post].
architecture : [obs-activit_sc].
activité : [obs-activit_sc].
largement : [IMPORTANCE].
donner : [lien-logique].
abondamment : [IMPORTANCE]
signifier : [PORTEE].
situer : [PORTEE].
cependant: [CONTRAST].
néanmoins: [CONTRAST].
mais: [CONTRAST].
examiner: [PORTEE].
constater: [CONTRAST].

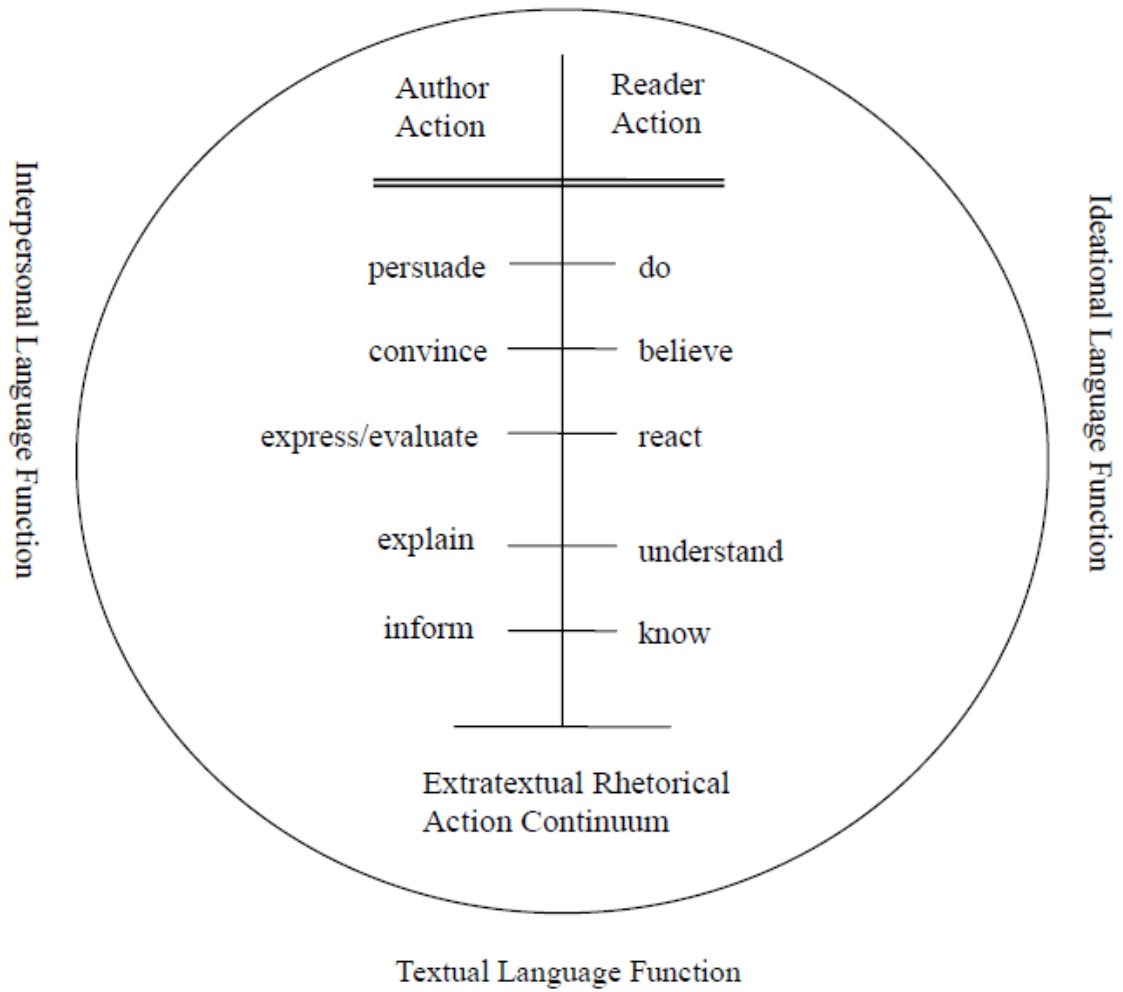
postulat : [OP_Mentale]

ANNEXE 3

ANNEXE 3.1 : W.J. Vande Kopple's Classification System for Metadiscourse (1985,82-92)

CATEGORY	FUNCTION
Textual metadiscourse	
Text connectives	Used to help show how parts of a text are connected to one another. Includes sequencers (first, next, in the second place), reminders (as I mentioned in chapter 2), and topicalizers, which focus attention on the topic of a text segment (with regard to, in connection with).
Code glosses	Used to help readers to grasp the writer's intended meaning. Based on the writer's assessment of the reader's knowledge, these devices reward, explain, define, or clarify the sense of a usage
Validity markers	Used to express the writer's commitment to the probability of or truth of a statement. These include hedges(perhaps, might, may), emphatics(clearly, undoubtedly), and attributers which enhance a position by claiming the support of a credible other(according to Einstein)
Narrators	Used to inform readers of the source of the information presented-who said or wrote something (according to smith, the Prime minister announced that).
Interpersonal metadiscourse	
Illocution markers	Used to make explicit the discourse acts the writer is performing at certain points(to conclude, I hypothesize, to sum up, we predict)
Attitude markers	Used to express the writer's attitudes to the propositional material he or she presents9unfortunately, interestingly, I wish that, how awful that).
Commentaries	Used to address readers directly, drawing them into an implicit dialogue by commenting on the reader's probable mood or possible reaction to the text(you will certainly agree that, you might want the third chapter first).

ANNEXE 3.2 : A.CRISMORE'S MODAL OF METADISCOURSE (1989, 91)



ANNEXE 3.3 : METADISOURSE CATEGORIZATION BY A. CRISMORE ET AL. (1993,47-54)

CATEGORY	FUNCTION	EXAMPLES
Textual metadiscourse		
<i>1. Textual markers</i>		
Logical connectives	Show connection between ideas	Therefore; so; in addition; and
Sequencers	Indicate sequence /ordering of material	First; next; finally; 1,2,3
Reminders	Refer to earlier text material	As we saw in chapter one
Topicalizers	Indicate a shift in topic	Well, now we discuss ...
<i>2. Interpretive markers</i>		
Code glosses	Explain text material	For example; that is
Illocution markers	Name the act performed	To conclude; in sum; I predict
Announcements	Announce upcoming material	In the next section
Interpersonal metadiscourse		
Hedges	Show uncertainty to the truth of assertion	Might; possible; likely
Certainty markers	Express full commitment to assertion	Certainly; know; shows
Attributers	Give source/support of information	Smith claims that ...
Attitude markers	Display writer's affective values	I hope/agree; surprisingly
Commentary	Build relationship with reader	You may not agree that

ANNEXE 3.4 : AN INTERPERSONAL MODEL OF METADISOURSE (K. HYLAND, 2005, 49)

CATEGORY	FUNCTION	EXAMPLES
Interactive	Help to guide the reader through the text	Resources
Transitions	Express relations between main clauses	In addition; but; thus; and
Frame markers	Refer to discourse acts, sequences and stages	Finally; to conclude; my purpose is
Endophoric markers	Refer to information in other parts of the text	Noted above; see figure; in section 2
Evidentials	Refer to information from other texts	According to X; Z states;
Code glosses	Elaborate propositional meaning	namely; e.g.; such as; in other words
Interactional	Involve the reader in the text	Resources
Hedges	Withhold commitment and open dialogue	Might; perhaps; possible; about
Boosters	Emphasize certainty and close dialogue	
Attitude markers	Express writer's attitude to proposition	Unfortunately; I agree; surprisingly
Self-mentions	Explicit reference to authors	I; we; my; me; our
Engagement markers	Explicitly build relationship with reader	Consider; note; you can see that

ANNEXE 3.5 : PERSONAL AND IMPERSONAL CONFIGURATIONS OF METATEXT AND WRITER-READER INTERACTION (A. ADÈL, 2006, 38)

<p>Metadiscourse Metatext Impersonal Text/Code-oriented</p>	<p>Personal Participant-oriented</p>	<p>Writer-oriented</p>	<p>Reader-oriented</p>	<p>Writer-reader interaction Personal Participant-oriented</p>	<p>Reader-oriented</p>
<p>In other words The question is A definition of involve This essay will primarily deal with.....</p>	<p>In our discussion above If we take As an example</p>	<p>My conclusion is that By..... I mean..... As a writer, I would like to argue that.....</p>	<p>AS YOU HAVE SEEN You might want to read the first section first....</p>	<p>I know that you think that... Correct me if i'm wrong....</p>	<p>Now, dear reader, you probably..... Does this sound..... to you ?</p>

ANNEXE 4 : ARTICLE TRANSMIS AUX EXPERTS POUR SURLIGNER LES PHRASES VEHICULANT UN PROBLEME DE RECHERCHE

Usage des TIC, développement et extension des compétences professionnelles chez les enseignants en formation initiale au Département d'Informatique et des Technologies Educatives de l'ENS de l'université de Yaoundé I

Use of ICT, development and extension of professional skills among future teachers in the Department of Computer Science and Educational Technology of Advance High School of Teachers of the university of Yaoundé I

Julia Ndibnu-Messina Ethé

Département des Sciences de l'Éducation, Université de Douala, Douala, Cameroun

Gérard Willy Nya Nouatcha

Département d'Informatique, Université de Yaoundé I, Yaoundé, Cameroun

Résumé

Cet article propose une typologie des usages des TIC par les enseignants en formation initiale à l'École Normale Supérieure (ENS) de l'université de Yaoundé I à partir d'une enquête de terrain reposant sur une méthode essentiellement qualitatif-quantitative appliquée à 38 élèves-professeurs du Département d'Informatique et des Technologies Educatives (DITE). Les entretiens par questionnaire et l'analyse descriptive des contenus catalysent une taxinomie d'emploi des outils TIC par les futurs enseignants du DITE et précisent les implications casuelles de son utilisation par les autres instructeurs, particulièrement, ceux de langues. Il est dès lors possible d'orienter les pédagogues en formation initiale quant au processus d'usage des TIC dans leurs pratiques d'auto-apprentissage et de transmission. En outre, les suggestions finales guideront les enseignants et les responsables des Écoles Normales sur les stratégies d'accompagnement utiles au processus d'insertion pédagogique des TIC dans l'éducation et d'extension à l'usage du web 2.0.

Mots clés : TIC, formation initiale, apprentissage, enseignement, compétences professionnelles, autoformation

Abstract

This paper proposes a typology of uses of ICT by student teachers at the Advance High School of Teachers (AHST) of the University of Yaounde I from a field survey based on an essentially qualitative-quantitative method applied to 38 future teachers of the Department of Computer Science and Educational Technologies (DITE). Interviews by questionnaire and descriptive content analysis catalyze a taxonomy of the use of ICT tools for future teachers called contingencies and specify the implications of its use by other instructors, specifically, language teachers. It is therefore possible to orient the teachers training on the use of ICT in their learning/teaching processes and practices. It aims also to guide teachers and leaders of training schools in the accompanying strategies useful in the process of educational integration ICT in education and extension to the use of web 2.0.

Keywords: ICT, initial training, learning, teaching, professional competences, self-training

Introduction

Depuis 2001, le processus de pénétration des TIC dans le système éducatif camerounais connaît une célérité incomparable au niveau de l'enseignement supérieur à travers les différentes universités et écoles de formation. De fait, le multimédia semble, depuis lors, l'instrument pédagogique le plus employé à l'école. Le cas le plus apparent semble l'École Normale Supérieure de l'Université de Yaoundé I qui déploie de remarquables efforts avec notamment l'ouverture en 2008 de la filière informatique. Elle se consacre à la formation des technopédagogues chevronnés dans le processus d'intégration, d'acquisition et de mise à disposition d'un ensemble de ressources numériques TIC au service de l'enseignement/apprentissage. Toutefois, d'après Salomon Tchameni Ngamo et Thierry Karsenti (2008), une meilleure intégration des TIC ne se polarise pas seulement à la mise à disposition des outils TIC aux enseignants et apprenants, mais nécessite un usage approprié de ceux-ci par les premiers acteurs du syncrétisme, à savoir les instructeurs pendant et après leur formation. Il ne s'agit pas non plus d'étudier, comme Marcelline Tchamabe (2013), l'impact de l'usage des TIC sur les apprenants d'un niveau équivalent au secondaire ou de s'interroger sur les applications réelles des textes de loi quant à l'enseignement de l'informatique ou de son usage pour enseigner, mais de se questionner si l'appropriation des TIC et TICE par ceux qui sont supposés former les élèves du secondaire. Nous souscrivons plutôt aux recherches empiriques effectuées par Karsenti et al (2011) sur l'usage des TIC en pédagogie universitaire. Les TIC apportent aux étudiants canadiens de nouvelles opportunités : en est-il de même pour les étudiants camerounais en formation initiale ? Aussi, paraît-il légitime de s'interroger sur l'ampleur des usages des outils TIC par les futurs enseignants du secondaire formés au DITE et des autres départements de l'ENS comme ceux de Français et de langues et cultures camerounaises. Les TICE sont-elles d'un apport incontesté à tous les futurs pédagogues ? Si aujourd'hui les pays africains ne se limitent plus à former les apprenants à la simple connaissance des outils numériques, mais s'étendent à l'usage du web social pour une autonomisation de l'apprentissage, qu'en est-il du Cameroun ? Le présent article, réalisé dans le contexte singulier de la formation initiale des enseignants de l'ENS de l'université de Yaoundé I, vise à essayer de déchiffrer, documenter et éclairer les perspectives pédagogiques qui se dessinent pour l'avenir des Écoles de formation initiales au Cameroun à travers l'usage des outils numériques et dans une moindre mesure du web 2.0. Après un point sur l'usage des TIC dans l'enseignement, nous disséquons l'accessibilité au multimédia, dresserons une typologie de l'usage des TIC par les futurs enseignants et envisagerons les possibilités de recyclage et d'extension de l'emploi spécifique du web 2.0 aux autres enseignants, spécifiquement, ceux de langues.

Contexte d'usage des TIC et numérisation pédagogique

Conformément à la formulation de Salomon Tchameni Ngamo et Thierry Karsenti (2009), l'expression « intégration pédagogique des TIC » désigne une incorporation des technologies de l'information et de la communication (TIC) pour faciliter le processus d'enseignement et d'apprentissage à l'école. Les recherches y relatives ne se restreignent plus à l'élucidation des politiques d'intégration des outils TIC dans le processus d'enseignement/apprentissage, mais ambitionnent selon certains chercheurs (Baron, Bruillard et Levy, 2000; Monaghan, 2004; Rinaudo, 2001) l'exploration des différents usages institués des outils TIC par les enseignants. L'exploitation des TIC exige

également des différents acteurs du système éducatif un pilotage et une aptitude en informatique, catalyseurs d'une meilleure intégration de ceux-ci pendant les enseignements et vecteurs d'un développement efficace des compétences professionnelles TIC.

De plus, l'usage des TIC dans la transmission des connaissances à l'université semble s'accroître d'une manière exponentielle (Buckley, Pitt, Norton et Owens, 2010). Dans les lycées camerounais, Béché (2012) souligne l'impact positif de l'usage des outils numériques sous surveillance pédagogique sur les acquis et les performances scolaires des lycéens. En outre, il ajoute que « *les résultats du Projet Panafricain d'intégration pédagogique des nouvelles technologies (PanAf) [...] présentent aussi les TIC comme produisant un impact positif sur l'apprentissage* ». En corroborant les affirmations précédentes sur l'apport extraordinaire des TICE sur la formation scolaire des élèves, d'autres chercheurs (Bailey, 1997; Basque, 2005; Dias 1999; Dockstader, 1999; Karsenti,

Zajc et Larose, 2001; Sandholtz, Ringstaff et Dwyer, 1997) relèvent que les TIC bénéficient d'un usage transversal et généralisé pour l'enseignement et l'apprentissage de toutes les disciplines. Dans l'optique d'une promotion intensifiée des TICE, le gouvernement camerounais a renforcé la formation à la manipulation de l'outil informatique par des textes stipulant l'existence des centres numériques dans les lycées et les universités. Nonobstant ces avancées, l'emploi des instruments classiques numériques ne construit pas encore des capacités professionnelles permettant aux futurs enseignants du Cameroun de s'autoformer ou d'assister leurs camarades dans l'appropriation des TIC. À ce jour, il n'est pas encore fait mention des constructions des plateformes pour des cours en ligne et encore moins, des Cours en Ligne Ouverts et Massifs (CLOM) par les étudiants en formation initiale. Comme on le voit, la question de l'efficacité pédagogique des TIC semble redoutable et difficile à cerner. Le Cameroun, encore dans une phase embryonnaire en la matière, met l'accent sur l'appropriation (pas toujours effective) des TIC.

D'après Millerand (1998), Basque et Lundgren-Cayrol (2003), les usages des TIC renvoient à des utilisations repérables et analysables, à travers des pratiques spécifiques d'une technologie dans le cadre de la formation des enseignants. En d'autres termes, l'usage des TIC dans le cadre de la formation et l'apprentissage des enseignants projette une description de leurs pratiques quotidiennes au cours des différentes utilisations qu'ils en font dans une perspective hybride de formation professionnelle. Aujourd'hui, les multiples usages des TIC en éducation ont permis d'établir différentes typologies, notamment celles de Raby (2004), Gauthier (2004), Karsenti et Ngamo (2007). Ces typologies d'usages, plus précisément celle de Karsenti et Ngamo (2007) mettent en évidence deux types d'emploi des TIC en classe. D'une part, les usages des TIC comme objets d'apprentissage et d'autre part, les usages pour l'acquisition des connaissances et l'enseignement des disciplines. Dans le monde, un troisième usage y paraît maintenant associé : l'autoformation et/ou l'autonomisation des apprentissages à travers un emploi plus approfondi des plateformes de formation interactives. Le web 2.0, socle de ladite formation construit et accroît les connaissances de millions de futurs diplômés en formation initiale. En 2002, l'un des plus grands centres de formation à distance de l'Afrique, situé à Madagascar, le Centre National de Télé-enseignement de Madagascar (CNTEMAD) a édifié près de 25% des étudiants de l'enseignement supérieur du pays¹. En Afrique centrale, le Cameroun semble l'un des pays les plus avancés en matière d'informatique, mais devrait encore faire des efforts pour se hisser à la hauteur des pays de l'Afrique de l'Ouest comme le Sénégal et du Maghreb tels que la Tunisie ou le Maroc qui associent quasi systématiquement toute formation universitaire aux TICE

(<http://www.observatoitic.org>).

En revenant à la typologie de l'usage des TIC, Karsenti et Ngamo (2007) déclarent que les usages des TIC se résument notamment à la recherche et au traitement d'informations, à l'interaction entre apprenant et machine, la communication avec les pairs, l'échange des ressources, la production des documents, la conception des supports multimédias, la présentation des contenus, etc. Les multiples usages des TIC professent de l'importance de ces derniers dans l'accompagnement pédagogique et des nombreuses études en la matière. Pour lever le pan sur la présente investigation, une méthode qualitatif-quantitatif, prenant son appui sur les méthodes d'analyse éprouvées, est privilégiée.

Méthodologie d'enquête

La méthode d'analyse ETED (Emploi Type Étudié dans sa Dynamique), Nicole Mandon (1990) présente trois points focaux complémentaires : la centration des démarches sur l'opérateur titulaire de l'emploi, la spécification des dimensions de compétence à explorer et la prise en compte de la diversité de situations. Ces trois éléments ont guidé le choix de la population, le type d'enquête et les items utilisés pour le recueil des informations.

Échantillonnage et population d'étude

L'échantillonnage non probabiliste et intentionnel (Merriam, 1988; Poupard, Deslauriers, Groulx, Laperrière, Mayer, et Pirès, 1997) servant d'approche dans la présente étude s'opère sur la base d'un groupe naturel de 38 futurs enseignants du DITE en formation initiale à l'ENS de Yaoundé I. Les critères subséquents prescrivent la grille de choix :

- être régulièrement inscrit à l'un des différents niveaux de formation initiale à l'ENS
- suivre régulièrement les cours dispensés
- avoir une adresse e-mail
- pouvoir utiliser Internet

Sur la base des critères énoncés ci-dessus, l'échantillon retenu se présente comme suit : 14 élèves au Niveau I dont 12 hommes et 2 femmes ; 15 élèves au Niveau II dont 12 hommes et 3 femmes et 9 élèves au Niveau IV dont 6 hommes et 3 femmes. L'absence des futurs éducateurs des niveaux 3 et 5 est justifiée par le fait qu'ils étaient en stage académique au moment de l'enquête.

Tableau I : Distribution de la population d'étude par niveau

<i>Classes ou niveaux</i>	<i>Effectifs</i>
Niveau I Niveau II Niveau IV	14
	15
	9

Total	38
--------------	-----------

2. Outils d'enquête et collecte de données

L'outil d'enquête prioritairement utilisé comprend le questionnaire et l'interview semi-directive. Le questionnaire a été élaboré en référence aux questionnaires de Karsenti et al. (2007) dans leur enquête sur les « conditions d'efficacité des TIC en pédagogie universitaire », de Djeumeni Tchamabe (2010) dans une investigation portant sur les pratiques pédagogiques des enseignants avec les TIC au Cameroun et enfin sur celui de Mian Bi Sehi Antoine (2010) tiré de son enquête sur

« Usages et compétence tic en formation initiale à l'ENS d'Abidjan (Côte d'Ivoire) : le cas des formateurs et des futurs enseignants ». Toutefois, les questions sur les compétences ont été élaborées à l'aide du standard des compétences pour les enseignants de l'UNESCO (2011). Les questions posées épousent une orientation quantitative, c'est-à-dire qu'elles sont exprimées sur une échelle graduée entre « moyennement d'accord », « assez d'accord », « complètement d'accord » et « pas d'accord du tout » Karsenti (2004). Après l'adoption du questionnaire et avant la validation de ce dernier, un prétest a été organisé auprès d'une douzaine de sujets appartenant à la population. La préenquête visait principalement à nous assurer de l'aspect fonctionnel du questionnaire. Plus précisément, il était question d'observer le niveau de compréhension des questions par les participants à l'enquête. C'est dans ce sillage que plusieurs questions ont été reformulées pour garantir la validité du questionnaire.

Thèmes de l'étude

Au terme de l'analyse théorique, deux principaux axes de l'étude sous forme de thèmes ont été adoptés. Il s'agit de la disponibilité, de l'accessibilité aux outils TIC et de la typologie d'usage des TIC par les futurs enseignants du DITE. Le tableau ci-dessous regroupe les thèmes assortis des modalités et des indicateurs.

Tableau II : Thèmes de l'étude

Thèmes	Indicateurs	Modalités
1. Accès aux TIC	<ul style="list-style-type: none"> ➤ Types d'outils ➤ Lieux d'accès ➤ Conditions d'accès 	Tout à fait disponible, plutôt disponible, tout à fait indisponible, plutôt indisponible
2. Typologie d'usages des TIC	<ul style="list-style-type: none"> ➤ Type d'activité ➤ Situations d'usage ➤ Type d'usage 	Tout à fait adéquat, plutôt adéquat, tout à fait inadéquat, plutôt inadéquat

3. Méthode d'analyse des données

L'analyse des données a été réalisée grâce à un masque de saisie élaboré dans le logiciel CSPRO 4. L'apurement des données a été effectué avec le logiciel SPSS. Des contrôles de cohérence sur les fichiers ont permis d'améliorer la qualité des données. En effet, la prépondérance des questions ouvertes dans le questionnaire a conduit au regroupement des modalités de ces questions selon leur sens pour créer soit de nouvelles modalités, soit de nouvelles variables. L'analyse des données avait deux dimensions : une dimension quantitative et une dimension qualitative. L'approche quantitative s'appuie sur la comptabilisation des avis et opinions de différentes catégories en termes de pourcentage. L'approche qualitative, quant à elle, s'appuie essentiellement sur les discours, les éléments du langage des différentes catégories d'analyse issues de l'approche quantitative.

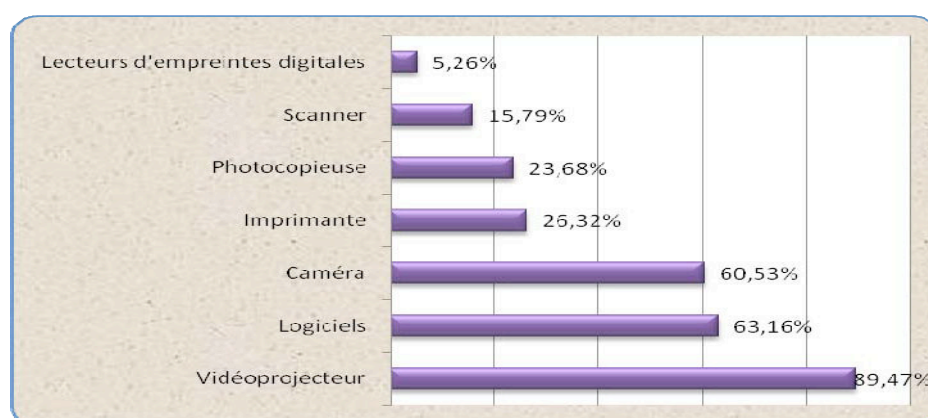
Résultats

Cette section présente les résultats en fonction de l'objectif de l'étude qui est, rappelons-le, de dresser la typologie d'usage des TIC par les futurs enseignants en formation initiale au DITE de l'ENS de Yaoundé I. La présentation des résultats s'opère sous forme de graphiques et de tableaux suivis de commentaires.

Disponibilité et état d'utilisation des outils TIC

Le graphique I retrace la présence des infrastructures TIC minimales. Ces dernières englobent une salle informatique connectée à Internet, un vidéo projecteur (89,47 %) et des logiciels éducatifs (63,16 %). Nous nous sommes limités au matériel existant pour déterminer la densité de leur utilisation par les enquêtés.

Graphique 1. Connaissance des outils présents dans le DITE



Source : ENS, 2012

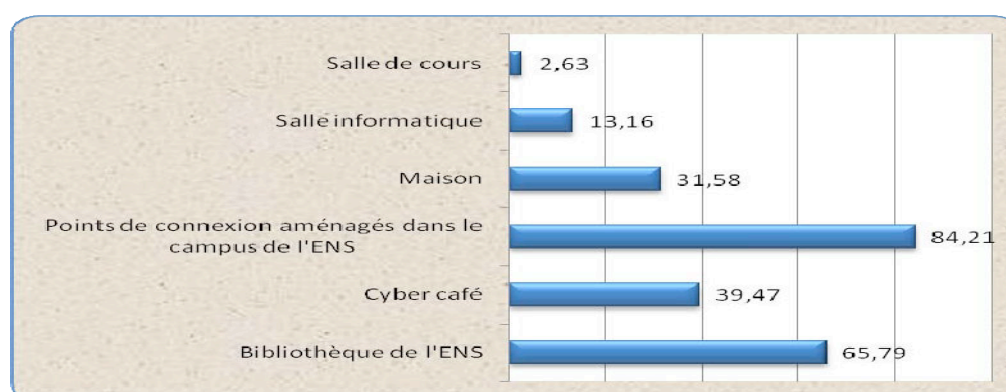
Les vidéoprojecteurs et les logiciels éducatifs servent à l'enseignement depuis plus d'une décennie. Aussi, est-il naturel de voir que presque tous les élèves-professeurs savent l'utiliser. En revanche, un seul vidéoprojecteur pour 38 apprenants semble insuffisant pour la transmission des connaissances. Le lecteur d'empreintes digitales intervient conjointement avec une fiche d'enregistrement manuelle. Les élèves préfèrent se diriger vers cette fiche qui, affirment-ils, paraît plus fiable que des empreintes inscrites quelque part. Les outils TIC méconnus ne gagnent pas la faveur des élèves. Malgré tout, l'École

normale s'attelle à généraliser son utilisation dans tous les départements. Néanmoins, certains enseignants demeurent réfractaires à l'apprentissage de leur utilisation. Il appert difficile d'utiliser internet et autres outils plus récents pour exploiter et expliciter leurs cours.

Accès à Internet

D'après Karsenti (2009), Internet est l'un des éléments minimum et indispensable au processus d'intégration des TIC dans l'éducation. C'est ainsi qu'au département d'informatique et des technologies éducatives et dans tout le campus de l'ENS de Yaouné I, le directeur a installé une connexion Internet libre sans fil et accessible à tous les futurs enseignants de lycée. Pour la disposition de la connexion Internet à l'ENS de Yaoundé I, tous les enquêtés répondent « oui », mais seulement 57,89 % d'entre eux la jugent d'un débit moyen, impropre à certains téléchargements.

Graphique 2. Lieux d'accès à la connexion Internet



Source : ENS, 2012- NB : lire les chiffres en pourcentage (2,63% ; 13,16%, etc.)

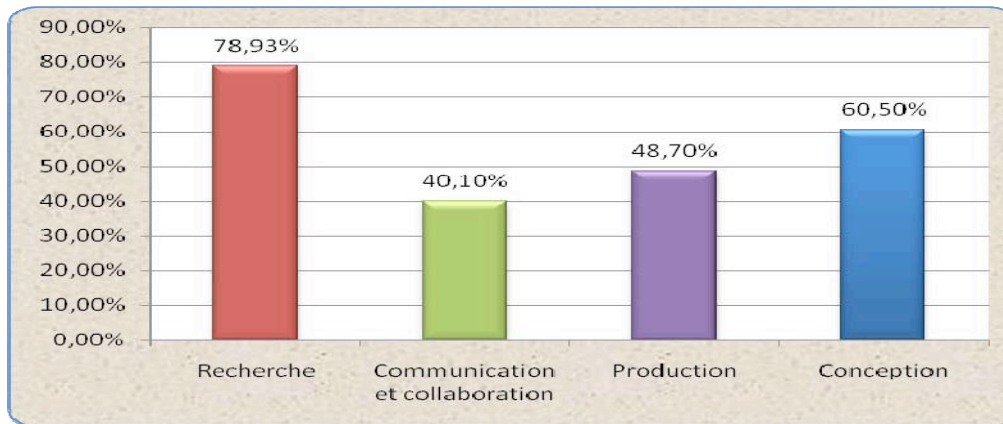
Le graphique ci-dessus dresse les statistiques du lieu d'accès à la connexion Internet. La distribution laisse apparaître que les lieux les plus accessibles sont les points de connexion aménagés sur le campus (84,21 %) et la bibliothèque de l'ENS (65,79 %). Ces forts taux se justifient d'une part, par le fait que ces lieux d'accès sont libres et ne demandent pas d'importants moyens financiers et d'autre part, par le fait de l'indisponibilité de la salle informatique et le nombre insuffisant de postes au laboratoire du DITE. L'observation montre que la majorité des apprenants disposent d'un ordinateur portable avec une connexion sans fil aussi bien au DITE qu'ailleurs.

Typologie d'usage des TIC par les futurs enseignants du DITE

La typologie des usages des TIC par les futurs enseignants dans le cadre de leur formation peut être regroupée dans la catégorie des usages des TIC pour l'apprentissage (Tardif, 1998). Il s'agit de la recherche d'informations (78,93%), la conception des

programmes d'applications, de tutoriels, de jeux éducatifs ou de pages Web (60,50%), la production de documents (48,70%), la communication ou la collaboration (40,10%).

Graphique 3. Typologie des usages des TIC par les futurs enseignants



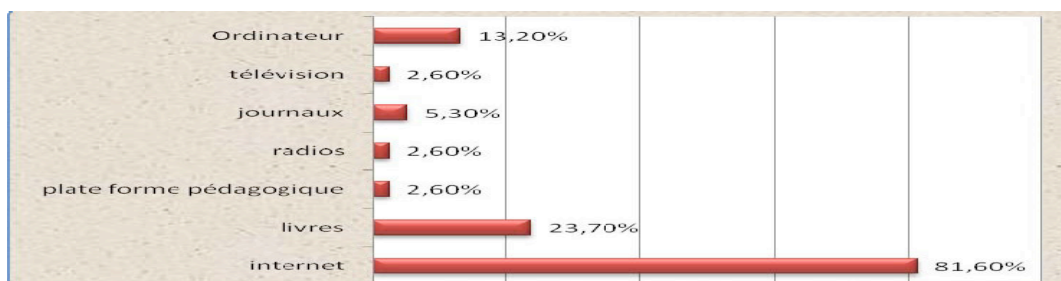
Source : ENS, 2012

La recherche d'information

La recherche d'information semble être la principale activité pour laquelle les futurs enseignants font usage des TIC. Elle s'effectue à l'aide de l'outil Internet avec un taux de 81,60 % comme le présente le graphique 4. Pour cette recherche, ils utilisent des moteurs de recherche tels que « Google » et

« Yahoo ». Seulement 23,7% des futurs éducateurs interrogés utilisent des livres pour la recherche à cause de leur coûts alors qu'Internet leur fournit une panoplie d'informations libres et gratuites.

Graphique 4. Outils TIC de recherche d'information par les futurs enseignants du DITE

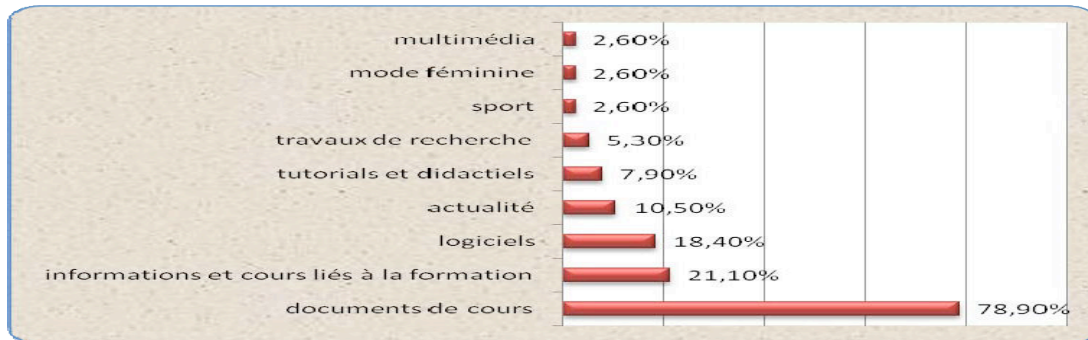


Source : ENS, 2012

En utilisant très peu les plateformes pédagogiques, les futurs pédagogues du DITE se dissocient de l'aspect pédagogique de leur apprentissage et seraient très peu compétents à initier des apprentissages auprès de leur homologue d'autres filières. En revanche, leur utilisation d'Internet prépare à une ouverture aux recherches universelles et ne se focalise

pas sur le Cameroun même si le graphique 5 illustre que les principales données ou informations recherchées sur Internet restent essentiellement en rapport avec le contenu des enseignements suivis en classe.

Graphique 5. Types d'informations recherchées sur Internet

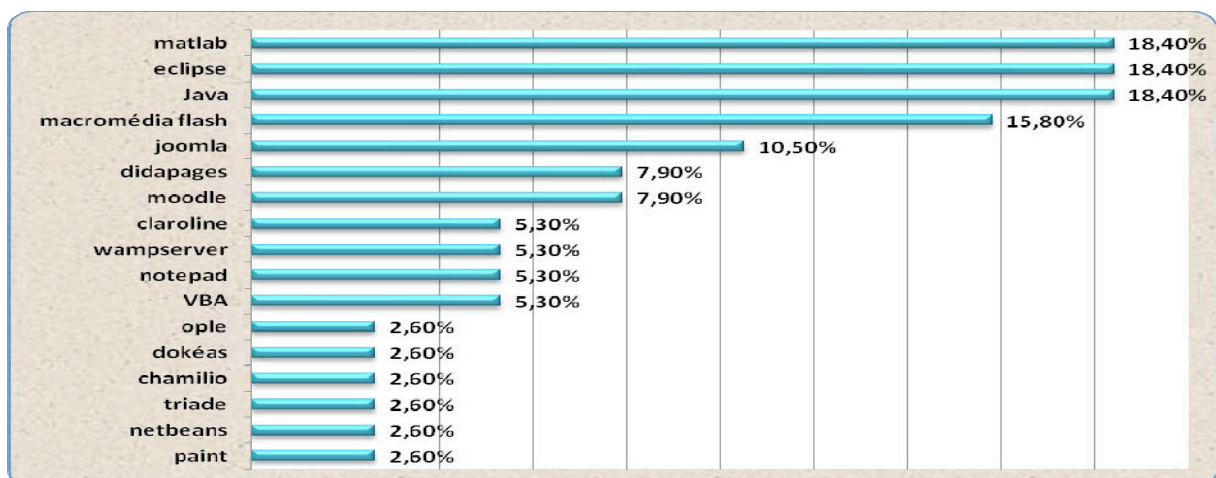


Source : ENS, 2012

La conception

La conception apparaît comme le deuxième usage préféré des TIC par les élèves-professeurs du DITE. Le penchant pour cette activité se justifie en partie par l'un des objectifs de formation au DITE qui aspire à former des technopédagogues aptes non seulement à enseigner l'informatique, mais aussi à produire des outils TIC d'auto-apprentissage et d'aide à la décision. L'examen de la distribution laisse apparaître que les données sont dispersées, car aucun indicateur n'a enregistré un résultat supérieur à 50 %. En poursuivant l'analyse, le graphique 6 laisse apparaître que les principaux outils utilisés pour la conception sont les logiciels *matlab* (18,40 %), *eclipse* (18,40 %), *java* (18,40 %) et *macromédia flash* (15,80 %).

Graphique 6. Outils de conception

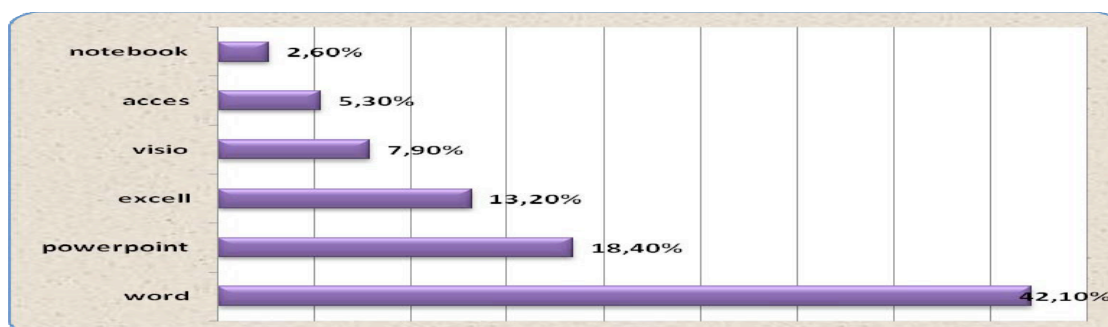


Source : ENS, 2012

La production de document

La production des documents marque la troisième position de la typologie des usages des TIC par les futurs enseignants modelés au DITE. Elle semble une activité indispensable au processus de formation et d'apprentissage de ceux-ci. Pour arriver à cet état de fait, comme le graphique 7 l'illustre, ils utilisent les logiciels standards de la *suite Microsoft Office* tels que "Word", "Excel" ou "PowerPoint" pour saisir un texte, effectuer des calculs et confectionner des présentations.

Graphique 7: Outils de création des contenus



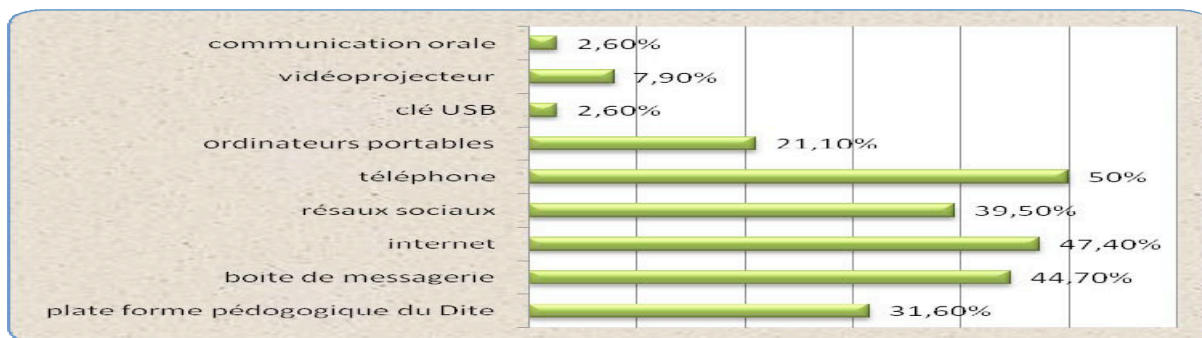
Source : ENS, 2012

Word paraît accessible à tous les apprenants, mais les autres programmes tout aussi utiles méritent une assimilation progressive par ceux-ci.

La communication et la collaboration entre le DITE et les autres départements

L'activité communicative interpelle l'usage des TIC par les élèves-professeurs du DITE. Toutefois, comme l'a indiqué le graphique 3, le taux de 40.10% montre que cette activité n'est pas des moindres. Les futurs enseignants utilisent les TIC pour communiquer et collaborer avec leurs condisciples, enseignants et pairs à travers les outils tels que le téléphone, Internet (gâce aux courriers électroniques et les réseaux sociaux) et via la plateforme du DITE (<http://dite-ens.cm/>). Ladite plateforme peut servir de lieu d'environnement numérisé pour le recyclage/la formation gratuite des autres enseignants et élèves professeurs de l'ENS.

Graphique 8. Outils de communication et de collaboration



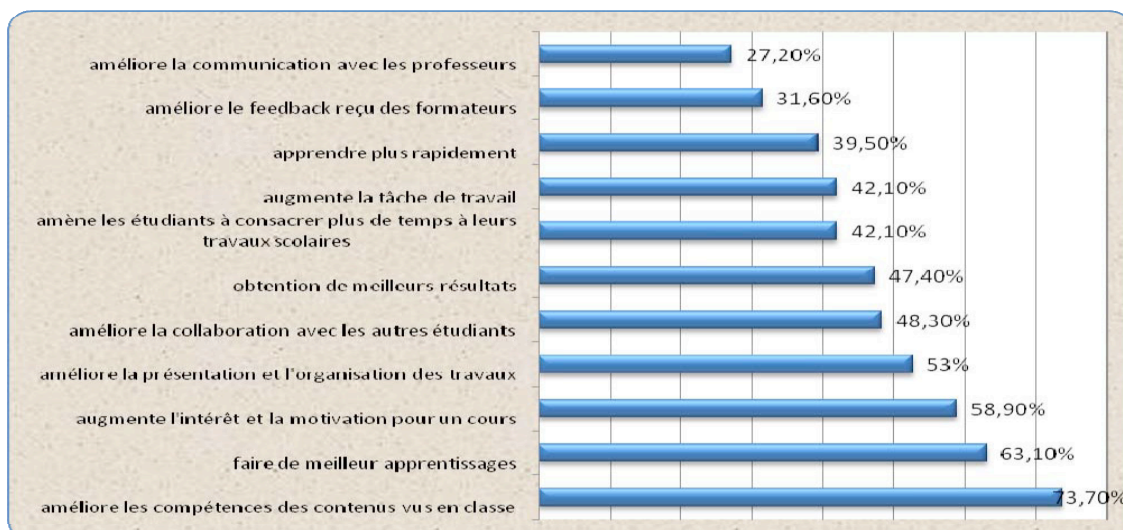
Source : ENS, 2012

La communication et la collaboration entre futurs enseignants portent sur le partage des liens de cours, des ressources, des travaux collaboratifs (exposé, montage des projets, échanges avec les autres départements) que donnent les enseignants. Par contre, les échanges entre élèves et enseignants du DITE se focalisent sur la clarification des zones d'ombre relatives au cours et à l'envoi des devoirs.

Représentation des élèves sur l'amélioration des connaissances et l'autonomisation en apprentissage

À la question de savoir quel était l'apport des TIC dans leur formation et dans leur autonomisation, les élèves-professeurs du DITE ont formulé plusieurs réponses qui ont été regroupées en termes de récurrence dans le graphique 9. Près de 73,70 % d'entre eux déclarent que l'utilisation des TIC est d'un apport incontestable. Grâce aux différentes plateformes et autres outils multimédias, les élèves du DITE ont approfondi et compris les contenus abordés en classe comme le laissaient présager certaines études antérieures (Karsenti et al. 2011). 63,10 % quant à eux pensent que l'usage des TIC leur permet de faire un meilleur apprentissage grâce à la faculté des TIC de susciter l'intérêt et la motivation pour un cours (58,90 %) malgré le faible pourcentage variant entre 47,40 % et 39,50 % des répondants pensant que les TIC permettaient l'obtention de meilleurs résultats et d'apprendre plus rapidement. Malgré les avantages qu'offrent les TIC, 42,10 % des apprenants pensent que l'intégration des outils TIC dans les pratiques augmente la tâche de travail. 53 % des futurs enseignants du fait de l'existence de nombreux outils de présentation et de production de travaux affichent de meilleures présentations des travaux de fin d'année. La plupart préfèrent soutenir en utilisant un PowerPoint. La communication, étant un élément important dans le processus de formation, il semble surprenant que seuls 31,60 % des répondants utilisent ces moyens pour interagir avec leurs enseignants et 27,20 % pensent que les TIC améliorent la communication avec les enseignants.

Graphique 9. Perception de la « valeur ajoutée des TIC » par les futurs enseignants du DITE



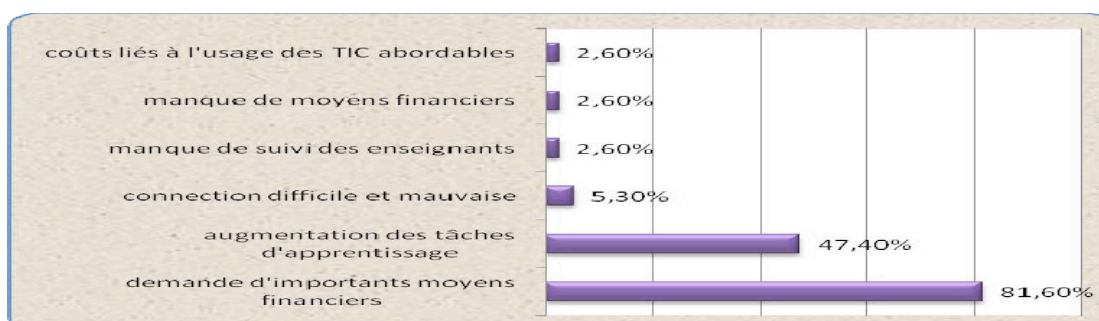
Source : ENS, 2012 (inspiré de karsenti et al. (2011) et modifié pour satisfaire les orientations du présent article)

Convaincus de l'apport des TIC pendant les enseignements, les élèves professeurs sont encouragés à en user dès leur affectation sur le terrain si le lycée dispose d'outils informatiques. En outre, ils espèrent impacter positivement sur leurs camarades malgré certaines obstructions infrastructurelles relatives à l'emploi des TIC.

Entraves à l'usage productif des TIC

L'usage des TIC dans les activités d'enseignement/apprentissage présente certes de nombreux avantages pour les futurs enseignants, mais impose aussi de nombreuses contraintes. Comme contraintes, 81,60 % des répondants sont unanimes que l'usage des TIC demande d'importants moyens financiers (achat d'ordinateur portable, coûts élevés pour une connexion personnelle et pour des appels téléphoniques, etc.) et 47,40 % des répondants affirment que l'usage des TIC augmente plutôt les tâches d'apprentissage.

Graphique 10 : Contraintes des futurs enseignants du DITE quant à l'utilisation des TIC



Source : ENS, 2012

Quelles implications et perspectives pour la formation initiale au Cameroun?

D'un point de vue global

L'usage des TIC favorise une bonne et rapide insertion pédagogique des TIC dans leur pratique pédagogique et développe par conséquent des compétences professionnelles accentuées auprès des élèves-professeurs. Aussi, les TIC doivent-elles être fortement considérées dans la formation des enseignants, car comme le constate Fonkoua (2006), les TIC qui constituent un dispositif important dans la formation des enseignants ne sont pas suffisamment élaborées ni prises en compte dans la formation initiale des enseignants à l'ENS. Même constat chez Cissé et Maïga (2006) lorsqu'ils déclarent que l'intégration des TIC dans la formation initiale des enseignants à l'ENS se synthétise à des cours de bureautique. Les formateurs de l'ENS, dans le cadre de leurs enseignements, doivent davantage faire usage des TIC et sensibiliser les futurs enseignants sur l'apport des TIC dans le processus d'apprentissage. De plus, selon l'OCDE (1998), pour la manifestation d'une estime tangible des TIC dans le secteur de l'éducation, il ne suffit pas de combiner l'utilisation de l'outil informatique avec les pédagogies existantes, mais il est pertinent d'adapter l'enseignement aux nouvelles possibilités qu'offrent ces dernières.

D'autres spécialistes insistent sur la nécessité de profiter de la « valeur pédagogique ajoutée » rendue possible avec l'usage des TIC pour permettre une amélioration de la pédagogie et du rapport au savoir (Perrenoud, 1998; Peraya, 2002; Karsenti, Peraya et Viens, 2002a). Bien que les outils seuls n'incitent pas les élèves-professeurs à en faire un bon usage, un effort doit néanmoins être amorcé quant à l'acquisition d'outils TIC supplémentaires pour catalyser un bon ratio outil/apprenant dans tous les départements de l'ENS. En outre, l'usage anecdotique des outils numériques classiques (CD, PowerPoint, ordinateurs, internet, etc.) ne conditionne pas l'appropriation de nouvelles formes d'enseignement, parmi lesquelles, les approches hybrides, le tutorat, la création des plateformes pédagogiques et les nouvelles formes d'évaluation. Les élèves-professeurs du DITE garantissent l'appropriation de ces orientations didactiques récentes par les autres étudiants en formation initiale.

Du point de vue de la didactique des langues

L'IFIC (2013) vient de lancer un appel à création des CLOM. L'AUF (2013), quant à elle, appelle à s'inscrire à des cours en ligne ouverts et massifs en anglais, pendant que les premières formations massives en français viennent de débiter (<http://fr.openclassrooms.com/mooc/html5-css5> ou encore les ProFLE). En France, des projets d'enseignement des langues sont légion dans un système d'interaction et de tutorat à distance. L'enseignement hybride gagne du terrain. Au Cameroun, des expérimentations voient le jour en la matière. La plus connue, mais pas encore optimisée ni amplement expérimentée est celle de Nkenlifack, Nangue, Demsong & Tchokomakoua (2012) qui proposent une plateforme éducative pour l'enseignement d'une langue de l'ouest Cameroun. Ladite plateforme naît d'une association entre informaticiens et enseignants de langues du secteur primaire désireux de promouvoir à grande échelle l'enseignement des langues camerounaises. Si les élèves-professeurs et les formateurs du DITE collaborent avec les linguistes, la maîtrise des outils numériques ne se limiterait pas à un usage de ceux-ci pour la recherche des informations, mais

s'étendrait à une optimisation de l'emploi des interfaces numériques et du web social pour une meilleure formation conduisant à une autonomisation de l'apprentissage. Les formations aux conceptions, mises en ligne et suivies d'apprentissages ne seraient plus l'apanage des organismes internationaux en charge de la promotion des langues, mais découleraient d'un département ayant une expertise avérée, moins coûteux et facilement accessible.

Conclusion

Au Cameroun, le secteur de l'éducation est de plus en plus marqué par l'intégration des outils TIC et des médias dans le processus d'enseignement et d'apprentissage. Il se pose la question de l'usage de ces outils TIC en éducation en général et, en particulier, dans la formation initiale des enseignants, domaine sur lequel cette étude est axée. Dans un tel contexte, une typologie d'usage des TIC par les futurs enseignants en formation au DITE de l'ENS de Yaoundé I a été élaborée. Les résultats de cette étude montrent d'une part d'importants efforts entrepris par le gouvernement et le Directeur de l'ENS de Yaoundé I en ce qui concerne l'acquisition et la mise à disposition de nombreux outils TIC dans le cadre de leur formation et apprentissage. D'autre part et loin d'être la moindre, une bonne intégration de ces outils a été observée par les futurs enseignants dans leurs différentes activités d'apprentissage notamment en ce qui concerne la recherche de l'information, les activités de conception, les activités de production et enfin les activités de communication et de collaboration avec les enseignants, les pairs, les camarades et les communautés d'apprentissage. De plus, une bonne perception de la « valeur ajoutée pédagogique » des TIC dans les activités d'enseignement/apprentissage des futurs enseignants est à noter même si ceux-ci se heurtent à des nombreux obstacles parmi lesquels le coût élevé des outils numériques et le manque de soutien ou de suivi de la part de leurs formateurs. L'éducation aux TICE étant l'intégration dans l'école et ses enseignements, de la connaissance des nouvelles formes d'apprentissage, de pensée, d'expression, de langages, de contenus et de comportements engendrés par les modes de production spécifiques des médias (Michel Pichette, 1995), les études sur la formation et les recyclages des autres apprenants en formation initiale se doit d'être sérieusement envisagé au Cameroun. À ce jour, il ne s'agit plus de collaborer pour une maîtrise des instruments classiques numériques, mais de les employer pour une autonomisation de l'apprentissage et un renforcement des capacités d'encadrement des étudiants à travers un suivi à distance et en présentiel, un système de tutorat et d'évaluation correspondant au degré de compétence en TIC. Les départements de langues amélioreraient les pratiques d'enseignement qui se verraient associées plus de ludique et de tâches interactives que de théories grammairiennes basées sur des œuvres littéraires parfois difficiles à décrypter.

Remerciements

à Marcelline Djeumeni Tchamabe (ENS de Yaoundé 1, Cameroun) pour l'aide apportée à cet article.

Références

- Bailey, G. D. (1997). What technology leaders need to know? The essential top 10 concepts for technology integration in the 21st Century. *Learning and Leading With Technology*, 25(1), 57-62.
- Basque, J., et Lundgren-Cayrol, K. (2002). Une typologie des typologies des applications des TIC en éducation. *Sciences et techniques éducatives*, 9(3-4), 263-289.
- Basque J. (2005). Une réflexion sur les fonctions attribuées aux TIC en enseignement universitaire. *Revue internationale des technologies en pédagogie universitaire*, 2(1), 30-41.
- Béché, E. (2012). Le rôle de l'usage de l'ordinateur dans le travail scolaire des apprenants : Opinions des élèves des lycées Général Leclerc et bilingue de Yaoundé (Cameroun). Dans C. Dili Pali (éd.), *Langages, Littérature et Éducation. De la poétique des savoirs endogènes aux mutations sociale* (p. 222-243). Paris : L'Harmattan.
- Buckley, C. A., Pitt, E., Norton, B. et Owens, T. (2010). Students approaches to study, conceptions of learning and judgments about the value of networked technologies. *Active Learning in Higher Education*, 11, 55-65.
- Cissé, D.-D. et Maïga, M. (2006). La formation des enseignants au Mali. *Bulletin formation et profession : chronique internationale*, 3(12). Consulté le 15 mai 2010 du site du bulletin Formation et Profession : <http://www.formationprofession.org/index.php>
- Dias, L. (1999). Integrating technology. *Learning and Leading with Technology*, 27 (3), 10-13,21.
- Djeumeni Tchamabe, M. (2010). *Pratiques pédagogiques des enseignants avec les TIC au Cameroun entre politiques publiques et dispositifs techno-pédagogiques, compétences des enseignants et compétences des apprenants, pratiques publiques et pratiques privées*. Thèse de Doctorat.
- Dockstader, J. (1999). Teachers of the 21st Century know the what, why, and how of technology integration. *T.H.E. Journal*, 26(6), 73-74.
- Fonkoua, P. (2005). Les dispositifs de formation des enseignants. Dans *Les institutions de formation des enseignants en Afrique subsaharienne : pour un renforcement des capacités* (p.108-119). Paris : UNESCO.
- Gauthier, P.-D. (2004). Taxonomies des outils TICE. *Portail des Technologies Educatives de la GEV, Mission Outils TICE*. Consulté le 4 juillet 2013 : <http://gev.industrie.gouv.fr/IMG/pdf/TaxonomieOutilsTICE-3.pdf>
- Karsenti, T et Ngamo, S.T (2007). Qualité de l'éducation en Afrique et rôle potentiel des TIC. *International review of education*, 53, 655-686.
- Karsenti, T., Savoie-Zajc, L. et Larose, F. (2001). Les futurs enseignants confrontés aux TIC : changements dans l'attitude, la motivation et les pratiques pédagogiques. *Éducation et francophonie*, 29, 1-29.
- Karsenti, Carole Raby, Hélène Meunier et Stéphane Villeneuve (2011). Usage des TIC en pédagogie universitaire : point de vue des étudiants. *International Journal of Technologies in Higher Education*, 8(3). Récupéré le 14 février 2014 du site: http://www.ritpu.org/IMG/pdf/RITPU_v08_n03_6.pdf
- Karsenti, T., Savoie-Zajc, L. et Larose, F. (2001). Les futurs enseignants confrontés aux TIC : changements dans l'attitude, la motivation et les pratiques pédagogiques. *Éducation et francophonie*, 29, 1-29.
- Karsenti, T. (2004). Les TIC et les futurs enseignants : les facteurs qui influencent leur utilisation. Dans D. Biron et M. Cividini (dir.), *La formation enseignante au temps des réformes* (p.3-16). Sherbrooke : Éditions du CRP.
- Karsenti, T. et Tchameni, N. (2009). *Intégration pédagogique des TIC en Afrique : Stratégies d'action et pistes de réflexion*. Harmattan-Cameroun.
- Mandon, N. (1990). *Analyser le sens et la complexité du travail : La méthode ETED*. Paris : L'Harmattan.
- Merriam, S.B. (1988). *Case study in education: A qualitative approach*. San Francisco (CA): Jossey- Bass.

- Mian Bi Sehi A. (2010). Usages et compétence TIC en formation initiale à l'ENS d'Abidjan (Côte d'Ivoire) : le cas des formateurs et des futurs enseignants OCDE (1998). *Compte rendu du séminaire sur les NTIC, Comité de l'Éducation*, OCDE. Récupéré du site : www.oecd.org
- Millerand, F. (1998). Usages des NTIC : les approches de la diffusion, de l'innovation et de l'appropriation. Dans *COMPOSITE*. [En ligne] http://composite.uqam.ca/98.1/articles/ntic_1.htm
- Nkenlifack, Nangue, Demsong & Tchokomakoua (2012). Approche de modernisation de l'enseignement des langues et cultures nationales du Cameroun à l'aide des TIC. *La Revue canadienne de l'apprentissage et de la technologie*, 38(2).
- Onguene Essono, L.M. et Onguene Essono, C. (2002). TIC et TIC à l'école : analyse des nouvelles pratiques enseignantes dans les salles de classes d'Afrique noire. *Langue et Communication*, 1(2).
- Peraya, D. (2002). Communication et nouvelles technologies. P. Perrig-Chiello et F. Darbellay (dir.), *Qu'est-ce que l'interdisciplinarité ? Les nouveaux défis de l'enseignement* (p. 117-143). Lausanne: Institut universitaire Kurt Bösch.
- Perrenoud, P. (1998). *Se servir des technologies nouvelles*. Genève : Faculté de psychologie et de sciences de l'éducation, Université de Genève.
- Poupart, J., Deslauriers, J.P., Groulx, L.H, Laperrière, A., Mayer, R. et Pires, A. (dir.) (1997). *La recherche qualitative : enjeux épistémologiques et méthodologiques*. Boucherville : Gaëtan Morin.
- Raby, C. (2004). *Analyse du cheminement qui a mené des enseignants du primaire à développer une utilisation exemplaire des technologies de l'information et de la communication (TIC) en classe*. Thèse de doctorat, Université du Québec à Montréal.
- Sandholtz, J.H., Ringstaff, C. et Dwyer, D.C. (1997). *La classe branchée. Enseigner à l'ère des technologies*. Montréal : Chenelière / McGraw-Hill.
- UNESCO (2011). *TIC UNESCO : Un référentiel des compétences pour les enseignants*. [En ligne] <http://www.unesco.org/new/fr/unesco/themes/icts/teacher-education/unesco-ict-competency-framework-for-teachers>

ANNEXE 4.1. : PHRASES EXPRIMANT UN PROBLEME DE RECHERCHE SURLIGNEES PAR LES EXPERTS AU SEIN DU MEME ARTICLE DE RECHERCHE

ARTICLE 1

- mais de se questionner si l'appropriation des TIC et TICE par ceux qui sont supposés former les élèves du secondaire.
- Les TIC apportent aux étudiants canadiens de nouvelles opportunités : en est-il de même pour les étudiants camerounais en formation initiale ?
- Le présent article, réalisé dans le contexte singulier de la formation initiale des enseignants de l'ÉNS de l'université de Yaoundé I, vise à essayer de déchiffrer, documenter et éclairer les perspectives pédagogiques qui se dessinent pour l'avenir des Écoles de formation initiales au Cameroun à travers l'usage des outils numériques et dans une moindre mesure du web 2.0.

ARTICLE 2

- Cet article propose une typologie des usages des TIC par les enseignants en formation initiale à l'École Normale Supérieure (ENS) de l'université de Yaoundé I à partir d'une enquête de terrain reposant sur une méthode essentiellement qualitatif-quantitative appliquée à 38 élèves-professeurs du Département d'Informatique et des Technologies Educatives (DITE).
- Les entretiens par questionnaire et l'analyse descriptive des contenus catalysent une taxinomie d'emploi des outils TIC par les futurs enseignants du DITE et précisent les

implications casuelles de son utilisation par les autres instructeurs, particulièrement, ceux de langues.

- Il est dès lors possible d'orienter les pédagogues en formation initiale quant au processus d'usage des TIC dans leurs pratiques d'auto-apprentissage et de transmission.

- En outre, les suggestions finales guideront les enseignants et les responsables des Écoles Normales sur les stratégies d'accompagnement utiles au processus d'insertion pédagogique des TIC dans l'éducation et d'extension à l'usage du web 2.0.

- Les TIC apportent aux étudiants canadiens de nouvelles opportunités : en est-il de même pour les étudiants camerounais en formation initiale ? Aussi, paraît-il légitime de s'interroger sur l'ampleur des usages des outils TIC par les futurs enseignants du secondaire formés au DITE et des autres départements de l'ENS comme ceux de Français et de langues et cultures camerounaises.

- Les TICE sont-elles d'un apport incontesté à tous les futurs pédagogues? Si aujourd'hui les pays africains ne se limitent plus à former les apprenants à la simple connaissance des outils numériques, mais s'étendent à l'usage du web social pour une autonomisation de l'apprentissage, qu'en est-il du Cameroun ?

- Le présent article, réalisé dans le contexte singulier de la formation initiale des enseignants de l'ÉNS de l'université de Yaoundé I, vise à essayer de déchiffrer, documenter et éclairer les perspectives pédagogiques qui se dessinent pour l'avenir des Écoles de formation initiales au Cameroun à travers l'usage des outils numériques

- Cette section présente les résultats en fonction de l'objectif de l'étude qui est, rappelons-le, de dresser la typologie d'usage des TIC par les futurs enseignants en formation initiale au DITE de l'ENS de Yaoundé I

ARTICLE 3

- Cet article propose une typologie des usages des TIC par les enseignants en formation initiale à l'École Normale Supérieure (ENS) de l'université de Yaoundé I à partir d'une enquête de terrain reposant sur une méthode essentiellement qualitatif-quantitative appliquée à 38 élèves-professeurs du Département d'Informatique et des Technologies Educatives (DITE).

- Les entretiens par questionnaire et l'analyse descriptive des contenus catalysent une taxinomie d'emploi des outils TIC par les futurs enseignants du DITE et précisent les implications casuelles de son utilisation par les autres instructeurs, particulièrement, ceux de langues

- Les TIC apportent aux étudiants canadiens de nouvelles opportunités : en est-il de même pour les étudiants camerounais en formation initiale ?
- Aussi, paraît-il légitime de s'interroger sur l'ampleur des usages des outils TIC par les futurs enseignants du secondaire formés au DITE et des autres départements de l'ENS comme ceux de Français et de langues et cultures camerounaises.
- Les TICE sont-elles d'un apport incontesté à tous les futurs pédagogues?
- Si aujourd'hui les pays africains ne se limitent plus à former les apprenants à la simple connaissance des outils numériques, mais s'étendent à l'usage du web social pour une autonomisation de l'apprentissage, qu'en est-il du Cameroun ?
- Le présent article, réalisé dans le contexte singulier de la formation initiale des enseignants de l'ÉNS de l'université de Yaoundé I, vise à essayer de déchiffrer, documenter et éclairer les perspectives pédagogiques qui se dessinent pour l'avenir des Écoles de formation initiales au Cameroun à travers l'usage des outils numériques et dans une moindre mesure du web 2.0.
- Comme on le voit, la question de l'efficacité pédagogique des TIC semble redoutable et difficile à cerner. Le Cameroun, encore dans une phase embryonnaire en la matière, met l'accent sur l'appropriation (pas toujours effective) des TIC.
- Pour lever le pan sur la présente investigation, une méthode qualitatif-quantitative, prenant son appui sur les méthodes d'analyse éprouvées, est privilégiée
- Dans un tel contexte, une typologie d'usage des TIC par les futurs enseignants en formation au DITE de l'ENS de Yaoundé I a été élaborée.
- Les résultats de cette étude montrent d'une part d'importants efforts entrepris par le gouvernement et le Directeur de l'ENS de Yaoundé I en ce qui concerne l'acquisition et la mise à disposition de nombreux outils TIC dans le cadre de leur formation et apprentissage.
- D'autre part et loin d'être la moindre, une bonne intégration de ces outils a été observée par les futurs enseignants dans leurs différentes activités d'apprentissage notamment en ce qui concerne la recherche de l'information, les activités de conception, les activités de production et enfin les activités de communication et de collaboration avec les enseignants, les pairs, les camarades et les communautés d'apprentissage. De plus, une bonne perception de la « valeur ajoutée pédagogique » des TIC dans les activités d'enseignement/apprentissage des futurs enseignants est à noter même si ceux-ci se heurtent à des nombreux obstacles parmi lesquels le coût élevé des outils numériques et le manque de soutien ou de suivi de la part de leurs formateurs.

ARTICLE 4

- Toutefois, d'après Salomon Tchameni Ngamo et Thierry Karsenti (2008), une meilleure intégration des TIC ne se polarise pas seulement à la mise à disposition des outils TIC aux enseignants et apprenants, mais nécessite un usage approprié de ceux-ci par les premiers acteurs du syncrétisme, à savoir les instructeurs pendant et après leur formation.

- Il ne s'agit pas non plus d'étudier, comme Marcelline Tchamabe (2013), l'impact de l'usage des TIC sur les apprenants d'un niveau équivalent au secondaire ou de s'interroger sur les applications réelles des textes de loi quant à l'enseignement de l'informatique ou de son usage pour enseigner, mais de se questionner si l'appropriation des TIC et TICE par ceux qui sont supposés former les élèves du secondaire.

- Nous souscrivons plutôt aux recherches empiriques effectuées par Karsenti et al (2011) sur l'usage des TIC en pédagogie universitaire.

- Les TIC apportent aux étudiants canadiens de nouvelles opportunités : en est-il de même pour les étudiants camerounais en formation initiale ? Aussi, paraît-il légitime de s'interroger sur l'ampleur des usages des outils TIC par les futurs enseignants du secondaire formés au DITE et des autres départements de l'ENS comme ceux de Français et de langues et cultures camerounaises.

- Les TICE sont-elles d'un apport incontesté à tous les futurs pédagogues?

- Comme on le voit, la question de l'efficacité pédagogique des TIC semble redoutable et difficile à cerner.

- Le Cameroun, encore dans une phase embryonnaire en la matière, met l'accent sur l'appropriation (pas toujours effective) des TIC.

- À la question de savoir quel était l'apport des TIC dans leur formation et dans leur autonomisation, les élèves-professeurs du DITE ont formulé plusieurs réponses qui ont été regroupées en termes de récurrence dans le graphique 9

- Il se pose la question de l'usage de ces outils TIC en éducation en général et, en particulier, dans la formation initiale des enseignants, domaine sur lequel cette étude est axée.

ARTICLE 5

- Cet article propose une typologie des usages des TIC par les enseignants en formation initiale à l'École Normale Supérieure (ENS) de l'université de Yaoundé I à partir d'une enquête de terrain reposant sur une méthode essentiellement qualitatif-quantitative appliquée à 38 élèves-professeurs du Département d'Informatique et des Technologies Educatives (DITE)
- Nous souscrivons plutôt aux recherches empiriques effectuées par Karsenti et al (2011) sur l'usage des TIC en pédagogie universitaire.
- Les TIC apportent aux étudiants canadiens de nouvelles opportunités : en est-il de même pour les étudiants camerounais en formation initiale ?
- Aussi, paraît-il légitime de s'interroger sur l'ampleur des usages des outils TIC par les futurs enseignants du secondaire formés au DITE et des autres départements de l'ENS comme ceux de Français et de langues et cultures camerounaises. Les TICE sont-elles d'un apport incontesté à tous les futurs pédagogues?
- Si aujourd'hui les pays africains ne se limitent plus à former les apprenants à la simple connaissance des outils numériques, mais s'étendent à l'usage du web social pour une autonomisation de l'apprentissage, qu'en est-il du Cameroun ?
- Le présent article, réalisé dans le contexte singulier de la formation initiale des enseignants de l'ÉNS de l'université de Yaoundé I, vise à essayer de déchiffrer, documenter et éclairer les perspectives pédagogiques qui se dessinent pour l'avenir des Écoles de formation initiales au Cameroun à travers l'usage des outils numériques et dans une moindre mesure du web 2.0.

ARTICLE 6

- Toutefois, d'après Salomon Tchameni Ngamo et Thierry Karsenti (2008), une meilleure intégration des TIC ne se polarise pas seulement à la mise à disposition des outils TIC aux enseignants et apprenants, mais nécessite un usage approprié de ceux-ci par les premiers acteurs du syncrétisme, à savoir les instructeurs pendant et après leur formation
- Aussi, paraît-il légitime de s'interroger sur l'ampleur des usages des outils TIC par les futurs enseignants du secondaire formés au DITE et des autres départements de l'ENS comme ceux de Français et de langues et cultures camerounaises
- Les TICE sont-elles d'un apport incontesté à tous les futurs pédagogues?
- Le présent article, réalisé dans le contexte singulier de la formation initiale des enseignants de l'ÉNS de l'université de Yaoundé I, vise à essayer de déchiffrer,

documenter et éclairer les perspectives pédagogiques qui se dessinent pour l'avenir des Écoles de formation initiales au Cameroun à travers l'usage des outils numériques et dans une moindre mesure du web 2.0.

- Ces typologies d'usages, plus précisément celle de Karsenti et Ngamo (2007) mettent en évidence deux types d'emploi des TIC en classe.
- D'une part, les usages des TIC comme objets d'apprentissage et d'autre part, les usages pour l'acquisition des connaissances et l'enseignement des disciplines. Dans le monde, un troisième usage y paraît maintenant associé : l'autoformation et/ou l'autonomisation des apprentissages

ARTICLE 7

- Les TIC apportent aux étudiants canadiens de nouvelles opportunités : en est-il de même pour les étudiants camerounais en formation initiale ? Aussi, paraît-il légitime de s'interroger sur l'ampleur des usages des outils TIC par les futurs enseignants du secondaire formés au DITE et des autres départements de l'ENS comme ceux de Français et de langues et cultures camerounaises.
- Les TICE sont-elles d'un apport incontesté à tous les futurs pédagogues?
- Si aujourd'hui les pays africains ne se limitent plus à former les apprenants à la simple connaissance des outils numériques, mais s'étendent à l'usage du web social pour une autonomisation de l'apprentissage, qu'en est-il du Cameroun ?
- Après un point sur l'usage des TIC dans l'enseignement, nous disséquons l'accessibilité au multimédia, dresserons une typologie de l'usage des TIC par les futurs enseignants et envisagerons les possibilités de recyclage et d'extension de l'emploi spécifique du web 2.0 aux autres enseignants, spécifiquement, ceux de langues.
- Il s'agit de la disponibilité, de l'accessibilité aux outils TIC et de la typologie d'usage des TIC par les futurs enseignants du DITE.

ANNEXE 4.2 : PHRASES EXPRIMANT UN PROBLEME DE RECHERCHE SURLIGNEES PAR LES EXPERTS AU SEIN DE 5 ARTICLES DE RECHERCHE DIFFERENTS

ARTICLE 1

- notre article tente de mettre en évidence les principaux facteurs qui ont permis à ce projet de renouveler, de l'intérieur, l'organisation de l'établissement, de lui redonner un nouveau souffle.
- notre article tente de mettre en évidence les principaux facteurs qui ont permis à ce projet de renouveler, de l'intérieur, l'organisation de l'établissement, de lui redonner un nouveau souffle.
- L'article est à la fois un retour sur une expérience et une tentative d'explication du succès de cette dernière, en utilisant principalement le cadre théorique du leadership partagé.
- Les différents éléments pris en compte permettent une interrogation sur les conditions de la réussite mais également une réflexion sur l'absence de duplication d'un tel dispositif au sein d'autres établissements de l'Université de Dakar.

ARTICLE 2

- Après une expérience de neuf ans, n'est-il pas temps de procéder à une évaluation du dispositif de l'UVT ?
- Ne faut-il pas que l'UVT s'engage dans une démarche qualité visant une amélioration continue des résultats par une utilisation plus efficace des moyens humains et matériels et une amélioration des processus pour les offres de formation qu'elle dispense ?
- En l'absence de normes, comment pourrait-on soutenir une politique nationale d'enseignement et de formation de grande qualité dans une optique de compétition et d'innovation internationale ?
- Quelles normes sont appliquées et quels éléments du dispositif sont concernés

ARTICLE 3

Aucune source spécifiée dans le document actif.

- Cet article est constitué de trois parties principales : Dans ce point

- **Aucune source spécifiée dans le document actif.ctif (I)**, nous souhaitons ouvrir le débat sur la place de l'innovation numérique à "l'école" en montrant **combien différentes tensions relativement stériles** tout à la fois l'animent et l'immobilisent _que ce soit au niveau des finalités, des methods, de l'évaluation et des outils.
- Nous montrerons que les perspectives ainsi dégagées conduisent à une **révision de la place centrale de l'humain** à la fois autonome et responsable
- Nous pourrions tout aussi bien argumenter que ce dernier ne fera que s'élargir entre ceux qui sont connectés à internet et les autres.
- *Avant de continuer notre analyse, il est bon de comprendre qu'il y a MOOC et MOOC !*

ARTICLE 4

- Toutefois, d'après Salomon Tchameni Ngamo et Thierry Karsenti (2008), une meilleure intégration des TIC ne se polarise pas seulement à la mise à disposition des outils TIC aux enseignants et apprenants, mais nécessite un usage approprié de ceux-ci par les premiers acteurs du syncrétisme, à savoir les instructeurs pendant et après leur formation.
- 14. Il ne s'agit pas non plus d'étudier, comme Marcelline Tchamabe (2013), l'impact de l'usage des TIC sur les apprenants d'un niveau équivalent au secondaire ou de s'interroger sur les applications réelles des textes de loi quant à l'enseignement de l'informatique ou de son usage pour enseigner, mais de se questionner si l'appropriation des TIC et TICE par ceux qui sont supposés former les élèves du secondaire. Nous souscrivons plutôt aux recherches empiriques effectuées par Karsenti et al (2011) sur l'usage des TIC en pédagogie universitaire.
- 15. Les TIC apportent aux étudiants canadiens de nouvelles opportunités : en est-il de même pour les étudiants camerounais en formation initiale ?
- 16. Aussi, paraît-il légitime de s'interroger sur l'ampleur des usages des outils TIC par les futurs enseignants du secondaire formés au DITE et des autres départements de l'ENS comme ceux de Français et de langues et cultures camerounaises.
- 17. Les TICE sont-elles d'un apport incontesté à tous les futurs pédagogues?
- 18. Comme on le voit, la question de l'efficacité pédagogique des TIC semble redoutable et difficile à cerner. Le Cameroun, encore dans une phase embryonnaire en la matière, met l'accent sur l'appropriation (pas toujours effective) des TIC.

19. À la question de savoir quel était l'apport des TIC dans leur formation et dans leur autonomisation, les élèves-professeurs du DITE ont formulé plusieurs réponses qui ont été regroupées en termes de récurrence dans le graphique 9
20. Il se pose la question de l'usage de ces outils TIC en éducation en général et, en particulier, dans la formation initiale des enseignants, domaine sur lequel cette étude est axée.

ARTICLE 5

21. Dans cet article, nous présentons les résultats d'une recherche exploratoire conduite autour d'un dispositif d'intégration de l'iPod Touch dans une école de l'école post-obligatoire.
22. Nos résultats mettent à jour plusieurs avantages pédagogiques de l'iPod Touch dont un élément central constitue la multimodalité l'appareil qui permet l'accès à des ressources variées mobilisables ponctuellement.
23. Néanmoins, le caractère portatif et les limitations de l'outil posent des défis aux niveaux institutionnel et pédagogique que nous présentons dans cette étude
24. Malgré ces limitations avérées des ANP utilisés à des fins d'apprentissage, les recherches effectuées mettent en évidence un certain nombre de potentialités et d'avantages qui méritent d'être soulignés
25. L'auteur insiste néanmoins sur les bénéfices que peut apporter ce support pour une partie des étudiants (Taylor, 2009).