

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LECTURE ET ANALYSE ARGUMENTAIRE DE TEXTE ASSISTÉES PAR ORDINATEUR (LAATAO):
DÉTECTION AUTOMATIQUE D'ARGUMENT DANS LES TEXTES DE HAUT NIVEAU THÉORIQUE

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

AU DOCTORAT EN INFORMATIQUE COGNITIVE

PAR

FRANCIS LAREAU

DÉCEMBRE 2023

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.12-2023). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

À l'achèvement de ce périple intellectuel et académique, je suis profondément reconnaissant envers toutes les personnes qui ont marqué mon chemin et apporté leur pierre à l'édification de cette thèse. Leur influence, variée et précieuse, a été une véritable source d'inspiration, de soutien et de sagesse pour moi.

Je souhaite avant tout exprimer ma gratitude sincère à mes directeurs de thèse, Jean-Guy Meunier et Roger Villemaire, pour leur accompagnement avisé, leur patience et leur appui constant durant cette aventure. Leurs conseils judicieux et leur habileté à stimuler ma pensée ont joué un rôle déterminant dans l'accomplissement de ce projet.

Ma profonde reconnaissance va également à Christophe Malaterre de la Chaire de recherche du Canada en philosophie des sciences de la vie, pour son mentorat précieux, son expertise approfondie et ses orientations éclairées qui ont considérablement enrichi ma compréhension des écrits philosophiques et scientifiques.

Un remerciement particulier est adressé à Élias Rizkallah ainsi qu'à l'ensemble de mes collègues du Centre ATO, du LANCI et du CIRST. Leur collaboration, les échanges intellectuels captivants et l'esprit de camaraderie qu'ils ont partagé avec moi ont été cruciaux pour élargir mes horizons et approfondir mes compétences dans le domaine de la LATAO.

Je suis aussi infiniment reconnaissant envers ma famille et ma conjointe, pour leur amour sans limites, leur soutien moral et leur encouragement continu. Leur présence et leur confiance infaillible ont été le pilier de cette réalisation.

Enfin, je souhaite remercier avec gratitude les Fonds de Recherche du Québec pour leur support financier essentiel tout au long de mon parcours. Un immense merci s'étend à toutes les personnes qui, de près ou de loin, ont apporté leur contribution à cette thèse, que ce soit à travers un soutien logistique, technique ou émotionnel. Chaque apport a été un trésor inestimable dans la construction de ce travail de recherche.

DÉDICACE

À l'intersection de la passion et de la persévérance, ce travail est dédié à ceux qui m'ont guidé et inspiré tout au long de ce voyage intellectuel.

À mon regretté ami et mentor, Jean-Guy Meunier, dont la sagesse et la bienveillance ont laissé une empreinte indélébile sur ma vie et mon travail. Ton héritage intellectuel et humain continue de m'inspirer chaque jour.

À mes chers parents, et ma conjointe, pour leur amour inconditionnel et leur soutien sans faille. Vous avez allumé la flamme de la curiosité en moi et m'avez toujours encouragé à poursuivre la connaissance avec ardeur et intégrité.

À mes professeurs et mentors, pour avoir éclairé mon chemin avec leur savoir et leur expertise. Votre guidance a été le phare dans les moments d'incertitude, illuminant les nuances complexes de la pensée.

À mes amis et collègues, pour les discussions enrichissantes et les débats stimulants. Vous avez été les compagnons indispensables dans cette exploration des profondeurs de l'argumentation et de la raison.

Et enfin, à tous ceux qui trouvent dans les textes philosophiques et scientifiques non seulement des réponses, mais des questions qui alimentent l'esprit critique et la réflexion continue. Que ce travail contribue modestement à notre quête commune de compréhension et de vérité.

Cette thèse est un hommage à votre influence inestimable et à notre partage d'un amour indéfectible pour la connaissance.

AVANT-PROPOS

La vérité est rarement pure et jamais simple.
Oscar Wilde

À l'aube de cette exploration académique, il me semble essentiel de prendre un moment pour refléter sur les origines et les aspirations qui ont façonné cette thèse doctorale dédiée au forage d'argument dans les textes philosophiques et scientifiques. Ce travail est le fruit d'une quête personnelle et intellectuelle, née d'une profonde fascination pour les méandres de la pensée humaine et la complexité des arguments qui façonnent nos discours les plus élaborés.

L'impulsion de ce projet trouve ses racines dans mon mémoire de maîtrise, réalisé sous la direction de Jean-Guy Meunier, où j'ai exploré l'analyse de concepts dans les textes philosophiques. Cette première étape a éveillé en moi un intérêt particulier pour la manière dont les idées se manifestent, interagissent et évoluent dans le discours. Elle fut le lieu de rencontres et d'amitiés essentielles dans le développement de mon parcours, notamment au LANCI (Laboratoire d'ANalyse Cognitive de l'Information). Elle a posé les bases de mon exploration ultérieure des textes philosophiques et scientifiques.

Mon cheminement s'est par la suite enrichi à la Chaire de recherche du Canada en philosophie des sciences de la vie, sous la direction de Christophe Malaterre. Ici, mon travail sur l'analyse thématique des textes philosophiques et scientifiques a affiné ma capacité à discerner les nuances et les complexités inhérentes à ces types de discours.

Au Centre ATO, collaborant avec Élias Rizkallah et mes collègues du CIRST (Centre Interuniversitaire de Recherche sur la Science et la Technologie), mon immersion dans l'enseignement de la LATAO (Lecture et Analyse de Texte Assistée par Ordinateur) a ouvert de nouveaux horizons. Cette expérience m'a permis de combiner ma passion pour l'enseignement, la technologie et une approche analytique rigoureuse du texte, enrichissant mes méthodes de recherche et élargissant mes perspectives.

Cette thèse est le fruit de ces diverses expériences formatrices, chacune contribuant à façonner ma vision et ma compréhension de l'importance des structures argumentatives dans les discours philosophiques et scientifiques. Elle se veut une exploration technique autant qu'une célébration de la pensée critique et de

la rigueur intellectuelle, avec la conviction que déchiffrer ces textes est essentiel pour avancer dans notre compréhension des débats qui influencent notre société.

Francis Lareau

TABLE DES MATIÈRES

REMERCIEMENTS	ii
DÉDICACE	iii
AVANT-PROPOS.....	iv
LISTE DES FIGURES.....	xii
LISTE DES TABLEAUX	xiii
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES.....	xvi
LISTE DES SYMBOLES ET DES UNITÉS	xxiv
RÉSUMÉ	xxv
ABSTRACT	xxvi
INTRODUCTION	1
CHAPITRE 1 La lecture et l'analyse de texte.....	6
1.1 Introduction	6
1.2 Le paradigme externaliste de l'analyse du discours	7
1.3 Le paradigme internaliste de l'analyse du discours	7
1.3.1 La théorie internaliste structuraliste	8
1.3.2 La théorie internaliste formelle ou cognitive.....	8
1.3.3 La théorie internaliste herméneutique.....	9
1.4 Le texte de haut niveau théorique.....	9
1.4.1 Les grands types structurels de textes de haut niveau théorique.....	10
1.4.1.1 La structure textuelle longue du texte de haut niveau théorique.....	10
1.4.1.2 La structure textuelle courte du texte de haut niveau théorique	10
1.4.2 Les grands objectifs des textes de haut niveau théorique	11
1.4.2.1 L'approche exploratoire et le texte de type analytique	12
1.4.2.2 L'approche résolutive et le texte de type argumentatif.....	12
1.4.3 Circonscription du discours de haut niveau théorique.....	16
1.5 Conclusion.....	19
CHAPITRE 2 La lecture et l'analyse de texte assistées par ordinateur	20
2.1 Introduction	20
2.2 La lecture et l'analyse de texte assistées par ordinateur.....	21
2.3 L'intersectionnalité de la LATAO	25

2.3.1	Les sciences cognitives.....	25
2.3.2	La fouille de texte.....	26
2.3.3	L'apprentissage machine	26
2.3.4	Le traitement automatique des langues naturelles.....	27
2.3.5	Les humanités numériques.....	27
2.4	Les sous-objectifs de la LATAO.....	28
2.4.1	L'extraction d'information	29
2.4.2	La catégorisation.....	29
2.4.3	La visualisation.....	30
2.4.4	La synthèse	30
2.5	Les spécialisations de la LATAO.....	30
2.5.1	L'analyse conceptuelle.....	31
2.5.2	L'analyse thématique.....	31
2.5.3	L'analyse d'argument.....	33
2.6	Conclusion.....	34
CHAPITRE 3 L'argumentation		35
3.1	Introduction	35
3.2	Les définitions du dictionnaire.....	35
3.3	Les définitions philosophiques ou scientifiques	36
3.3.1	L'approche rhétorique antique et celle classique.....	37
3.3.2	L'approche rhétorique moderne	38
3.3.3	L'approche épistémologique	39
3.3.4	L'approche consensualiste.....	40
3.3.5	L'approche communicationnelle	41
3.4	Conclusion.....	44
CHAPITRE 4 Les modèles computationnels de l'argumentation.....		45
4.1	Introduction	45
4.2	Les modèles rhétoriques.....	45
4.2.1	Les applications computationnelles des modèles rhétoriques	48
4.3	Les modèles épistémologiques	49
4.3.1	Les modèles épistémologiques à base de graphes.....	49
4.3.2	Les modèles épistémologiques informels.....	53
4.3.3	Les applications computationnelles des modèles épistémologiques.....	55
4.4	Les modèles consensualistes	55
4.4.1	Les applications computationnelles des modèles consensualistes	57
4.5	Les modèles communicationnels.....	57
4.5.1	Les modèles communicationnels en sciences cognitives	57
4.5.2	Les modèles communicationnels en linguistique	61
4.5.3	Les applications computationnelles des modèles communicationnels.....	63

4.6	Les <i>corpora</i> sur lesquels sont appliqués les différents modèles.....	64
4.7	Conclusion.....	65
CHAPITRE 5 La modélisation computationnelle de la tâche d’analyse d’argument		67
5.1	Introduction	67
5.2	La décomposition de la tâche d’analyse d’argument	68
5.2.1	La définition du corpus	70
5.2.2	Le décodage du corpus	70
5.2.3	La détection des structures argumentatives	71
5.2.4	La représentation des résultats	72
5.2.5	L’interprétation et évaluation des résultats	72
5.3	Conclusion.....	72
CHAPITRE 6 Implémentation de la tâche de détection des arguments.....		74
6.1	Introduction	74
6.2	Les différentes représentations vectorielles du texte	74
6.2.1	Les représentations vectorielles classiques.....	75
6.2.2	Les représentations vectorielles par plongement lexical	76
6.2.2.1	Les représentations vectorielles <i>word2vec</i> et <i>doc2vec</i>	76
6.2.2.2	Les représentations vectorielles de type BERT.....	77
6.2.2.3	Les représentations vectorielles inspirées des modèles génératifs comme GPT ou ChatGPT.....	79
6.3	Les différentes approches de détection des structures argumentatives	80
6.3.1	Les approches déductives de la détection des structures argumentatives.....	80
6.3.2	Les approches inductives de la détection des structures argumentatives.....	80
6.3.3	Les approches hybrides de la détection des structures argumentatives	81
6.4	Les représentations visuelles des structures argumentatives	82
6.5	L’évaluation des outils de détection	82
6.5.1	L’évaluation qualitative.....	82
6.5.2	L’évaluation quantitative	83
6.5.3	Les <i>corpora</i> annotés.....	84
6.6	Conclusion.....	84
CHAPITRE 7 Les hypothèses de recherche		86
7.1	Introduction	86
7.2	Le contexte de citation.....	87
7.3	La structure argumentaire des textes de haut niveau théorique	88
7.3.1	Les microstructures argumentaires	88
7.3.2	Les macrostructures argumentaires	89
7.3.2.1	La macrostructure « discussion »	89
7.3.2.2	La macrostructure « introduction ».....	90
7.3.2.3	La macrostructure « méthodes ».....	90

7.3.2.4	La macrostructure « résultats »	90
7.3.2.5	La structure « corps du texte » et autres structures adjacentes	91
7.3.3	Les structures argumentaires étendues	92
7.4	Les hypothèses	96
7.4.1	L'identification de l'appartenance d'un contexte de citation à la macrostructure IMRAD.....	96
7.4.2	L'identification de la thèse principale	97
7.4.2.1	L'identification de la thèse principale à l'aide de l'emplacement	97
7.4.2.2	L'identification de la thèse principale à l'aide de connecteurs de discours	98
7.4.2.3	L'identification de la thèse principale à l'aide de propriétés syntaxiques ou sémantiques.....	99
7.4.2.3.1	L'identification de la thèse principale à l'aide du titre	99
7.4.2.3.2	L'identification de la thèse principale à l'aide du résumé	101
7.4.2.3.3	L'identification de la thèse principale à l'aide de l'introduction	102
7.4.2.3.4	L'identification de la thèse principale à l'aide de la discussion	103
7.5	Conclusion	103
CHAPITRE 8 La détection de l'appartenance des contextes de citation aux macrostructures argumentatives		
		105
8.1	Introduction	105
8.2	Contexte et état de l'art.....	106
8.2.1	Les techniques de représentation	107
8.2.2	Les méthodes de classification	108
8.2.2.1	Les modèles de classification basée sur la catégorisation.....	110
8.2.2.1.1	Le modèle des <i>K</i> -moyennes	110
8.2.2.1.2	Le modèle de mélange gaussien.....	111
8.2.2.2	Les modèles de classification instanciée	112
8.2.2.2.1	La méthode des <i>K</i> plus proches voisins	112
8.2.2.2.2	La méthode des voisins de rayon	113
8.2.2.2.3	La méthode du plus proche centroïde.....	114
8.2.2.3	Les modèles de classification linéaire.....	114
8.2.2.3.1	La descente de gradient stochastique	115
8.2.2.3.2	Les machines à vecteurs de support (SVM).....	116
8.2.2.3.3	La régression logistique	117
8.2.2.3.4	La régression <i>Ridge</i>	117
8.2.2.3.5	La classification passive-agressive	118
8.2.2.4	Les modèles de classification arborescente	119
8.2.2.4.1	Les arbres de décision.....	119
8.2.2.4.2	Les forêts aléatoires.....	120
8.2.2.4.3	Les classifieurs d'arbre supplémentaire	121
8.2.2.4.4	L'amplification de gradient.....	121
8.2.2.5	Les modèles de classification neuronale	122
8.2.2.5.1	Le perceptron.....	122
8.2.2.5.2	Le perceptron multicouche.....	123
8.2.2.5.3	La classification de séquence BERT.....	124
8.3	Les méthodes	125
8.3.1	La constitution du corpus	125

8.3.2	Le décodage du corpus	126
8.3.3	Les méthodes de détection des macrostructures argumentatives	127
8.3.4	L'évaluation des résultats	127
8.4	Les résultats	128
8.4.1	Les caractéristiques importantes de chaque classe IMRAD	129
8.4.2	Les résultats du modèle de classification de séquences BERT	130
8.4.3	Les résultats des modèles de classification basée sur la catégorisation	134
8.4.4	Les résultats des modèles de classification instanciée	137
8.4.5	Les résultats des modèles de classification linéaire	139
8.4.6	Les résultats des modèles de classification arborescente	142
8.4.7	Les résultats des modèles de classification neuronale	144
8.4.8	La détection de l'appartenance des cas indéterminés aux macrostructures IMRAD.....	147
8.5	Conclusion	148
CHAPITRE 9 La détection des énoncés de thèse principale		154
9.1	Introduction	154
9.2	Contexte et état de l'art.....	155
9.2.1	Approche fondée par le positionnement	157
9.2.2	Approche fondée par les marqueurs	157
9.2.3	Approche fondée par l'analyse des structures syntaxiques ou sémantiques	159
9.3	Les méthodes	160
9.3.1	La constitution du corpus	161
9.3.1.1	La complétion des données textuelles	163
9.3.2	Le décodage du corpus	164
9.3.3	Les méthodes de détection de la thèse principale	165
9.3.3.1	L'identification de la thèse principale dans le corps du texte à l'aide de l'énonciation de la thèse principale dans le résumé.	165
9.3.3.2	L'identification de la thèse principale à l'aide de l'emplacement	166
9.3.3.3	L'identification de la thèse principale à l'aide de connecteurs de discours	166
9.3.3.4	L'identification de la thèse principale à l'aide de propriétés syntaxiques ou sémantiques... ..	166
9.3.4	L'évaluation des résultats	167
9.4	Les résultats	170
9.4.1	Les résultats de l'identification de la thèse principale dans le corps du texte à l'aide de l'énonciation de la thèse principale dans le résumé.....	172
9.4.2	Les résultats de l'identification de la thèse principale à l'aide de propriétés syntaxiques ou sémantiques	177
9.4.2.1	Les résultats de l'identification de la thèse principale à l'aide du titre	178
9.4.2.1.1	Les résultats des modèles fondés par le titre appliqués au résumé	178
9.4.2.1.2	Les résultats des modèles fondés par le titre appliqués à la discussion	180
9.4.2.2	Les résultats de l'identification de la thèse principale à l'aide du résumé.....	181
9.4.2.2.1	Les résultats des modèles fondés par le résumé appliqués au résumé	182
9.4.2.2.2	Les résultats des modèles fondés par le résumé appliqués à la discussion	183
9.4.2.3	Les résultats de l'identification de la thèse principale à l'aide de l'introduction	185
9.4.2.3.1	Les résultats des modèles fondés par l'introduction appliqués au résumé	185

9.4.2.3.2	Les résultats des modèles fondés par l'introduction appliqués à la discussion	187
9.4.2.4	Les résultats de l'identification de la thèse principale à l'aide de la discussion.....	188
9.4.2.4.1	Les résultats des modèles fondés par la discussion appliqués au résumé.....	189
9.4.2.4.2	Les résultats des modèles fondés par la discussion appliqués à la discussion.....	190
9.4.3	Les résultats de l'identification de la thèse principale à l'aide de l'emplacement.....	192
9.4.3.1	Les résultats des modèles fondés par l'emplacement appliqués au résumé.....	193
9.4.3.2	Les résultats des modèles fondés par l'emplacement appliqués à la discussion.....	195
9.4.4	Les résultats de l'identification de la thèse principale à l'aide de connecteurs de discours.....	197
9.4.4.1	Les résultats des modèles fondés par les connecteurs de discours appliqués au résumé	197
9.4.4.2	Les résultats des modèles fondés par les connecteurs de discours appliqués à la discussion 199	
9.4.5	Les résultats de la combinaison des approches les plus performantes.....	200
9.4.5.1	Les résultats du modèle ensembliste appliqué au résumé	201
9.4.5.2	Les résultats du modèle ensembliste appliqué à la discussion	202
9.5	Conclusion	204
CHAPITRE 10 Conclusion et perspectives.....		206
10.1	Introduction	206
10.2	Résumé de la méthode	207
10.2.1	Les modèles computationnels de détection de l'appartenance des contextes de citation aux macrostructures IMRAD.....	208
10.2.2	Les modèles computationnels de détection des thèses principales	209
10.3	La synthèse des résultats	210
10.3.1	La synthèse de la détection de l'appartenance des contextes de citation aux macrostructures 210	
10.3.2	La synthèse de la détection des thèses principales.....	213
10.4	Les faiblesses et les limites de la recherche.....	216
10.5	Les avenues et recherches ultérieures.....	219
CONCLUSION		230
ANNEXE A Modèle révisé de la RST.....		234
ANNEXE B PDTB3 : Connecteur explicite et leur sens.....		251
APPENDICE A Les thèses principales		263
APPENDICE B Les connecteurs de discours contenus dans les énoncés de thèse principale du résumé ..		274
APPENDICE C Les connecteurs de discours contenus dans les énoncés de thèse principale de la discussion 275		
BIBLIOGRAPHIE.....		277

LISTE DES FIGURES

Figure 3.1 Modèle rhétorique de Wigmore	50
Figure 3.2 Modèle épistémologique de Toulmin	54
Figure 3.3 Cadre d'argumentation	56
Figure 3.4 Modèle communicationnel	58
Figure 3.5 Modèle de Horn	60
Figure 3.6 Modèle RST.....	61
Figure 4.1 Diagramme de classes	69
Figure 6.1 Organisation des macrostructures IMRAD.....	92
Figure 6.2 Exemple de relation d'un pseudotexte structuré IMRAD	94
Figure 6.3 Exemple de structures étendues.....	95
Figure 7.1 Valeurs de la fonction de perte (<i>loss</i>) à différentes étapes (<i>steps</i>) de l'affinage pour le sous-ensemble d'entraînement (<i>train</i>) et pour le sous-ensemble d'évaluation (<i>loss</i>).....	132
Figure 7.2 Valeurs de F-mesure moyenne pondérée (<i>weighted_avg</i>) à différentes sous-étapes de l'affinage pour le sous-ensemble d'entraînement (<i>train</i>) et le sous-ensemble d'évaluation (<i>loss</i>)	133
Figure 8.1 Nombre de phrases contenues dans les résumés.....	171
Figure 8.2 Nombre de phrases contenues dans les discussions.....	172
Figure 8.3 Distribution des positions relatives des énoncés de thèse principale dans les résumés.....	193
Figure 8.4 Distribution des positions relatives des énoncés de thèse principale dans les discussions	195
Figure 8.5 Précision, rappel et F-mesure des top- <i>n</i> connecteurs sélectionnés au moyen de la précision et conjointement appliqués sur le résumé.....	198
Figure 8.6 Précision, rappel et F-mesure des top- <i>n</i> connecteurs sélectionnés au moyen de la précision et conjointement appliqués sur la discussion.....	199

LISTE DES TABLEAUX

Tableau 4.1 Shèmes argumentatifs d’Hastings	46
Tableau 7.1 Mots de coefficient est plus élevée pour les macrostructures IMRAD	129
Tableau 7.2 Valeurs de précision, de rappel, de F-mesure (F1) et de support de chaque classe IMRAD du modèle BERT optimal sur les données d’évaluation	133
Tableau 7.3 Valeurs de précision, de rappel, de F-mesure (F1) et de support de chaque classe IMRAD du modèle BERT optimal sur les données d’apprentissage.....	134
Tableau 7.4 Performance des modèles de classification basée sur la catégorisation appliquées aux diverses représentations vectorielles exprimée en termes de F-mesure pour chacune des classes (notées « F1 (D) » pour discussion, « F1 (I) » pour introduction, « F1 (M) » pour méthodes et « F1 (R) » pour résultats), d’exactitude (notée « Exactitude »), de F-mesure moyenne (notée « F1 (moy.) ») et de F-mesure pondérée (notée « F1 (pond.) »)	135
Tableau 7.5 Performance des modèles de classification instanciée appliquées aux diverses représentations vectorielles exprimée en termes de F-mesure pour chacune des classes (notées « F1 (D) » pour discussion, « F1 (I) » pour introduction, « F1 (M) » pour méthodes et « F1 (R) » pour résultats), d’exactitude (notée « Exactitude »), de F-mesure moyenne (notée « F1 (moy.) ») et de F-mesure pondérée (notée « F1 (pond.) »)	137
Tableau 7.6 Performance des modèles de classification linéaire appliquées aux diverses représentations vectorielles exprimée en termes de F-mesure pour chacune des classes (notées « F1 (D) » pour discussion, « F1 (I) » pour introduction, « F1 (M) » pour méthodes et « F1 (R) » pour résultats), d’exactitude (notée « Exactitude »), de F-mesure moyenne (notée « F1 (moy.) ») et de F-mesure pondérée (notée « F1 (pond.) »)	139
Tableau 7.7 Performance des modèles de classification arborescente appliquées aux diverses représentations vectorielles exprimée en termes de F-mesure pour chacune des classes (notées « F1 (D) » pour discussion, « F1 (I) » pour introduction, « F1 (M) » pour méthodes et « F1 (R) » pour résultats), d’exactitude (notée « Exactitude »), de F-mesure moyenne (notée « F1 (moy.) ») et de F-mesure pondérée (notée « F1 (pond.) »)	142
Tableau 7.8 Performance des modèles de classification neuronale appliquées aux diverses représentations vectorielles exprimée en termes de F-mesure pour chacune des classes (notées « F1 (D) » pour discussion, « F1 (I) » pour introduction, « F1 (M) » pour méthodes et « F1 (R) » pour résultats), d’exactitude (notée « Exactitude »), de F-mesure moyenne (notée « F1 (moy.) ») et de F-mesure pondérée (notée « F1 (pond.) »)	144
Tableau 7.9 Résultat de prédiction du modèle BERT_imrad sur les cas indéterminés.....	147
Tableau 8.1 Exemples de relations discursives organisées en trois niveaux typologiques et exemplifiées par des connecteurs.....	158

Tableau 8.2 Répartition des types de composantes et d'annotations du corpus AbstrCT en fonction du thème principal.....	162
Tableau 8.3 Exemple tiré du corpus AbstrCT	163
Tableau 8.4 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) d'un modèle parfait ...	169
Tableau 8.5 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) d'un modèle de base pour l'identification de l'énoncé de thèse principale dans le résumé.....	177
Tableau 8.6 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) d'un modèle de base pour l'identification de l'énoncé de thèse principale dans la discussion	178
Tableau 8.7 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_T_DBW	179
Tableau 8.8 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_T_OAI ...	179
Tableau 8.9 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_T_DBW	180
Tableau 8.10 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_T_OAI.	181
Tableau 8.11 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_R_DBW	182
Tableau 8.12 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_R_OAI.	183
Tableau 8.13 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_R_DBW	184
Tableau 8.14 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_R_OAI.	184
Tableau 8.15 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_I_DBW	186
Tableau 8.16 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_I_OAI..	186
Tableau 8.17 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_I_DBW	187
Tableau 8.18 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_I_OAI..	188
Tableau 8.19 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_D_DBW	189
Tableau 8.20 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_D_OAI.	190
Tableau 8.21 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_D_DBW	190

Tableau 8.22 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_D_OAI	191
Tableau 8.23 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_Début.	194
Tableau 8.24 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_Alt-début	194
Tableau 8.25 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_Fin.....	196
Tableau 8.26 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_Alt-fin-début	196
Tableau 8.27 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_Ensemble_1	201
Tableau 8.28 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_Ensemble_2	202
Tableau 8.29 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_Ensemble_1	202
Tableau 8.30 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_Ensemble_2	203

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

AA	Apprentissage actif
Abox	Assertional box
AbstrCT	Abstract randomized controlled trial corpus
ACL	Attributive concept langage
Adam	Adaptive moment estimation
AM	Apprentissage machine
Arg	Argument
ATP	Adenosine 5'-triphosphate
Att	Attack
AUC	Area under the curve
BBC	British broadcasting corporation
BERT	Bidirectional encoder representations from transformers
BERT_base	BERT de base
BERT_imrad	BERT de base affiné sur la tâche de détection de l'appartenance des contextes de citation aux macrostructures IMRAD
BERT_large	BERT large
BERT_mpnet	BERT de base affiné sur une tâche MPNet
BERT_nli_stsb	BERT large affiné sur une tâche de NLI et STS
B-I	Billroth I
BONG	Bag of n-gram

BOW	Bag of words
BPE	Byte-Pair Encoding
CA	Carbonic anhydrase
CapsNet	Capsule neural network
CARA	Composante argumentative et relation argumentative
CARS	Create a research space
CBOW	Continuous bag of words
C-CRNE	Clustering-based method for collecting reliable negative examples
ChatGPT	Chat generative pre-trained transformer
CNN	Convolutional neural network
CV	Cross-validation
D	Discussion
D_Alt-fin-début	Modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen de l'alternance de la position à la fin et au début de la structure
D_CD_9	Modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen des 9 connecteurs de discours les plus précis
D_D_OAI	Modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen de la discussion et de la représentation textuelle d'OpenAI (text-embedding-ada-002)
D_D_DBOW	Modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen de la discussion et de la représentation textuelle DBOW
D_Ensemble_1	Modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen de D_I_DBOW (vote prépondérant), D_Alt-fin-début et D_CD_9
D_Ensemble_2	Modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen de D_I_DBOW (vote prépondérant), D_T_DBOW, D_Alt-fin-début et D_CD_9

D_Fin	Modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen de la position à la fin de la structure
D_I_OAI	Modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen de l'introduction et de la représentation textuelle d'OpenAI (text-embedding-ada-002)
D_I_DBOW	Modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen de l'introduction et de la représentation textuelle DBOW
D_R_OAI	Modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen du résumé et de la représentation textuelle d'OpenAI (text-embedding-ada-002)
D_R_DBOW	Modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen du résumé et de la représentation textuelle DBOW
D_T_OAI	Modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen du titre et de la représentation textuelle d'OpenAI (text-embedding-ada-002)
D_T_DBOW	Modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen du titre et de la représentation textuelle DBOW
DBOW	Distributed bag of words
DILCA	Reliable negatives are selected based on a trainable distance measure distance learning for categorical attributes
DM	Distributed memory
DNF	Disjunctive normal form
Doc2vec	Document to vector
DOMIF	Domaine d'information
eval	Evaluation
Evi	Evidence
Extra-S	External to sentence
F1-tk	F-measure top-k

FT	Fouille de texte
GAN	Adversarial generative network
GPU	Generative positive-unlabeled
HCL	Health-related quality of life
HN	Humanités numériques
I	Introduction
Idf	Inverse document frequency
IMRAD	Introduction, methods, results and discussion
InTeReC	In-text reference corpus
Intra-S	Internal to sentence
IRI	Internationalized Resource Identifier
KNN	k-nearest neighbors
LATAO	Lecture et analyse de texte assistées par ordinateur
LD	Logique de description
LIME	Local interpretable model-agnostic explanations
LOOCV	Leave one out cross-validation
M	Méthodes
MajCl	Major claim
MDLF	Multidictionnaire de la langue française
MLP	Multilayer perceptron
MPNet	Masked and permuted language modeling network

MR	Méthodes et résultats
NB	Naive Bayes
NLI	Natural language inference
NLP	Natural language processing
NSCLC	Non-small-cell lung cancer
OAI	Modèle text-embedding-ada-002 d'OpenAI
OWL	Ontology Web langage
PDTB	Penn discourse treebank
PE	Positive enlargement
PEBL	Positive exemple based learning
PLOS	Public library of science
PNLH	Positive examples and negative examples labeling heuristic
P-tk	Précision top-k
QOL	Quality of life
R	Résultats
R_Alt-début-fin	Modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen de l'alternance de la position au début et à la fin de la structure
R_CD_9	Modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen des 9 connecteurs de discours les plus précis
R_D_OAI	Modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen de la discussion et de la représentation textuelle d'OpenAI (text-embedding-ada-002)
R_D_DBOW	Modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen de la discussion et de la représentation textuelle DBOW

R_Début	Modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen de la position au début de la structure
R_Ensemble_1	Modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen de R_Alt-début-fin (vote prépondérant), R_I_DBOW et R_CD_9
R_Ensemble_2	Modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen de R_Alt-début-fin (vote prépondérant), R_I_DBOW, R_D_DBOW et R_CD_9
R_I_OAI	Modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen de l'introduction et de la représentation textuelle d'OpenAI (text-embedding-ada-002)
R_I_DBOW	Modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen de l'introduction et de la représentation textuelle DBOW
R_R_OAI	Modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen du résumé et de la représentation textuelle d'OpenAI (text-embedding-ada-002)
R_R_DBOW	Modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen du résumé et de la représentation textuelle DBOW
R_T_OAI	Modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen du titre et de la représentation textuelle d'OpenAI (text-embedding-ada-002)
R_T_DBOW	Modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen du titre et de la représentation textuelle DBOW
RBF	Radial basis function
RCT	Randomized controlled trial
RD	Résultats et discussion
RDF	Ressource description framework
Regex	Regular expression
ReLU	Rectified linear unit
RNN	Recurent neural network

ROC	Receiver operating characteristic
RST	Rhetorical structure theory
R-tk	Rappel top-k
R-Y	Roux-en-Y
SG	Skip-gram
SGD	Stochastic gradient descent
SHAP	Shapley additive explanations
STS	Semantic textual similarity
Sup	Support
SVC	Support vector machine classifier
SVM	Support vector machine
TAL	Traitement automatique du langage
TALN	Traitement automatique du langage naturel
TBox	Terminological box
TECHNETIUM	Texte court de haut niveau théorique publié dans une revue à comité de lecture
Tf	Term frequency
TfIdf	Term frequency inverse document frequency
TFIPNDF	Term frequency inverse positive-negative document frequency
TLFi	Trésor de la langue Française informatisé
TP	Thèse principale
UE	Unité élémentaire

UML	Unified modeling language
UNIF	Unité d'information
VAE	Variational autoencoder
W3C	World Wide Web Consortium
Word2vec	Word to vector
XML	Extensible markup language

LISTE DES SYMBOLES ET DES UNITÉS

mg Milligramme

RÉSUMÉ

Ce travail de recherche vise à appliquer les méthodes de lecture et d'analyse de texte assistées par ordinateur (LATAO) à des textes de haut niveau théorique en explorant plus particulièrement les dimensions d'assistance à la fouille d'argument (*argument mining*). Il vise le développement d'un système informatique dont le but est d'effectuer la tâche de détection automatique d'arguments. Dans un premier temps, nous présentons les assises théoriques d'un tel système dont, notamment, l'analyse du discours, la lecture et l'analyse de texte assistées par ordinateur, l'argumentation, les modèles computationnels de l'argumentation ainsi que leurs différentes applications. Dans un deuxième temps, nous présentons le système de fouille d'argument de manière à saisir les différents objets informatiques en jeu ainsi que les étapes opératoires comme la définition du corpus, la préparation du corpus, la détection des arguments et la représentation des résultats. Dans un troisième temps, nous explorons plus en détail l'étape de détection des composantes argumentatives et des relations à l'intérieur et entre les arguments en explicitant les différentes représentations et les divers outils informatiques permettant une telle détection. Au terme de cette exploration, une théorie novatrice de l'argumentation est présentée ainsi que diverses hypothèses de recherche à propos de composantes argumentatives importantes telles que les contextes de citation et les thèses principales. À cet effet, deux expériences sont effectuées afin de tester les hypothèses de recherche. Les résultats de la première expérience montrent que la détection automatique de l'appartenance des contextes de citation à une structure argumentative particulière est possible. La combinaison d'une représentation textuelle basée sur « text-embedding-ada-002 » sans affinage et d'un classifieur de type SVM semble donner de très bons résultats, mais l'approche la plus performante apparaît être une combinaison de la représentation BERT de base affinée sur nos données et de la régression logistique avec validation croisée. Les résultats de la deuxième expérience montrent que la détection automatique de la thèse principale est possible. L'approche la plus performante pour la détection de la thèse principale apparaît être un modèle ensembliste combinant une mesure de similitude entre la section « introduction » et les énoncés de la section « discussion », l'emplacement de la thèse principale dans le texte et la présence (ou non) de certains connecteurs de discours. Dans la structure du résumé, l'approche fondée sur l'emplacement s'est avérée plus performante.

Mots-clés : argumentation; fouille d'argument; texte de haut niveau théorique; analyse du discours; Lecture et Analyse de Texte Assistées par Ordinateur (LATAO); Lecture et Analyse Argumentaire de Texte Assistées par Ordinateur (LAATAO).

ABSTRACT

This research aims to apply computer-assisted reading and text analysis (CARAT) methods to high-level theoretical texts by exploring more particularly the dimensions of argument mining assistance. It involves the development of a computer system designed for the automatic detection of arguments. First, the theoretical foundations of such a system are presented, including discourse analysis, computer-assisted text reading and analysis, argumentation, computational models of argumentation, and their various applications. Secondly, the argument mining system is detailed, covering the various computational objects involved and operational steps such as corpus definition, corpus preparation, argument detection, and results representation. Thirdly, the detection phase of argumentative components and relationships within and between arguments is explored in greater detail, clarifying the different representations and computational tools enabling such detection. This exploration culminates in the presentation of an innovative theory of argumentation and various research hypotheses concerning significant argumentative components like citation contexts and major claim. To this end, two experiments are conducted to test the research hypotheses. Results from the first experiment indicate that automatic detection of citation contexts' belonging to a specific argumentative structure is feasible. A combination of “text-embedding-ada-002” text representation without further training and an SVM classifier yields good results, but the most effective approach appears to be a combination of basic BERT representation trained on our data and a cross-validated logistic regression classifier. Results from the second experiment show that automatic detection of the main thesis is possible. The most effective approach for detecting the main thesis appears to be an ensemble model that combines a measure of similarity between the “introduction” section and the statements in the “discussion” section, the placement of the main thesis within the text, and the presence (or absence) of certain discourse connectors. In the abstract section, the approach based on the placement proved to be more effective.

Keywords : argumentation; argument mining; high-level theoretical text; discourse analysis; Computer-Assisted Reading and Analysis of Text (CARAT); Computer-Assisted Reading and Argument Analysis of Text (CARAAT).

INTRODUCTION

Ce projet doctoral s'inscrit dans la mouvance des humanités numériques avec une application des méthodologies de Lecture et d'Analyse de Texte Assistées par Ordinateur (LATAO) (Lebart et Salem, 1994 ; Meunier, 1997 ; Moretti et Piazza, 2007), mais appliquées à des textes de haut niveau théorique. La LATAO est un programme de recherche en pleine effervescence et un de ces objectifs est de décrire la lecture et l'analyse de texte au moyen d'une décomposition de cette tâche en sous-tâches, et d'expliquer par modélisation informatique toutes celles qui sont computationnellement modélisables. La lecture et l'analyse de texte sont des tâches importantes de l'activité de recherche et leur modélisation computationnelle permet le développement d'algorithmes utiles à l'assistance du travail de recherche en philosophie et en science. Ce paradigme scientifique se situe à l'intersection de plusieurs disciplines, notamment les sciences cognitives, la fouille de texte, la linguistique, l'informatique, et les sciences de l'information. Les outils informatiques utilisés en LATAO peuvent inclure des logiciels de traitement de texte, des bases de données textuelles, des systèmes de reconnaissance de la parole, et des programmes d'analyse sémantique ou syntaxique. Elle permet une analyse à la fois quantitative (par exemple, la fréquence des mots, la longueur des phrases, les motifs syntaxiques) et qualitative (comme l'analyse de thèmes, de motifs, ou de style). Avec l'avènement des grands ensembles de données (*Big Data*), la LATAO devient de plus en plus pertinente, car les chercheurs peuvent analyser de vastes ensembles de données textuelles, y compris des livres numériques, des articles académiques, des blogues, des tweets, etc. La LATAO a des applications dans de nombreux domaines, comme la linguistique, la littérature, l'histoire, la sociologie, la philosophie, la médecine, etc. La recherche utilise la LATAO pour modéliser computationnellement la tâche de lecture et d'analyse de textes, ce qui permet d'explicitier les méthodes ainsi que la reproduction de leurs applications. Elle donne lieu à divers modèles et méthodes permettant d'exploiter la puissance de calcul des ordinateurs pour analyser de grandes quantités de textes plus efficacement et précisément qu'il ne serait possible manuellement. Plusieurs chercheurs l'utilisent, notamment pour l'extraction de connaissances (Meunier et Nault, 1997), l'exploration de classifieurs s'inspirant de l'approche génétique (Rialle *et al.*, 1998), connexionniste (Meunier *et al.*, 1997) ou statistique (Benhadid *et al.*, 1998), la philologie (Seffah et Meunier, 1997), la classification des textes (Forest et Meunier, 2000), l'apprentissage, notamment par renforcement (Nguyen *et al.*, 2007), le génie logiciel (Biskri et Meunier, 2002), l'analyse de thèmes (Meunier *et al.*, 2005), de concepts (Chartrand *et al.*, 2017 ; Danis *et al.*, 2010 ; Danis et Meunier, 2012 ; Meunier *et al.*, 2011 ; Meunier et Forest, 2008 ; Sainte-Marie *et al.*, 2010, 2011) et de représentations sociales (Chartier et Meunier, 2011) ou artistiques (Chartier

et al., 2019 ; Rompré *et al.*, 2017). Le champ de LATAO évolue constamment avec le développement de nouvelles technologies et méthodes. L'intelligence artificielle et l'apprentissage profond, par exemple, offrent de nouvelles façons d'analyser et d'interpréter les textes. La LATAO est donc un paradigme de recherche dynamique et en évolution, qui capitalise sur les avancées technologiques pour enrichir et approfondir la lecture et l'analyse textuelle dans divers domaines philosophiques et scientifiques. Ce programme de recherche a abouti à plusieurs retombées significatives, la première étant l'élaboration d'un cadre théorique pour la LATAO. Notre contribution à ce domaine s'est manifestée à travers une série d'activités académiques et de collaborations, incluant des séminaires, des écoles d'été, des présentations lors de colloques, la publication d'articles et de chapitres de livres, ainsi que des participations à des projets de recherche variés, et bien sûr, ce travail de doctorat. Dans le cadre spécifique de la LATAO, nos recherches ont contribué à affiner l'analyse de thèmes (Bertoldi *et al.*, 2023 ; Lamirel *et al.*, 2023a, 2023b, 2024 ; Lareau *et al.*, 2023a, 2023b, 2024, s. d. ; Lareau et Malaterre, 2023a, 2023b ; Malaterre *et al.*, 2019, 2020b, 2020c, 2020a, 2021 ; Malaterre et Lareau, 2022, 2023d, 2023a, 2023e, 2023b, 2023c, 2024b, 2024a), de concepts (Chartrand *et al.*, 2016, 2018 ; Lareau, 2016, 2022 ; Lareau *et al.*, 2018 ; Le *et al.*, 2016 ; Pulizzotto *et al.*, 2016, 2016, 2018) et d'arguments (Kaiser *et al.*, à paraître ; Lareau, 2023a, 2023b), comme en témoignent diverses publications réalisées par nous et nos collaborateurs. Une seconde retombée de ce projet de recherche est le développement d'outils computationnels à l'analyse de thèmes, de concepts et d'arguments. Ceux-ci peuvent prendre la forme d'une chaîne de traitements algorithmiques ou cognitifs (correspondant à la suite de sous-tâches de la LATAO) ou prendre la forme de *packages* adaptés à des langages informatiques spécifiques. Plusieurs de nos contributions technologiques sont publiquement accessibles via GitHub (une entreprise de développement et services logiciels) et ont été partagées lors de divers événements académiques et projets de recherche. Une troisième retombée de ce projet de recherche est l'application de la LATAO à des domaines spécifiques tels que la science, la philosophie, l'épistémologie, l'astrobiologie, la statistique, entre autres. Ces contributions scientifiques consistent en la LATAO de *corpora* largement constitués de textes de haut niveau théorique de manière à en dégager les principaux thèmes, concepts et arguments. Lors de ces applications, les données textuelles ont souvent été mises en relation à des données contextuelles comme l'année de publication, l'éditeur, les auteurs.trices, etc., contribuant ainsi à une compréhension plus approfondie des domaines et des corpus étudiés.

Dans notre thèse doctorale, qui constitue une innovation au sein du programme de recherche de la LATAO, nous approfondissons les aspects de l'assistance informatique dans l'analyse d'arguments. Nous nous

concentrons particulièrement sur la fouille d'arguments (*argument mining*), c.-à-d. l'extraction automatique de structures argumentatives à partir de données textuelles non structurées (Peldszus et Stede, 2013a). Ce domaine suscite un intérêt croissant, car il ouvre la voie à l'exploitation des vastes connaissances disponibles dans les bases de données textuelles. Les progrès récents en apprentissage machine ouvrent des perspectives prometteuses non seulement dans les domaines des sciences sociales, économiques et politiques, mais aussi dans les technologies de l'information. L'argumentation est également un enjeu clé en intelligence artificielle, notamment pour sa capacité à relier les modèles computationnels aux modèles d'agents cognitifs, comme l'ont souligné Bench-Capon et Dunne (2007). En philosophie, Brandom (1994) a mis en évidence l'importance de l'inférence, du raisonnement et de l'argumentation dans la conceptualisation. De même, en sciences cognitives, Mercier et Sperber (2011b, 2011a) ont montré la fonction communicationnelle du raisonnement humain. Il est à noter que la fouille d'arguments est un domaine relativement récent, avec de premiers travaux se concentrant sur la détection de zones textuelles argumentatives (Teufel *et al.*, 2009) et des arguments dans les textes légaux (Mochales Palau et Moens, 2011), les revues en ligne (Cabrio et Villata, 2012b) et les débats (Saint-Dizier, 2012). Ce domaine est maintenant une spécialité de la fouille de texte avec ses revues et ses conférences (Grasso *et al.*, 2010). Ces avancées nous serviront de base pour développer l'analyse d'arguments comme spécificité de la LATAO, utilisant des outils spécifiques pour analyser la formulation et la structuration des arguments dans les textes, afin d'enrichir notre compréhension des textes étudiés.

Notre projet doctoral se développe en 10 chapitres. Le premier chapitre – La lecture et l'analyse de texte – pose les bases de notre exploration. Il présente deux paradigmes fondamentaux de l'analyse du discours, à savoir celui externaliste et celui internaliste, et examine en détail les théories structuralistes, formelles et herméneutiques qui façonnent notre compréhension des textes. Il présente également les caractéristiques des textes de haut niveau théorique, mettant en évidence leur structure et leurs objectifs spécifiques.

Le chapitre suivant – La lecture et l'analyse de texte assistées par ordinateur – examine la lecture et l'analyse de texte assistées par ordinateur (LATAO). Il retrace les origines historiques de la LATAO et l'intérêt de la modélisation des processus cognitifs ainsi que de l'interaction entre l'homme et l'ordinateur pour la lecture et l'analyse de texte. Ce chapitre souligne la contribution des diverses disciplines connexes à l'enrichissement de la LATAO et son application à l'analyse argumentaire de textes de haut niveau théorique.

Le troisième chapitre – L'argumentation – se penche sur la nature même de l'argumentation. En commençant par les définitions du dictionnaire, il explore les approches rhétoriques, épistémologiques, consensualistes et communicationnelles de l'argumentation. L'accent est mis sur la manière dont les différentes conceptions de l'argumentation influencent la façon dont nous analysons les textes. Cette exploration permettra de comprendre les fondements de l'argumentation et ses différentes approches.

Le quatrième chapitre – Les modèles computationnels de l'argumentation – franchit une étape supplémentaire en examinant les modèles informatiques qui ont été développés pour représenter et comprendre l'argumentation dans les textes. Il aborde les modèles rhétoriques, épistémologiques, consensualistes et communicationnels, en soulignant leurs applications pratiques pour analyser les arguments présents dans les textes.

Le cinquième chapitre – La modélisation computationnelle de la tâche d'analyse d'argument – se concentre sur les étapes clés impliquées dans l'analyse d'argument. Il explore la décomposition de cette tâche en différentes étapes, allant de la définition du corpus à sa représentation informatique, puis l'analyse des arguments et l'interprétation des résultats.

Le sixième chapitre – Implémentation de la tâche de détection des arguments – examine les diverses représentations vectorielles du texte utilisées dans les méthodes de détection des structures argumentatives. Des approches déductives, inductives et hybrides sont explorées, ainsi que l'évaluation qualitative et quantitative des outils de détection.

Le septième chapitre – Les hypothèses – est dédié à la formulation des hypothèses, en se basant sur les connaissances acquises dans les chapitres précédents à propos de la structure argumentaire des textes de haut niveau théorique. Il aborde la distinction entre les microstructures, les macrostructures et les structures argumentaires étendues, en mettant l'accent sur le rôle structural crucial joué par les contextes de citation et la thèse principale.

Le huitième chapitre – La détection de l'appartenance des contextes de citation aux macrostructures argumentatives – présente les méthodes utilisées pour détecter l'appartenance des contextes de citation aux macrostructures IMRAD (Introduction, Méthodes, Résultats, et Discussion). Différents modèles de classification sont explorés, et leurs résultats sont discutés en détail.

Le neuvième chapitre – La détection de la thèse principale – se concentre spécifiquement sur l'identification des segments textuels exprimant la thèse principale dans les textes de haut niveau théorique. Différentes approches basées sur l'emplacement, les propriétés syntaxiques et sémantiques, ainsi que les connecteurs de discours, sont évaluées et comparées.

Enfin, le dernier chapitre – Conclusion et perspectives – récapitule les principaux résultats et discussions des chapitres précédents. Les limites de la thèse sont également abordées, ainsi que les implications des découvertes pour la recherche future dans le domaine de la lecture et de l'analyse de texte assistées par ordinateur.

Cette thèse doctorale vise à approfondir notre compréhension des méthodes et des approches utilisées dans la lecture et l'analyse de texte, ouvrant ainsi de nouvelles perspectives pour la recherche et l'application pratique dans ce domaine en constante évolution.

CHAPITRE 1

La lecture et l'analyse de texte

Dans ce chapitre, nous entreprenons de définir et d'explorer les concepts clés nécessaires pour comprendre la lecture et l'analyse de texte. Nous commençons par examiner la notion de discours, avant de nous focaliser sur celle de texte. L'analyse du discours est abordée à travers deux paradigmes distincts. D'une part, le paradigme externaliste, qui considère que le sens d'un texte est situé en dehors de celui-ci, intégré dans le discours. D'autre part, le paradigme internaliste, qui soutient que le sens réside au sein même des textes. Ce paradigme internaliste est lui-même décomposé en plusieurs approches théoriques: la théorie structuraliste, qui analyse les structures sous-jacentes du texte; la théorie formelle, qui s'attache aux aspects formels et cognitifs; et la théorie herméneutique, qui se concentre sur l'interprétation et la compréhension des textes. Nous nous penchons ensuite sur les textes de haut niveau théorique, les classifiant selon leur structure en deux catégories principales: les formes longues, qui offrent une exploration détaillée et approfondie, et les formes courtes, qui présentent des idées de manière concise. Ces textes sont également analysés en fonction de leurs objectifs principaux et secondaires, explorant comment ils visent à informer, persuader, ou susciter la réflexion.

1.1 Introduction

L'objectif de cette thèse doctorale est l'analyse des arguments dans le discours de haut niveau théorique. L'analyse d'argument est un cas particulier de l'analyse du discours et, évidemment, toute analyse du discours de haut niveau théorique implique une certaine théorie du discours, notamment du discours de haut niveau théorique. Ce dernier, qui est principalement de nature scientifique ou philosophique, se distingue des autres formes de discours tels que ceux politiques, religieux, littéraires ou esthétiques. Son objectif est de présenter de façon rationnelle et réfléchie une perspective originale sur un sujet faisant l'objet de questionnement, en tenant compte des opinions et des arguments existants et en s'appuyant sur les contributions d'autres auteurs dans le domaine. Subséquemment, l'analyse du discours de haut niveau théorique vise à déceler et comprendre les idées, les thèses ou les doctrines qui y sont exprimées, d'examiner comment elles s'agencent ou sont développées, en particulier à travers des arguments, et cela en examinant la forme linguistique qui le sous-tend, c.-à-d. le texte (Panaccio, 2019). Deux paradigmes théoriques permettent de conceptualiser la relation texte-discours, ainsi que son analyse.

1.2 Le paradigme externaliste de l'analyse du discours

La théorie externaliste présente le discours comme un acte de langage où la notion de texte est vague et inopérante (Guespin, 1971). Le texte y est considéré comme une manifestation concrète du discours, tandis que le discours est vu comme le contexte plus large et dynamique dans lequel le texte prend forme et signification (Rastier, 2005). L'analyse s'intéresse alors aux contextes d'énonciation du discours dont elle vise à découvrir les conditions de production socio-économiques (Angenot, 1985 ; Bakhtine, 1978 ; Pêcheux, 1969), sociohistoriques (Burman et Parker, 2016 ; Foucault, 1969), sociopolitiques (Bart, 1998 ; Berger et Luckmann, 1966 ; Bourdieu, 1998) ou psychologiques (Edwards, 1997 ; Molder et Potter, 2005 ; Potter et Wetherell, 1987).

1.3 Le paradigme internaliste de l'analyse du discours

À l'opposé des théories externalistes, qui avancent que la signification des textes est extérieure à ceux-ci, partiellement du moins, les théories internalistes affirment que le sens réside dans les textes eux-mêmes. Par conséquent, l'analyse des textes est possible et peut se concevoir comme autonome, sans nécessiter de faire appel à des éléments externes. Dans cette perspective, les discours ne sont pas vus comme des entités abstraites ou des forces extérieures influençant les textes, mais plutôt comme des ensembles composés de textes individuels. La signification d'un discours spécifique est donc vue comme émanant des textes qui le composent, chaque texte contribuant à la formation du discours par ses caractéristiques uniques et ses éléments de signification. Le discours, soutient Rastier, se conçoit comme une pluralité, car il n'est pas composé d'autre chose que de ce qui a été dit ou écrit, c.-à-d. des textes :

[L]es textes et les discours (au pluriel !) se trouvent [...] exactement au même niveau ontologique : par exemple, le discours littéraire est fait de tous les textes littéraires ; le texte littéraire n'est pas considéré comme l'énoncé produit par le discours littéraire, et l'étude de ce discours n'est pas censée expliquer les textes qui en relèvent. (Rastier, 2005)

À l'instar de Rastier, si l'on admet que les discours sont des collections de textes au même niveau ontologique, alors le discours en tant que totalité, s'il en est un, correspondrait à l'ensemble de tous les textes. Toutefois, l'analyse effective d'une telle totalité est douteuse, voire impossible, de sorte que la notion de discours se doit d'être appréhendée de manière plurielle. En somme, la théorie internaliste présente le discours comme un contenu énonciatif singulier exprimé dans une forme linguistique complexe – un texte – qui est ainsi une notion opératoire (Galatanu, 2007, 2018). Les théories internalistes, qui comprennent les approches structuraliste, formelle ou cognitive et herméneutique, offrent diverses

façons de modéliser les textes ainsi que leur analyse. Chacune de ces théories propose une conception différente de la manière dont les textes contribuent et façonnent les discours.

1.3.1 La théorie internaliste structuraliste

Une première théorie est structuraliste. Le texte soutenant le discours y apparaît comme une structure langagière complexe constituée de divers niveaux interreliés (Chabrol, 1973 ; Halliday et Hasan, 1976 ; Mellet et Longrée, 2009 ; Rastier, 1972 ; van Dijk, 1972 ; Werlich, 1976). Le structuralisme considère le discours comme un système de signes qui obéit à des règles et des structures sous-jacentes. Cette approche met l'accent sur les relations entre les éléments constitutifs du discours, tels que les phonèmes, les mots, les phrases, les paragraphes, etc. Selon le structuralisme, la signification d'un discours est déterminée par la structure globale du système linguistique auquel il appartient. Le structuralisme cherche à identifier les schémas récurrents et les règles qui organisent le discours. Les structures peuvent être représentées sous forme de diagrammes ou de modèles pour mieux visualiser les relations entre les éléments.

1.3.2 La théorie internaliste formelle ou cognitive

Une deuxième théorie est formelle et cognitive. Le texte est régi par certaines règles contrôlant la formation du discours, lesquelles déterminent les inférences au sein de suites de phrases (Kintsch et van Dijk, 1978), assurent la cohérence de l'ensemble du texte (Ghadessy, 1999), et contrôlent la distribution des contextes (Firth, 1964 ; Harris, 1951) de manière à former des régularités (Bruni *et al.*, 2014 ; Lenci, 2008 ; Turney et Pantel, 2010). D'une part, la théorie formelle du discours se concentre sur les aspects logiques ou mathématiques du langage. Elle cherche à décrire les règles formelles qui régissent la construction des énoncés et la logique de leur argumentation. Cette approche considère le discours comme un système de propositions et d'inférences. D'autre part, la théorie formelle du discours cherche à modéliser et à comprendre les processus cognitifs jouant un rôle dans la compréhension et la production du discours.¹ La théorie formelle du discours s'intéresse particulièrement à la logique et aux règles formelles qui gouvernent la construction des énoncés et la validité des arguments. Elle examine comment

¹ Par exemple, la compréhension et la production du discours peuvent impliquer la perception, la mémoire, le raisonnement, l'attention, le langage, la résolution de problèmes, etc. Lorsque nous nous engageons avec le discours, ces processus cognitifs entrent en jeu pour comprendre le sens, traiter l'information et évaluer les arguments.

les individus utilisent la logique et les opérations mathématiques pour évaluer la cohérence et la validité des propositions dans le discours.

1.3.3 La théorie internaliste herméneutique

Une troisième théorie est herméneutique. Le discours est un objet épistémique d'interprétation (Apel, 1975 ; Derrida, 1967a, 1967b ; Dilthey, 1992 ; Føllesdal, 1979 ; Gadamer, 1982, 2018 ; Grondin, 1993 ; Habermas, 1981, 1987 ; Ricœur, 1969, 1986 ; Taylor, 1992). Le texte est entendu comme une forme d'expression linguistique de cet objet d'interprétation qui peut être marquée par la subjectivité, la rationalité, la communication et la culture. Cette approche s'appuie sur des méthodes d'analyse textuelle telles que l'examen des symboles, des références culturelles et des structures narratives. Elle cherche à mettre en évidence les différents niveaux de signification et à comprendre comment le discours est influencé par des facteurs sociaux, historiques et culturels. Lorsqu'appliquée à l'analyse des textes de haut niveau théorique, la relation texte-discours permet l'exploration matérielle des idées, des thèses et des doctrines puisque celles-ci ne sont en fait accessibles que par leurs formes linguistiques textuelles. Ces objets d'investigation sont en quelque sorte des constituants sémantiques accessibles via des thèmes, des concepts et des arguments.

1.4 Le texte de haut niveau théorique

Le discours de haut niveau théorique s'exprime au moyen de la forme matérielle qu'est le texte. Les textes de haut niveau théorique peuvent être catégorisés selon différents types. Selon Aston (1977), les types discursifs se distinguent selon leurs fonctions illocutoires que l'on conçoit comme l'objectif d'un acte de langage. De plus, Werlich (1976) postule que les types de texte se distinguent en fonction de leurs caractéristiques linguistiques telles que le type de phrase, l'expansion de clause, la séquence, la structuration ou la temporalité. Autrement, certains chercheurs comme Smith (1985) et Biber (1989) soutiennent que les caractéristiques intrinsèques aux textes ne sont pas indépendantes de celles extrinsèques comme la fonction illocutoire du texte. De plus, notons que plusieurs théoriciens de la typologie textuelle soutiennent la porosité des limites entre les divers types, notamment parce que les objectifs peuvent être multiples (de Beaugrande et Dressler, 1981 ; Hatim, 1991 ; Parret, 1987 ; Smith, 1985 ; Youga, 1989).

En résumé, la structure textuelle et la fonction illocutoire sont des éléments distinctifs de certains types de discours, y compris ceux de haut niveau théorique. Cependant, l'identification de l'appartenance d'un

texte à un discours spécifique peut être complexe, en raison de l'objectif parfois implicite derrière sa production et la variabilité des structures. Certains chercheurs, comme Fishelov (1991) et Fowler (1982), proposent une approche inspirée de la notion wittgensteinienne d'air de famille, qui offre plus de flexibilité par rapport aux définitions traditionnelles, et qui peut mieux rendre compte du phénomène linguistique en jeu. Selon cette approche, les textes appartenant à un discours particulier sont liés par divers moyens, sans qu'aucun élément ne soit nécessaire et suffisant à tous :

Representations of a genre may then be regarded as making up a family whose septs and individual members are related in various ways, without necessarily having any single feature shared in common by all. (Fowler, 1982, p. 41)

Ainsi, pour identifier l'appartenance d'un texte au discours de haut niveau théorique, l'analyse se concentrera sur sa fonction illocutoire explicite dans le texte et sa structure interne spécifique. Dans un premier temps, nous aborderons la structure du texte de haut niveau théorique et, ensuite, son principal objectif.

1.4.1 Les grands types structurels de textes de haut niveau théorique

1.4.1.1 La structure textuelle longue du texte de haut niveau théorique

Traditionnellement, le texte de haut niveau théorique prend une forme longue (le livre par exemple), mais celui-ci tend de plus en plus vers une forme plus courte, notamment celui de l'article. Comme le souligne Lewenstein (2007), il est aujourd'hui commun de circonscrire la science à ce qui se publie sous forme d'article dans les journaux scientifiques ou philosophiques. Or, soutient l'auteur, même si le livre joue un rôle moins central que jadis en science, celui-ci contribue de manière importante au développement de la science, notamment en ce qui a trait au développement intellectuel de la science elle-même, au recrutement des gens en science, à la culture scientifique et aux débats publics. De plus, il est un fait que les auteurs d'articles scientifiques ou philosophiques citent des livres, ce qui montre que leur discours s'inscrit dans un discours plus large qui inclut des textes longs pouvant prendre entre autres la forme du livre.

1.4.1.2 La structure textuelle courte du texte de haut niveau théorique

Le court texte de haut niveau théorique peut prendre la forme de chapitre de livres, de blogue, d'article de conférence professionnelle ou d'article de revue spécialisée. La forme privilégiée par la communauté scientifique et philosophique est celle de l'article publié dans les revues à comité de lecture qui, selon

Thyer (2008), restera le principal moyen de diffusion des résultats de recherche dans un avenir prévisible, car comparativement aux autres moyens de diffusion, celui de l'article de revue est plus prestigieux et mieux reconnu au sein de la communauté philoscientifique en raison du processus de sélection rigoureux (examen à l'aveugle par les pairs, commentaires et suggestions pour améliorer l'article), la facilité d'accès (via Internet), la rapidité de diffusion et son influence relativement élevée dans les décisions de promotion et de titularisation au sein du milieu académique.

1.4.2 Les grands objectifs des textes de haut niveau théorique

L'objectif d'un texte de haut niveau théorique est de fournir une perspective rationnelle et réfléchie sur un sujet de débat, une perspective unique sur un problème en questionnement, en se positionnant par rapport au discours existant, et en s'appuyant sur les contributions d'autres auteurs. Cet objectif se réalise en défendant une ou plusieurs thèses, ancrées dans un champ d'investigation spécifique, à l'aide de sources textuelles et justifiées par deux types d'approches, donnant lieu à des textes différents. La première approche, plus courante en philosophie, est exploratoire et met l'accent sur l'analyse, tandis que la seconde, privilégiée en science, est résolutive et se concentre sur la résolution argumentative. La distinction entre science et philosophie n'est pas clairement délimitée, mais ces domaines tendent à se démarquer par des questionnements distincts. Le même principe s'applique à l'intérieur de ces domaines, ce qui donne lieu à des disciplines philosophiques distinctes (épistémologie, éthique, philosophie de l'esprit, etc.) ainsi que diverses sciences (physique, chimie, biologie, sciences humaines, mathématiques, etc.). Comme le souligne John Searl (1999, p. 1), les problèmes philosophiques présentent généralement trois caractéristiques particulières. D'abord, ils portent souvent sur des cadres conceptuels abstraits plutôt que sur des questions spécifiques à l'intérieur de ces cadres. Ensuite, ce sont des questions pour lesquelles il n'existe pas de méthode de résolution généralement acceptée, qu'elle soit formelle ou empirique. Enfin, ils impliquent souvent des enjeux conceptuels non résolus. Subséquemment, l'approche utilisée est souvent exploratoire et met en jeu des méthodes d'exploration, notamment analytiques. En revanche, un problème philosophique peut se transformer en un problème scientifique lorsqu'une méthode de résolution scientifique est adoptée. Cette transition s'opère lorsqu'un questionnement philosophique, initialement abstrait et théorique, se concrétise en une problématique expérimentale ou empirique, permettant ainsi sa résolution par des méthodes scientifiques. Le cas échéant, l'approche est résolutive et met en jeu des méthodes de résolution, notamment argumentatives.

1.4.2.1 L'approche exploratoire et le texte de type analytique

L'approche exploratoire vise à parcourir les limites conceptuelles d'un problème en questionnement afin de mieux le circonscrire ou d'en clarifier la nature. Traditionnellement, il existe plusieurs méthodes permettant une telle exploration. Les plus connues sont les méthodes herméneutique (Gadamer, 2018 ; Ricœur, 1986), déconstructiviste (de Man, 1982 ; Derrida, 1988) et analytique. La philosophie analytique est une tradition qui commence avec Frege, se poursuit avec Carnap, Wittgenstein, Russell, Ryle, Quine, Strawson, et qui se présente aujourd'hui sous plusieurs variantes (voir Glock, 2008). Une caractéristique importante de l'approche analytique est la visée d'une approche plus scientifique des questions philosophiques. Elle est sans doute la méthode exploratoire la plus utilisée par la communauté philoscientifique et donne lieu à des textes de type analytique. Habituellement, ces derniers commencent avec une question de recherche sur laquelle on ne prend pas immédiatement position, car le but sous-jacent en est un d'exploration et d'évaluation. L'énoncé de la thèse principale dans un article analytique est en quelque sorte dynamique, car il peut évoluer au fil de la recherche présentée par le texte.

1.4.2.2 L'approche résolutive et le texte de type argumentatif

L'approche résolutive vise la persuasion, la résolution d'une problématique ou d'un questionnement. Selon Aston (1977), la fonction illocutoire du texte de type argumentatif est de convaincre. D'après Kummer (1972), le texte argumentatif vise à résoudre un problème pouvant se concevoir comme l'assomption que le lecteur conçoit inadéquatement un certain état du monde, ce que le texte vise à remédier.² Autrement, Ram (1991) soutient que le but sous-jacent de la production et la consommation de ce type de texte est de résoudre une question. Sous sa forme courte de l'article publié dans les revues à comité de lecture, le texte de type argumentatif répond habituellement à une structure fonctionnelle standardisée. Cette structure argumentative textuelle est souvent nommée IMRAD (de l'anglais *Introduction, Methods, Results, And Discussion*). À l'instar de Pontille (2007), on peut concevoir l'introduction comme une section où on propose les principales hypothèses de recherche, la méthodologie comme une section où on présente les matériaux et les méthodes permettant l'investigation de ces

² Notons que cet objectif de concevoir adéquatement un problème est partagé avec l'approche exploratoire. La tradition analytique visant à « penser clairement à propos de » (*thinking clearly about ...*) quelque chose exemplifie très bien cela en ce que l'objectif, habituellement explicite dans le titre, est de conceptualiser adéquatement quelque chose au terme d'une exploration conceptuelle.

hypothèses, les résultats comme l'aboutissement de l'investigation et la discussion comme un retour sur les hypothèses de départ, l'évaluation du travail effectué et ses implications :

Le dispositif textuel IMRAD constitue ainsi une expression particulière de la preuve. Il procède par une exposition linéaire des opérations de recherche. Les divers éléments empiriques et théoriques sont agencés dans un enchaînement logique : la partie « résultats » est l'aboutissement de l'investigation décrite dans la section « matériels et méthodes », et elle corrobore les hypothèses qui sont posées dans l'« introduction » [...] À ce stade de l'argumentation, les auteurs sont alors en mesure d'évaluer et d'interpréter les implications de leur travail. La section « discussion » est réservée à cet effet. Ils y comparent leurs résultats avec les hypothèses avancées dans l'introduction. Ils discutent également avec les travaux d'autres chercheurs pour renforcer leurs propres conclusions, ou spécifier en quoi leurs résultats s'inscrivent en faux par rapport à des études antérieures. Ce faisant, ils identifient les implications pratiques et théoriques de leur analyse. Ils peuvent ainsi faire des inférences quant à leur propre travail, et ouvrir des perspectives de futures recherches. (Pontille, 2007, p. 237)

Robert A. Day (1989) attribue la première utilisation de la structure IMRAD à Louis Pasteur en 1876 dans ces « Études sur la bière ». À cette époque, les travaux scientifiques sont largement publiés sous forme d'essais ou de traités. La structure IMRAD a commencé à être adoptée par les scientifiques autour des années 1940 et est rapidement devenue le format dominant pour les articles de recherche dans la majorité des principales revues scientifiques à la fin des années 1970 (Wu, 2011). Un élément clé de l'adoption rapide de ce format par la communauté scientifique est l'adoption en 1972 d'une norme nationale américaine pour la préparation des documents scientifiques (ANSI Z39.16-1972) par la National American Standards Institute (ANSI). En médecine par exemple, l'adoption du format IMRAD passe de 0% en 1935 à 100% en 1985 pour quatre (4) grandes revues du domaine (Sollaci et Pereira, 2004).³ Selon Suppe (1998), la structure IMRAD est nécessaire en science dans la mesure où l'espace de publication est limité et où chaque élément textuel doit contribuer à l'objectif principal du texte qui est d'établir une ou plusieurs thèses principales :

Scientific articles exemplify standard functional units constraining argumentative structures. Severe space limitations demand every paragraph and illustration contribute to establishing the paper's claims. Philosophical testing and confirmation models should take into account each paragraph, table, and illustration. (Suppe, 1998, p. 381)

³ Le corpus d'étude était composé des revues *British Medical Journal*, *Journal of the American Medical Association (JAMA)*, *The Lancet*, et *New England Journal of Medicine*. Notons que 100% des articles de JAMA sont au format IMRAD dès 1975.

Certains textes publiés dans les revues à comité de lecture sont composés de deux structures parallèles qui répondent à une même logique de justification d'une interprétation particulière des données de recherche, soit une structure discursive et une structure graphique :

When papers involve substantial recourse to graphics there typically are distinct textual and graphical argumentative structures which, while different, evidence the same basic logic of brokering among competing interpretations of data or results. (Suppe, 1998, p. 403)

Spécifions que l'auteur ne réfute pas la thèse selon laquelle les textes scientifiques peuvent se conformer aux schémas épistémologiques classiques⁴, notamment celui hypothético-déductif⁵, celui bayésien⁶ et celui de l'inférence à la meilleure explication⁷, mais il postule un schéma épistémologique additionnel, lequel serait fondé par une meilleure interprétation des résultats par rapport à d'autres interprétations possibles :

Scientific papers from diverse disciplines display a common organizational structure and exhibit similar argumentative structures. Once a scientific paper has located its place in intellectual space (the discipline and domain to which it intends to contribute) and presented its results, the main business of the paper is putting forth a favored interpretation of the results and impeaching competing interpretations. (Suppe, 1998, p. 403)

⁴ L'argument de Suppe (1998) contre la thèse selon laquelle les textes scientifiques se conforment aux schémas épistémologiques classiques consiste à montrer qu'un cas paradigmatique en science – le texte de Morgan à propos des plaques tectoniques – ne se conforme pas aux schémas épistémologiques classiques. De plus, il remarque que la majorité des textes scientifiques qu'il a lus ne se conforment pas. Subséquemment, on ne peut pas conclure que tous les textes scientifiques ne se conforment pas aux schémas épistémologiques classiques.

⁵ Le schéma épistémologique hypothético-déductif repose sur la formulation d'hypothèses explicatives, puis de déduire des prédictions vérifiables à partir de ces hypothèses. Ces prédictions peuvent ensuite être testées empiriquement pour évaluer la validité des hypothèses. Si les prédictions sont confirmées, cela renforce la confiance envers les hypothèses, mais si les prédictions sont réfutées, cela remet en question les hypothèses. Le schéma hypothético-déductif suit donc une logique déductive pour évaluer la validité des hypothèses.

⁶ Le schéma épistémologique bayésien repose sur le théorème de Bayes permettant de mettre à jour nos croyances en fonction de nouvelles informations. Cette approche débute avec des croyances initiales (*priors*) sur la probabilité d'une hypothèse donnée. Lorsque de nouvelles données sont disponibles, le théorème de Bayes est utilisé pour calculer les probabilités postérieures (*posteriors*) des hypothèses. Les probabilités postérieures sont mises à jour en combinant les connaissances initiales avec les nouvelles données. L'approche bayésienne intègre ainsi une logique probabiliste pour évaluer et mettre à jour des croyances.

⁷ Le schéma épistémologique de l'inférence à la meilleure explication met en jeu la recherche de l'explication la plus plausible ou la meilleure pour un ensemble de données ou de faits. Cette approche cherche à identifier l'explication qui rend le mieux compte des données disponibles et qui est la plus cohérente, complète et économique. Elle implique de peser différentes explications concurrentes et de les comparer en termes de leur capacité à expliquer les faits de manière satisfaisante.

Le schéma épistémologique d'une inférence à la meilleure interprétation⁸ est non ampliatif et met en relation des connaissances préalables, des données ou des résultats qui ensemble supportent une conclusion, et cette conclusion est en fait une interprétation des résultats qui s'oppose à d'autres interprétations concurrentes :

Interpretative arguments are non-ampliative. The evidential basis for the paper's claims consists of the paper's results or data augmented by background knowledge. Interpretative arguments have as their focus finding a correct interpretation of the results, not a strengthening of the evidential basis. In effect the situation is that the results are evidence for something and the argumentative task of the paper is to determine what that something is - calibrating one's conclusions to one's findings. (Suppe, 1998, p. 403)

Ainsi, l'objectif sous-jacent le texte publié dans une revue à comité de lecture est de résoudre une question de recherche et cette résolution se déploie habituellement selon une même structure argumentative générale qui se présente de manière linéaire en réponse à quatre sous-objectifs principaux. Le premier est de positionner le texte ou la question de recherche dans le domaine auquel on entend offrir une contribution. Le second est de présenter les matériaux et méthodes qui seront utilisés. Le troisième est de présenter les résultats de la recherche. Le dernier sous-objectif, dépendamment du schéma épistémologique adopté, est de confirmer ou infirmer les hypothèses de départ, ou encore privilégier une explication ou une interprétation en mettant en échec des explications ou des interprétations concurrentes. Bref, l'objectif est d'offrir une solution à la question de recherche.

L'introduction, la méthodologie, les résultats et la discussion forment ensemble le corps du texte entendu comme récipient du contenu argumentatif, mais la structure textuelle complète inclut le titre, le résumé (*abstract*), les mots-clés (*keywords*), les remerciements (*acknowledgments*), la bibliographie (*references*) et les appendices. Chaque partie IMRAD peut se subdiviser ou se nommer différemment selon la tradition auquel appartient le texte (Thelwall, 2019).⁹ Selon Suppe (1998), la section « introduction » peut se

⁸ Notons que le statut épistémique de l'inférence à la meilleure interprétation est matière à débat. Kitcher (1991), Pera et Shea (1991), par exemple, distinguent la fonction rhétorique de la fonction épistémique dans les textes scientifiques. Or, comme le fait remarquer Suppe (1998), cette thèse pose problème en ce que son acceptation implique l'exclusion d'une part importante du discours dit « scientifique » du domaine épistémique.

⁹ À l'aide d'une étude statistique, Thelwall (2019) montre que les sous-titres dans les textes scientifiques diffèrent d'un domaine de recherche à l'autre. Son interprétation des résultats est que les textes scientifiques possèdent effectivement une structure, mais que celle-ci est hétérogène. Cette conclusion met à mal la thèse selon laquelle les textes scientifiques suivent une même structure rhétorique. Or, selon Suppe (1998), ce qui permet d'identifier une structure IMRAD n'est pas son titre, bien que celui-ci en soi l'indice, mais plutôt sa fonction rhétorique. Puisque

subdiviser en une introduction et un arrière-plan théorique (*theoretical background*); la section « méthodes » peut se diviser en techniques d'expérimentation ou d'observation (*experimental or observational techniques*) et présentation des données (*samples*); la section « résultats » peut se partager en analyse des données (*data analysis*) et résultats ou observations (*results or observations*) et la section « discussion » peut se diviser en discussion et conclusion (*summary/conclusions*). Selon Teufel (1999), les textes scientifiques peuvent être divisés en sept « zones argumentatives », c.-à-d. le contexte, le sujet, les travaux connexes, l'objectif (ou problème), la méthode (ou solution), les résultats; la conclusion (ou thèse principale). En bref, si le but sous-jacent au texte philoscience publié dans une revue à comité de lecture est de résoudre une question de recherche et que cette résolution constitue ce qui, *de facto*, est soutenu par le texte – nommons cela la thèse principale – et si on accepte les schémas épistémologiques classiques ainsi que celui de l'inférence à la meilleure interprétation, alors ce qui est soutenu par ce type de texte (la thèse principale) peut constituer une conclusion du processus d'investigation de l'hypothèse de départ (déductive ou bayésienne), une meilleure explication ou une meilleure interprétation des résultats.¹⁰ Enfin, cette résolution s'effectue habituellement à l'aide de quatre sous-objectifs traditionnellement nommés « introduction », « méthodologie », « résultats » et « discussion », lesquels donnent lieu à une structure textuelle nommée IMRAD. D'ailleurs, notons que cette structure textuelle tend à être reprise dans le résumé. À cet effet, John Swales (1981) propose une analyse structurelle des résumés dans les articles scientifiques. Son modèle original, connu sous le nom de *Create a Research Space (CARS)*, a été affiné par Swales (1996) pour englober une structure en cinq étapes : situer la recherche, présenter la recherche, décrire la méthodologie, résumer les résultats et discuter des résultats. Éventuellement, Swales lui-même et d'autres chercheurs comme Bhatia (2013) et Lorés (2004) ont formulé un modèle IMRAD du résumé.

1.4.3 Circonscription du discours de haut niveau théorique

Ce projet de recherche se concentre sur l'analyse du discours dans les Textes Courts de Haut Niveau Théorique publiés dans une Revue à Comité de lecture (TECHNETIUM). Plus spécifiquement, les textes analysés répondront au format IMRAD. Ces derniers pourront appartenir au domaine philosophique ou

l'étude de Thelwall ne procède pas à l'analyse fonctionnelle des sections, elle ne permet pas de réfuter la thèse d'une structure homogène. Parallèlement, Medawar (1964) conteste l'ordre dans lequel les fonctions IMRAD devrait être présentée, argumentant l'ordre IMRAD ne reflète pas celui dans lequel s'effectue la recherche. Notons que l'auteur ne conteste pas les fonctions elles-mêmes, seulement l'ordre dans lequel les présenter.

¹⁰ Puisqu'un même texte peut utiliser plusieurs schémas épistémologiques et contenir plusieurs thèses principales, alors un texte peut contenir des thèses de types distincts.

scientifique, car certains textes philosophiques répondent au format IMRAD. Par exemple, un TECHNETIUM où on fait l'analyse d'un discours scientifique ou philosophique peut être considéré comme un travail philosophique. Un tel travail philosophique peut être publié dans une revue scientifique qui porte sur la science investiguée (p. ex. Malaterre et Lareau, 2023d), un cas d'investigation philosophique à l'intérieur d'une science particulière, ou encore dans une revue scientifique portant sur les méthodes d'investigation (p. ex. Bertoldi *et al.*, 2023 ; Lamirel *et al.*, 2024 ; Malaterre et Lareau, 2023a). Autrement, un travail d'analyse du discours philosophique ou scientifique peut être publié dans une revue philosophique. Le cas échéant, lorsque les textes analysés sont philosophiques (p. ex. Lareau, 2022), alors le travail philosophique porte sur une pratique en philosophie, un cas de métaphilosophie, et lorsque les textes analysés sont scientifiques (p. ex. Malaterre *et al.*, 2019 ; Malaterre et Lareau, 2022), alors le travail philosophique porte sur une pratique scientifique, un cas d'étude en épistémologie. Que ce soit en philosophie ou en science, un travail d'analyse philosophique du discours peut impliquer des méthodes computationnelles (Kaiser *et al.*, à paraître) et, le cas échéant, le texte produit applique habituellement les fonctions IMRAD, c.-à-d. qu'on pose une question de recherche, on présente les méthodes, puis les résultats, et, enfin, on discute et présente les conclusions de la recherche. Notons, cependant, que les travaux en philosophie ne se limitent pas à l'analyse du discours scientifique ou philosophique, mais ils visent à répondre à une question de recherche, d'une manière ou d'une autre. Que ce soit en science ou en philosophie, les textes qui appliquent les fonctions IMRAD ne le font pas nécessairement de manière explicite (dans les sous-titres ou dans le corps du texte) et ne suivent pas inévitablement l'ordre séquentiel habituelle IMRAD. En fait, même quand le texte est explicitement et chronologiquement IMRAD, cela ne signifie pas que tous les énoncés à l'intérieur des sections IMRAD remplissent exclusivement la fonction explicitée dans le sous-titre. Autrement dit, il peut y avoir un peu de méthodologie dans l'introduction, un peu de discussion dans les résultats, etc. Bref, les sections textuelles, titrées IMRAD ou non, ne sont pas des vases clos. D'ailleurs, la recherche montre que les textes scientifiques ne sont pas sous-titrés uniformément relativement aux sections IMRAD (Thelwall, 2019). Par exemple, les TECHNETIUM en mathématiques et en philosophie ont tendance à utiliser des sous-titres concrets, remplissant parfois des sous-objectifs fins, plutôt que des objectifs génériques comme IMRAD (Graves *et al.*, 2013 ; Lucas, 2022). En philosophie, Lucas (2022) montre que la majorité des TECHNETIUM suivent une structure IMRAD, partiellement du moins, mais pas toujours de manière explicite et/ou chronologique. D'une part, les TECHNETIUM en philosophie contiennent généralement une section « introduction » ainsi que « discussion » ou « conclusion », mais rarement les sections « méthodes » ou « résultats ». Cependant,

remarque l'auteur, la section « introduction » n'est pas aussi approfondie en philosophie que celle utilisée en science :

[T]he philosophy RA [research article] is similar to IMRD structured RAs in that both usually contain an introduction and conclusion, here too philosophy RAs exhibit notable differences in the typical moves and steps that they make. Most outstanding, one seldom finds elaborate literature reviews in the introduction that are used to provide background and motivate the project. Philosophers also do not typically motivate their work by noting gaps in previous studies. (Lucas, 2022, p. 239)

D'autre part, lorsque les sections « méthodes » et « résultats » sont présentes dans les travaux philosophiques, celles-ci ne suivent pas toujours l'ordre séquentiel IMRAD traditionnel :

While there were sometimes sub-sections [...] labeled "methods" and "results", these sections showed notable differences from more conventionally structured IMRD articles. For instance, in some cases, authors present several different experiments, with a "methods" and "results" sub-section for each experiment rather than including the entire methodology in one section. (Lucas, 2022, p. 238)

D'ailleurs, l'auteur remarque que la méthodologie se trouve parfois en bas de page dans les travaux philosophiques, et non pas dans une section « méthodes » :

In other cases, parts of the methodology are actually described in footnotes, e.g., the data analysis procedures. (Lucas, 2022, p. 238)

En résumé, puisque les textes de haut niveau théorique sont par définition des textes qui visent à répondre à une question de recherche, nous postulons que les TECHNETIUM répondent également aux sous-objectifs qu'implique l'exploration ou la résolution d'une question de recherche. Parmi ces sous-objectifs, on trouve ceux génériques, à un haut niveau d'abstraction, qui sont élaborés de manière explicite et linéaire dans le format IMRAD. Cependant, ces fonctions IMRAD ne sont pas toujours limitées à des sections du texte, ni explicitées par des sous-titres, ni ne suivent exclusivement l'ordre séquentiel habituel. Néanmoins, cette étude se concentrera sur les TECHNETIUM répondant explicitement à l'ordre séquentiel du format IMRAD. On dispose aujourd'hui d'outils informatiques efficaces pour procéder à la lecture et l'analyse de textes de haut niveau théorique et un traitement plus exhaustif de la spécificité de ce discours incluant les textes structurés IMRAD.

1.5 Conclusion

En somme, l'analyse du discours peut être abordée à travers deux paradigmes principaux : le paradigme externaliste et le paradigme internaliste. Chacun de ces paradigmes offre une perspective différente sur la manière dont le discours peut être analysé et interprété. Le paradigme externaliste considère que le sens d'un discours est déterminé par des facteurs externes au texte lui-même. Il met l'accent sur le contexte social, politique, historique et culturel dans lequel le discours est produit. Selon ce paradigme, pour comprendre pleinement un discours, il est nécessaire d'analyser les conditions dans lesquelles il a été créé, les intentions de l'auteur, les relations de pouvoir en jeu et les idéologies dominantes. Le paradigme internaliste, en revanche, se concentre sur les caractéristiques internes du discours, en accordant moins d'importance au contexte externe. Il examine la structure et la forme du texte ainsi que les éléments linguistiques, syntaxiques et sémantiques utilisés. Les trois approches couramment associées au paradigme internaliste sont la théorie structuraliste, la théorie formelle et la théorie herméneutique. Selon la théorie structuraliste, le discours est un système de signes organisés dans une structure. Elle se concentre sur les relations entre les divers éléments du discours, tels que les phonèmes, les mots, les phrases, les paragraphes, etc. Selon la théorie formelle, le discours peut s'analyser au moyen d'outils mathématiques ou logiques. On cherche à décrire les règles formelles qui régissent la construction des énoncés et la logique de leur argumentation. L'accent est mis sur la cohérence interne et les structures logiques du discours. La théorie herméneutique met l'accent sur l'interprétation et la compréhension du discours. Elle considère le discours comme un moyen de communication qui peut être interprété de différentes manières en fonction des expériences et des connaissances du lecteur. La théorie herméneutique cherche à comprendre les significations et les intentions à travers une analyse approfondie de la forme matérielle du discours, c.-à-d. le texte. Ainsi, le texte de haut niveau théorique peut être analysé en fonction de ses grandes structures. Deux formes couramment utilisées sont la forme longue et la forme courte. La forme longue se réfère à un texte détaillé et exhaustif, qui explore en profondeur un sujet spécifique. La forme courte, quant à elle, est plus concise et synthétique, présentant les principaux points et arguments. Le texte de haut niveau théorique privilégié par la communauté philoscience est de forme courte publiée dans les revues à comité de lecture, et sa structure tend vers un format IMRAD, qui est un acronyme pour Introduction, Méthodes, Résultats et Discussion. Cette structure est souvent utilisée dans les articles scientifiques et les travaux de recherche philosophique pour organiser et présenter les informations de manière cohérente. Elle permet aux lecteurs de comprendre clairement les objectifs de l'étude, la méthodologie utilisée, les résultats obtenus et l'interprétation/discussion des résultats.

CHAPITRE 2

La lecture et l'analyse de texte assistées par ordinateur

Le précédent chapitre traitait de notions importantes sous-jacentes à toute lecture et analyse de texte, c.-à-d. le discours et sa forme matérielle (le texte). Dans ce chapitre, nous nous intéressons à une autre notion importante, c.-à-d. l'assistance de l'ordinateur pour la tâche de lecture et d'analyse de texte.

2.1 Introduction

La Lecture et l'Analyse de Texte Assistées par Ordinateur (LATAO) est un domaine de recherche fascinant et complexe, situé à l'intersection de plusieurs disciplines telles que l'informatique, les sciences cognitives, et les humanités numériques. Cet espace interdisciplinaire explore l'évolution et l'impact des technologies de l'information sur la lecture et l'analyse de textes, en se penchant notamment sur l'émergence et le développement de méthodes assistées par ordinateur. La genèse de cette approche remonte aux travaux pionniers de Roberto Busa et de Lauren B. Doyle.

La première démonstration de l'analyse de texte assistée par ordinateur a eu lieu dans les années 1950, après la rencontre de Roberto Busa avec Thomas J. Watson, le fondateur d'IBM. Busa a proposé l'idée d'utiliser des ordinateurs pour automatiser l'analyse de textes anciens, un concept révolutionnaire à l'époque. Son projet phare, l'*Index Thomisticus*, visait à indexer et analyser l'ensemble des œuvres de Saint Thomas d'Aquin. Cette tâche monumentale impliquait la gestion de millions de mots de textes latins médiévaux, ce qui était extrêmement difficile et fastidieux à réaliser manuellement. Avec le soutien d'IBM, Busa a utilisé des ordinateurs pour traiter le texte, marquant l'un des premiers usages de l'informatique pour le traitement du langage naturel et l'analyse de texte. Cela a conduit à la création d'une base de données exhaustive des œuvres de Thomas d'Aquin, facilitant grandement leur étude et leur recherche (Mounier, 2018).

Lauren B. Doyle est une figure importante dans le développement de la théorie et de la critique littéraires, en particulier en ce qui concerne l'application des méthodes informatiques à l'analyse de la littérature (Doyle, 1961b). Son travail a contribué à montrer comment les outils informatiques pouvaient être utilisés pour analyser des structures narratives, des thèmes récurrents, et des motifs stylistiques dans la littérature (Doyle, 1959, 1961b, 1961a, 1962, 1963, 1965). Elle a été une des premières chercheuses à reconnaître le potentiel des technologies informatiques dans les études littéraires et a contribué à l'émergence des

humanités numériques, un domaine qui combine les technologies numériques avec les disciplines des sciences humaines.

Ces contributions ont ouvert la voie à une multitude de recherches interdisciplinaires et de développements techniques, menant à l'établissement de la lecture et l'analyse de texte assistées par ordinateur (LATAO) comme un domaine de recherche distinct. Ce chapitre met en lumière l'importance de la modélisation des processus cognitifs humains et l'interaction dynamique entre l'ordinateur et l'être humain dans l'analyse de texte. Il aborde également les différentes applications de la LATAO, montrant son rôle essentiel dans l'exploration textuelle, notamment dans le cadre de l'analyse de textes de haut niveau théorique. En intégrant des techniques issues de la fouille de texte, de l'apprentissage machine, du traitement automatique des langues naturelles et des humanités numériques, la LATAO offre une perspective riche et nuancée sur la lecture et l'analyse de texte.

2.2 La lecture et l'analyse de texte assistées par ordinateur

La notion de lecture de textes assistée par ordinateur a été introduite en 1959 par Lauren B. Doyle, envisageant cette approche comme un soutien computationnel à la lecture. Cette méthode, qualifiée par l'auteur de « recherche visuelle assistée par ordinateur » (*computer-assisted eye searching*), promettait une forme innovante de lecture distante:

Our hope is that the state of the art of programming to produce highly condensed, information-rich formats for search by the human eye will improve at such a rate that pure machine searching of natural text may not for many years catch up in convenience or effectiveness with computer-assisted eye searching. (Doyle, 1959, p. 61)

Cette idée est contemporaine de la première démonstration du système d'indexation KWIC (*KeyWord In Context*) par Hans Peter Luhn en 1958, lors d'une conférence internationale sur l'information scientifique (*International Conference on Scientific Information*). KWIC indexait les documents selon la fréquence des mots, et utilisait une liste de « *stopwords* » pour filtrer les mots peu informatifs. Luhn est considéré comme un des pères fondateurs du traitement automatique de l'information (Stevens, 2018). En 1958 également, il a développé des concepts clés tels que l'indexation automatique, le résumé automatique, et le système de recherche d'information (Luhn, 1958). Parallèlement, Gerard A. Salton obtint son doctorat à Harvard, en 1958 où il enseigna avant de fonder, en 1965, le département d'informatique de l'université de Cornell ainsi que le premier grand groupe de recherche en analyse de texte assistée par ordinateur. Éventuellement, le domaine de l'apprentissage scolaire assisté par ordinateur se développe et des

chercheurs comme David Wray (1983) et O. J. Obrist (1983) vont s'intéresser à l'utilisation de l'ordinateur pour faciliter l'apprentissage de la lecture. Wray percevait l'interaction homme-machine comme un moyen d'améliorer les compétences de lecture chez les enfants :

It has to be said first of all that the computer is a significant advance in this respect on the teaching machine. It opens up much greater possibilities of a varied response based upon a child's answer. It is thus capable of interactive, rather than simply reactive, teaching, and this possibility needs to be exploited to the full. (Wray, 1983, p. 31)

À partir de là, il n'y avait qu'un petit pas à franchir pour envisager une amélioration similaire de la tâche de lecture pour les chercheurs eux-mêmes (Biskri *et al.*, 2013 ; Sinclair, 2003 ; Virbel, 1994).

La lecture et l'analyse de texte assistée par ordinateur (LATAO) sont proposées en 1997 dans un article de Jean-Guy Meunier intitulé « La lecture et l'analyse de texte assistées par ordinateur (LATAO)¹¹ comme système de traitement d'information ». Le chercheur s'inspire de travaux en analyse du discours, en informatique et en sciences cognitives afin de poser les bases théoriques d'une modélisation de la tâche complexe de lecture et d'analyse de texte au moyen d'un système de traitement de l'information :

Dans les sciences cognitives avec orientation technologique, il est devenu classique de concevoir une opération cognitive comme un traitement d'information. (STI). [...] Dans une telle perspective, on peut alors voir les actions cognitives d'accès au contenu d'un texte comme un traitement d'information (dont les données d'entrée sont des chaînes de caractères et les données de sortie des "interprétations". Ce type d'approche permettrait de quitter une analyse qui ne porterait que sur la dimension physique de la machine (ordinateur ou cerveau) et pour aborder les fonctions effectuées dans le traitement. (Meunier, 1997, p. 212)

Cette approche peut entretenir des similarités avec l'analyse de corpus assistée par ordinateur (*Computer-Assisted Corpus Analysis – CACA*), l'analyse de texte assistée par ordinateur (*Computer-Assisted Text Analysis – CATA*), l'analyse de contenu assistée par ordinateur (*Computer-Assisted Content Analysis*)¹², l'analyse du discours assistée par ordinateur (*Computer-assisted discourse analysis*), l'analyse textuelle du discours assistée par ordinateur, mais s'en distingue, entre autres, par son objectif explicite de

¹¹ La traduction anglophone du terme « LATAO » – *Computer-Assisted Reading and Analysis of Text (CARAT)* – apparaît en 1999 dans une communication intitulée « Use of classifiers in computer-assisted reading and analysis of text (CARAT) » (Meunier *et al.*, 1999).

¹² Le terme peut référer à l'analyse assistée par ordinateur du contenu en général, ou de thèmes plus particulièrement.

modélisation de la tâche cognitive de lecture et d'analyse (Bross et Stermole, 1973 ; Chuang *et al.*, 2014 ; Klobucar et O'Neill, 2021 ; Marchand, 1998 ; Mergenthaler, 1996 ; Popping, 2000 ; Rizkallah, 2013). À cet effet, Meunier (2022) circonscrit le programme de recherche de la LATAO à l'intérieur de la science normale, et le définit comme visant à rendre compte du phénomène complexe de lecture et d'analyse de textes au moyen de divers types de modélisation scientifique, notamment conceptuel, formel, computationnel et matériel :

Pour comprendre le rôle de l'informatique dans une pratique scientifique particulière comme la LATAO, il faut en appeler à la philosophie contemporaine des sciences. Pour celle-ci, les projets scientifiques ne visent pas avant tout la construction de théories comme telles ; ils viseraient plutôt la construction de modèles. C'est par eux qu'on explique ce qui doit être fait, en l'occurrence les tâches ou opérations propres à ce phénomène ou activité que sont la lecture et l'analyse de texte assistées par ordinateur. [...] Et lorsque le phénomène est complexe, de nombreux épistémologues proposent de construire non pas un seul modèle, mais plusieurs modèles. (Meunier, 2022)

La relation dynamique entre l'ordinateur et l'être humain est conceptualisée dès les premiers travaux de Meunier (1976) et de son groupe de recherche, notamment lors de la modélisation systématique de l'analyse de textes, qui mena en 1973 à une première version du Système d'Analyse des Textes par Ordinateur (SATO). Ce système vise l'interaction machine-humain via des descriptions de données et des interprétations :

SATO (Système d'analyse des textes par Ordinateur), has been implemented with a view of providing an interaction between textual data description and its interpretation. (Meunier *et al.*, 1976, p. 281)

La LATAO se distingue donc par sa vocation d'assistantat, c.-à-d. que bien que l'objectif global soit de modéliser la lecture et l'analyse de texte, certains travaux en LATAO peuvent s'intéresser plus particulièrement à certaines sous-opérations, et, ultimement, l'interprétation est considérée comme une opération qui n'est pas exécutée par l'ordinateur :

In CARAT, literary critics, philologists, content analysers, philosophers, theologians, psychologists, historians, and many other types of professional text readers (lawyers, journalists, etc.) require computer tools that assist them in their own and often personal reading and analysis expertise. And they cannot accept automatic text "interpretation" tools under any form whatsoever. (Meunier *et al.*, 2005, p. 956)

Similairement, Rastier (2018) suggère que le rôle essentiel de l'ordinateur est d'assister l'interprétation, ce qui est explicite dans le titre de son chapitre « *Computer-Assisted Interpretation of Semiotic Corpora* ». Selon Meunier (2009), l'interprétation n'est pas une fonction computable (réalisable par un algorithme), mais une fonction non-computable exécutable par une machine herméneutique (humaine probablement):

[I]nterpretation belong to or originate in the humans that use the technology as an extension of their own cognitive operations. In this sense, it [the computer] is not a hermeneutical machine as such, for linguistic and textual interpretation are uniquely human competences. (Meunier, 2009)

En fait, soutient Meunier (2012), la cognition humaine, dans toute sa complexité, est conçue comme un mélange de fonctions computables et non computables. Dans les termes de Turing, une telle fonction peut être exécutée (et, donc, modélisée) par une machine de Turing composée d'une machine « A » capable de résoudre des fonctions computables ainsi que d'une machine « O » – un oracle – capable de résoudre des fonctions non computables :

[I]n this model of computation, a Turing machine, that is, an A-machine, is now changed into a O-machine which is an A-machine plus a Oracle. The A-machine can always process a classical computational function. But if it is confronted to a non computable, or non-recursive function, it may the consult the oracle so that a solution can be found. And this oracle can be external to the A-machine. (Meunier, 2012)

Si la thèse selon laquelle la lecture et l'analyse de texte impliquent des fonctions computables et non computables est juste, alors elles sont modélisables par une machine « A » ainsi qu'une machine « O ».¹³ Subséquemment, les premières fonctions peuvent être systématiquement exécutées par un ordinateur et les secondes heuristiquement exécutées par un être humain.

Enfin, une autre caractéristique notable de la LATAO est la considération du texte comme entité linguistique distincte. Le texte, en tant qu'unité sémiotique, ne tire pas seulement sa signification de ses propres composants, mais aussi de ses interactions avec d'autres textes. D'une part, la signification ou la compréhension du texte sont médiatisées par l'acte de lecture et d'interprétation, lui-même préalablement façonnée par d'autres textes et donc enrichie par une dynamique intertextuelle. D'autre

¹³ Spécifions que la question de savoir si la tâche de lecture et d'analyse de texte est une fonction entièrement computable est objet de débat. Le cas échéant, une machine « A » serait suffisante pour exécuter cette tâche. Dans tous les cas, une bonne pratique est d'explicitement quelles opérations sont modélisées par des fonctions computables, et lesquelles ne le sont pas.

part, le texte est décomposable et, donc, analysable en ses sous-structures et ses unités lexicales (sections, paragraphes, phrases, mots, etc.). En d'autres termes, la signification d'un texte particulier ne peut être réduite à ses seuls constituants, puisqu'elle est surdéterminée par d'autres textes :

[L]a lecture n'est pas uniquement interprétation et compréhension intratextuelle, elle est aussi une dynamique intertextuelle. [...] En effet, la définition lexicale, c'est-à-dire la signification, n'épuise pas tout le sens émergeant de la formation des phrases, des paraphrases, mais surtout du texte. [U]n texte est plus que des mots et des phrases : c'est une entité linguistique originale, dont le contenu signifiant est structuré et intégré. Ce n'est pas une unité grammaticale, comme une unité syntaxique ou une phrase, et il n'est pas défini par sa taille. (Meunier, 2022, p. 389-90)

En bref, étant donné que la signification d'un texte ne réside pas uniquement en lui-même, mais aussi dans d'autres textes, il est judicieux d'aborder la sémantique d'un texte au sein d'un ensemble de textes – un corpus comme échantillon d'un discours spécifique par exemple – au sein duquel le texte acquiert son sens, du moins partiellement. La pratique de la lecture et l'analyse de texte assistées par ordinateur (LATAO) va donc bien au-delà de l'usage d'outils informatiques pour l'exploration de textes, y compris ceux de haut niveau théorique ou structurés selon le format IMRAD. D'une part, la LATAO est une activité interdisciplinaire impliquant la modélisation computationnelle, en totalité ou en partie, du processus de lecture et d'analyse de texte, processus pouvant varier selon la discipline qui la met en jeu. D'autre part, les différentes applications de ces modèles computationnels sur des textes particuliersinstancient une activité de lecture et d'analyse de textes de manière plus ou moins distante (Moretti, 2000 ; Underwood, 2017). En outre, la LATAO est un champ disciplinaire à part entière se situant à l'intersection des sciences cognitives, de la fouille de texte, de l'apprentissage machine, du traitement automatique des langues naturelles et des humanités numériques, car différentes techniques et méthodes utilisées en LATAO sont développées dans ces sous-domaines de l'intelligence artificielle. Ces techniques mettent en jeu des procédures déductives, inductives ou hybrides.

2.3 L'intersectionnalité de la LATAO

2.3.1 Les sciences cognitives

Le lien entre la lecture et l'analyse de texte assistées par ordinateur (LATAO) et les sciences cognitives réside dans la manière dont ces technologies aident à modéliser, comprendre et à analyser les processus cognitifs humains, en particulier ceux liés à la compréhension du langage et à l'interprétation des textes. D'une part, la LATAO est une méthode qui utilise des outils informatiques pour lire et analyser des textes

en s'inspirant de certains principes ou théories élaborés en sciences cognitives. D'autre part, les sciences cognitives s'intéressent à la manière dont les êtres humains pensent, apprennent et traitent l'information, et la LATAO offre un cadre de recherche permettant de modéliser, en tout ou en partie, ces activités. En combinant la LATAO avec les sciences cognitives, il est possible de gagner des intuitions sur la façon dont les individus et les groupes perçoivent et traitent l'information, ce qui est essentiel pour comprendre les processus de communication et de pensée humaine. Cette combinaison permet également de développer des modèles plus précis et des théories sur le fonctionnement cognitif en relation avec le langage et la communication.

2.3.2 La fouille de texte

La fouille de texte (FT), parfois nommée fouille de données textuelles (*text data mining*) ou l'analyse de texte (*text analytics*), est une spécialisation de la fouille de données qui se présente habituellement comme une suite d'opérations algorithmiques modulaires modélisant de manière simplifiée des théories linguistiques ou différentes tâches cognitives comme la lecture et l'analyse. La FT se développe à partir des travaux de Salton et de son équipe de recherche sur la recherche d'information (*information retrieval*), travaux qui mèneront à la représentation vectorielle du texte (Salton *et al.*, 1975). La modularité de la FT découle de l'analyse de la tâche complexe de fouille de texte qui est difficilement réalisable sans une décomposition en une série de sous-opérations. L'une d'entre elles, essentielle de la FT, est la transformation du texte, entendu comme un ensemble de données non structurées, en une représentation structurée permettant l'identification de régularités linguistiques quantifiables et pouvant correspondre à des relations cachées, c.-à-d. des relations difficilement identifiables lors d'une lecture traditionnelle sur les données non structurées.

2.3.3 L'apprentissage machine

L'apprentissage machine, parfois nommé apprentissage automatique, artificiel ou statistique, est un champ d'études où le processus d'apprentissage est modélisé de manière informatique à l'aide de différentes approches mathématiques et statistiques afin de résoudre des tâches qui ne sont pas explicitement programmées. Elle se présente sous forme d'apprentissages supervisé, semi-supervisé ou non supervisé et inclut les populaires réseaux de neurones artificiels dont l'apprentissage profond est une variante performante pour les larges *corpora* (Cao *et al.*, 2015 ; Hinton et Salakhutdinov, 2011 ; Ranzato et Szummer, 2008). Lors de l'apprentissage supervisé, les modèles sont entraînés sur un jeu de données étiquetées (où les réponses sont connues) et apprennent à prédire les réponses pour de nouvelles

données. Lors de l'apprentissage non supervisé, les modèles explorent les données non étiquetées pour trouver des régularités. Lors de l'apprentissage semi-supervisé, on utilise un mélange de données étiquetées et non étiquetées. L'apprentissage profond (*deep learning*) est un sous-domaine de l'apprentissage machine supervisé qui utilise des réseaux de neurones artificiels avec de nombreuses couches (d'où le terme « profond ») pour apprendre à reconnaître des régularités complexes dans de grandes quantités de données.

2.3.4 Le traitement automatique des langues naturelles

Le traitement automatique de la langue naturelle (TALN), parfois nommé traitement automatique des langues (TAL), NLP (de l'anglais : *natural language processing*), informatique linguistique, linguistique computationnelle, textométrie ou lexicométrie, est un champ d'études multidisciplinaire impliquant la linguistique, l'informatique et l'intelligence artificielle. Il vise la création d'outils facilitant le traitement informatique de la parole ou du texte pour diverses applications. Entre autres, ceux-ci permettent la reconnaissance des unités linguistiques (caractère, phonème, mot, parole, etc.), l'analyse morphologique (lemmatisation, fonction du mot dans la phrase, etc.), l'analyse syntaxique (induction grammaticale, segmentation, arbre syntaxique, etc.), l'analyse sémantique (le sens d'un mot en contexte, entité nommée, analyse de sentiment, désambiguïsation, etc.), la traduction automatique, etc.

2.3.5 Les humanités numériques

Le développement des Humanités Numériques (HN) a permis le déploiement de nouvelles traditions de lecture et d'analyse de texte. D'une part, elle met en place les assises théoriques d'une représentation ou d'une modélisation informatique du texte (Meunier, 2017). D'autre part, elle s'interroge sur le processus d'analyse et de lecture, notamment celui distant (*distant reading*) qui permet la lecture et l'analyse de larges *corpora* autrement inaccessibles à une lecture attentive traditionnelle (Moretti, 2000). Elles représentent un domaine de recherche et de pratique qui se situe à l'intersection des sciences et des technologies numériques, et hérite de certaines ambitions et problématiques propres aux humanités traditionnelles, comme la littérature, l'histoire, l'art et la philosophie. Comme le souligne Busa (2004), le but des HN est l'analyse computationnelle de toutes les expressions sémiotiques humaines, notamment celles discursives sous forme de textes :

Humanities computing is precisely the automation of every possible analysis of human expression (therefore, it is exquisitely a 'humanistic' activity), in the widest sense of the word,

from music to the theater, from design and painting to phonetics, but whose nucleus remains the discourse of written texts (Busa, 2004, p. xvii)

Dans le contexte de la LATAO, les HN jouent un rôle fondamental, car elles permettent de traiter et d'analyser de vastes ensembles de données textuelles, qui peuvent varier considérablement en fonction de la discipline concernée, sans compter le pluralisme des méthodes d'investigation. Meunier (2017) souligne justement le rôle essentiel des HN dans le développement de modèles adaptés à la réalité des humanités numérisées, lequel n'est possible qu'en tenant compte des dimensions conceptuelles, formelles, et informatiques de cette modélisation :

[L]es HN utilisent plusieurs types de modèles. [...] Un premier contient des modèles formels. Un de ses sous-modèles est de type mathématique. Il identifie des entités et des relations structurantes dans les objets d'étude et s'exprime dans des langages mathématiques spécifiques. Un autre sous-modèle est computationnel. Celui-ci retient des modèles mathématiques les fonctions structurantes qui sont computables. [...] Une deuxième classe de modèle est de nature matérielle qui, dans le hn, est surtout de nature informatique. Celle-ci renvoie à une machine électronique, c'est-à-dire un ordinateur[.] Enfin, une troisième classe de modèles est de type conceptuel. Ces derniers modèles, exprimés dans un langage naturel, permettent diverses formes de conceptualisation de l'objet de recherche et, à leur tour, ils assistent une compréhension plus limpide de la complexité de la recherche et de ses explications. C'est par ce langage naturel que le modèle conceptuel peut être communiqué, échangé et discuté dans des communautés épistémiques. (Meunier, 2017)

À cet effet, Élias Rizkallah, avec son équipe au centre d'Analyse de Testes assistée par Ordinateur (ATO), tente de mettre sur pied une infrastructure logicielle suffisamment malléable pour que les opérations exécutées par le système correspondent à l'activité scientifique d'analyse théorisée par la recherche. En ce sens, l'auteur critique les solutions informatiques « clés en main », très utiles dans certains cas, mais peu malléables, où le procédé opératoire est parfois opaque, ou qui ne correspond pas tout à fait à l'activité d'analyse envisagée, ce qui va à l'encontre même de la méthode scientifique (Rizkallah, 2013). En outre, la recherche en HN tend à favoriser les approches ouvertes, malléables, modulaires, et capables de rendre compte de différentes méthodes d'analyse. La LATAO, enrichie par les approches des HN, permet une exploration textuelle approfondie et plus nuancée, intégrant à la fois les aspects pragmatiques et théoriques de la recherche.

2.4 Les sous-objectifs de la LATAO

Généralement, une LATAO se présente de manière modulaire comme un enchaînement de sous-opérations pouvant répondre à des sous-objectifs, lesquelles peuvent elles-mêmes se décomposer en plus

petites opérations. Parmi ces sous-objectifs se trouvent l'extraction d'information, la catégorisation, la visualisation et la synthèse.

2.4.1 L'extraction d'information

L'extraction d'information permet d'identifier les unités textuelles les plus significatives ainsi que les relations qu'elles entretiennent (Grishman, 2015). Ainsi, le texte est décomposé en unités prédéfinies – caractères, groupe de caractères, mots, groupe de mots, phrases, etc. – lesquelles constituent les intrants de processus algorithmiques qui, à terme, produisent l'extrait correspondant aux informations recherchées. Ces techniques peuvent se spécialiser pour générer des bases de connaissances, des requêtes sous forme de questions et réponses (Bouziane *et al.*, 2015), identifier des relations conceptuelles (Cowling *et al.*, 2010) ou identifier des règles d'association déterminant la probabilité de trouver un thème, un concept ou un argument étant donné la présence d'un autre (Solanki et Patel, 2015).

2.4.2 La catégorisation

La catégorisation est le processus de regroupement d'éléments ou de données en fonction de caractéristiques ou d'attributs communs. Elle peut être basée sur une compréhension intuitive ou qualitative des données. Le cas échéant, les catégories peuvent être inférées à partir d'un ensemble d'exemples. Quant à elle, la classification implique d'attribuer des éléments ou des données à des classes préétablies en suivant un ensemble défini de règles ou de critères. Elle est souvent plus formelle et, ce faisant, elle ne nécessite pas d'exemple (Jacob, 2004). La catégorisation ou la classification sont des tâches pouvant se spécialiser lors de la détection du langage, la détection d'intentions, l'analyse de sentiments, de thèmes, de concepts ou d'arguments (Raju *et al.*, 2017). Les analyses mettant en jeu la catégorisation ou la classification peuvent bénéficier grandement de l'approche de l'analyse multidimensionnelle. En intégrant des annotations telles que la date de publication, l'appartenance institutionnelle ou nationale, les auteurs, l'éditeur, etc., cette approche permet de créer des perspectives riches et dynamiques. L'analyse multidimensionnelle facilite l'examen des données catégorisées à travers différentes dimensions, permettant ainsi des analyses plus approfondies et contextuelles, notamment par le biais de tableaux croisés dynamiques. Cela ouvre la voie à une compréhension plus nuancée des tendances et modèles inhérents aux données catégorielles ou aux métadonnées.

2.4.3 La visualisation

La visualisation permet aux chercheurs de saisir, de raisonner et d'interpréter une grande quantité d'informations textuelles en représentant des relations entre diverses unités d'analyse (Dou et Liu, 2016). Entre autres, il est possible de représenter les résultats d'analyse à l'aide de typographies, de diagrammes, de graphes, de lignes temporelles, de projections spatiales, de cartes topologiques, de visualisations radiales, de visualisations en 3D ou de visualisations en haute dimensionnalité (Liu *et al.*, 2019). Pour réaliser ces visualisations, plusieurs outils et techniques mathématiques sont utilisés. Par exemple, la théorie des graphes permet de représenter des réseaux de relations, où les nœuds symbolisent des entités et les arêtes représentent les liens entre elles. Les algorithmes de clustering, basés sur des méthodes statistiques ou des techniques d'apprentissage machine, sont utilisés pour regrouper des données similaires, facilitant ainsi leur visualisation et leur interprétation. Les fonctions de dimensionnalité réduite, comme l'analyse en composantes principales (PCA) ou t-SNE (*t-distributed Stochastic Neighbor Embedding*), sont souvent employées pour visualiser des données de haute dimensionnalité dans un espace à deux ou trois dimensions. Ces techniques aident à identifier des modèles ou des structures sous-jacents dans de grands ensembles de données.

2.4.4 La synthèse

La synthèse permet de présenter de manière condensée et structurée les informations importantes trouvées à l'aide de différentes techniques d'analyse (Ben Ayed *et al.*, 2020). On procède ainsi par des résumés, des condensés, des reformulations et des tableaux (Allahyari *et al.*, 2017).

2.5 Les spécialisations de la LATAO

La lecture et l'analyse de textes de haut niveau théorique ont pour objectif d'examiner ou de détecter des éléments de contenu textuels pertinents. Parmi ces éléments se trouvent les thèmes, les concepts et les arguments. Subséquemment, une LATAO peut être orientée spécifiquement sur les concepts (analyse conceptuelle), les thèmes (analyse thématique), ou les arguments (analyse argumentaire) (Kaiser *et al.*, à paraître). Ces spécialisations ouvrent également la voie à d'autres formes d'analyse, en intégrant, par exemple, des perspectives issues de la linguistique, de la sémiotique, de la psychologie cognitive, de l'informatique, de la philosophie, des humanités, enrichissant ainsi la compréhension et l'interprétation des textes.

2.5.1 L'analyse conceptuelle

L'analyse conceptuelle vise à identifier les concepts importants et leurs relations dans le texte. Elle dépend évidemment des théories du concept qui la fondent. Peu de travaux appliquent des stratégies informatiques à des textes de haut niveau théorique (Forest et Meunier, 2004 ; Loiseau, 2005 ; Thagard, 1988 ; Valette, 2003). Mais récemment, quelques études ont démontré la pertinence d'une telle application pour la LATAO, notamment sur les segments elliptiques, implicites ou périsegments (Chartrand *et al.*, 2016 ; Lareau, 2022 ; Pulizzotto *et al.*, 2016), ainsi que sur divers *corpora*, par exemple Descartes (Forest et Meunier, 2004), Darwin (Sainte-Marie *et al.*, 2011), Bergson (Danis et Meunier, 2012), Peirce (Chartrand *et al.*, 2016 ; Le *et al.*, 2016 ; Pulizzotto *et al.*, 2016, 2018) et Evans (Lareau, 2016 ; Lareau *et al.*, 2018). Dans le domaine de la LATAO, certaines analyses conceptuelles assistées par ordinateur ont été spécifiquement définies sous l'appellation de « Lecture et Analyse Conceptuelle de Texte Assistées par Ordinateur (LACTAO) », connue en anglais sous le nom de « *Computer-Assisted Reading and Conceptual Analysis of Text (CARCAT)* » (Danis, 2012 ; Danis et Meunier, 2012 ; Lareau, 2016, 2022 ; Lareau *et al.*, 2018).

2.5.2 L'analyse thématique

L'analyse thématique considère un texte comme un ensemble de thèmes et vise à repérer les divers segments qui participent à leur construction. Vilém Mathesius, fondateur du cercle linguistique de Prague, est à l'origine d'une théorie structurelle de l'information conçue en termes de thème et de rhème¹⁴. Le rhème peut être compris comme un rassemblement d'énoncés ou de mots exprimant diverses informations (concepts, notions, propriétés, traits caractéristiques, etc.) à propos de quelque chose qui est dit, c.-à-d. le thème (Marandin, 1988). Ainsi, le thème est le point de départ de ce qui va être dit et fait implicitement appel à l'information connue de l'auteur et du lecteur à son propos, tandis que le rhème est la description explicite du thème ainsi que les nouvelles informations que l'auteur a l'intention d'exprimer à son égard (Halliday et Matthiessen, 2004 ; Thompson, 2004). De nos jours, la notion de thème peut varier selon le domaine spécifique où l'analyse thématique est appliquée. Cette diversité se retrouve dans de multiples disciplines, telles que la linguistique, les sciences cognitives, les sciences sociales, les études littéraires, les sciences de l'information, la bibliométrie, la philosophie, entre autres (Forest, 2006). Les premiers travaux d'analyse thématique assistée par ordinateur ont utilisé des techniques de classification, qui permettent de regrouper des documents selon leur similarité lexicale. Depuis, d'autres

¹⁴ Les termes « thème » et « rhème » sont originellement introduits par Ammann (1925), et élaborés par N. S. Troubetskoï, R. Jakobson et S. O. Kartsevski.

méthodes ont vu le jour, dont celles fondées par l'algèbre linéaire, tels la décomposition en valeurs singulières ou les modèles de nature probabiliste comme l'analyse sémantique probabiliste latente, l'allocation de Dirichlet latente, les modèles de corrélation ou les réseaux de neurones artificiels (Alghamdi et Alfalqi, 2015 ; Lamirel *et al.*, 2024 ; Zhao *et al.*, 2021). Les techniques de *topic modeling* sont aujourd'hui très utilisées en FT (Chen et Liu, 2016 ; Malaterre *et al.*, 2019, 2020b, 2021 ; Malaterre et Lareau, 2022, 2023d, 2023a). Ces analyses permettent de repérer des régularités qui échappent à l'observation ordinaire et qui peuvent être interprétées comme des thèmes.¹⁵ Ces thèmes peuvent être identifiés à différents niveaux structurels du texte, notamment microstructurelle (la phrase par exemple), macrostructurelle (la section notamment) ou le texte dans son entièreté. Cependant, l'analyse thématique ne se réduit pas à l'identification des thèmes, car elle ne cherche pas seulement à savoir quels thèmes se trouvent où dans un texte ou un corpus de textes, mais aussi à comprendre comment les thèmes se relient entre eux :

La spécificité qui semble caractériser le processus d'analyse thématique (par opposition à l'identification des thèmes d'un corpus) réside principalement dans l'identification de la structure et des liens possibles entre les différents thèmes. (Forest, 2006, p. 12).

Aujourd'hui, la recherche développe l'analyse thématique avec l'émergence de nouvelles technologies et approches. Cela inclut l'identification des communautés cachées d'intérêts partagés (*hidden communities of interest*) permettant de découvrir des groupes d'individus qui discutent des mêmes thèmes, révélant ainsi des réseaux de communication et des connexions sous-jacentes (Malaterre et Lareau, 2023a, 2024a). Le développement de techniques de visualisation permettant de représenter graphiquement des thèmes et leurs relations, facilitant ainsi la compréhension des structures complexes des données (Bertoldi *et al.*, 2023 ; Kaiser *et al.*, à paraître ; Malaterre et Lareau, 2023d, 2023a, 2024a). L'analyse thématique peut également se concentrer sur l'évolution des thèmes au fil du temps, et même traiter des données dans plusieurs langues, ce qui est crucial pour les études internationales et multiculturelles (Bertoldi *et al.*, 2023 ; Malaterre *et al.*, 2021 ; Malaterre et Lareau, 2022). Afin de maintenir une cohérence terminologique avec la LATAO ainsi que la LACTAO, nous suggérons de nommer l'analyse thématique dans le contexte de la LATAO comme « Lecture et Analyse Thématique de Texte Assistées par Ordinateur (LATTAO) », ou en anglais « *Computer-Assisted Reading and Thematic Analysis of Text (CARTAT)* ».

¹⁵ Les *topics* identifiés par les techniques de *topic modeling*, ou ce qui est dit être un « thème » pour la majorité des autres approches issues de l'AM, correspondent généralement à un ensemble de variables ou d'éléments partageant certaines régularités linguistiques. La question de savoir si ces régularités indiquent effectivement des thèmes est objet de débat.

2.5.3 L'analyse d'argument

L'argumentation peut se concevoir différemment selon la posture épistémologique. D'un point de vue empirique, l'analyse de l'argumentation peut viser l'identification des procédés discursifs généraux ou des arguments particuliers par lesquels on réussit effectivement à atteindre l'objectif d'une argumentation. D'un point de vue normatif, celle-ci vise l'identification des conditions ou des critères grâce auxquels on peut obtenir ce pour quoi on argumente, mais de manière rationnelle. Autrement, l'analyse peut être à la fois empirique et normative lorsque, par exemple, on identifie les arguments d'autrui pour ensuite montrer en quoi ceux-ci respectent ou dérogent de la norme. Bien entendu, toute approche mettant en jeu des normes de la rationalité se confronte aux débats à ce sujet en sciences cognitives et en épistémologie. Enfin, l'analyse d'arguments peut être assistée par ordinateur et, le cas échéant, certains chercheurs nomment cette tâche « fouille d'argument » (*argumentation mining* ou *argument mining*). Comme le soulignent Habernal et Gurevych, la fouille d'argument est une tâche d'analyse du discours qui consiste à appliquer sur des données textuelles un modèle traduisant computationnellement une certaine théorie de l'argumentation :

Argumentation mining [is] the general task of analyzing discourse [...] and applying a certain argumentation theory to model and analyze textual data [...] (Habernal et Gurevych, 2017, p. 126)

Parmi les tâches importantes de la fouille d'argument se trouvent la détection des composantes argumentatives ainsi que la détection des relations entre ces composantes (Lippi et Torroni, 2016b). Récemment, la recherche a montré l'effectivité de certains outils informatiques à cet effet, notamment les classifieurs naïfs bayésiens (Biran et Rambow, 2011 ; Duthie *et al.*, 2016), les machines à support vectoriel (Bar-Haim *et al.*, 2017 ; Duthie *et al.*, 2016 ; Habernal et Gurevych, 2017 ; Lippi et Torroni, 2016a, 2016c ; Menini *et al.*, 2018 ; Mochales Palau et Moens, 2011 ; Naderi et Hirst, 2016 ; Niculae *et al.*, 2017 ; Park et Cardie, 2014 ; Stab et Gurevych, 2017), les modèles d'entropie maximale (Duthie *et al.*, 2016 ; Mochales Palau et Moens, 2011), la régression logistique (Dusmanu *et al.*, 2017 ; Ibeke *et al.*, 2017 ; Levy *et al.*, 2014 ; Nguyen et Litman, 2018 ; Rinott *et al.*, 2015), les arbres de décisions et les forêts aléatoires (Dusmanu *et al.*, 2017 ; Eckle-Kohler *et al.*, 2015 ; Stab et Gurevych, 2014b), les algorithmes d'analyse syntaxique (Eger *et al.*, 2017 ; García Villalba et Saint-Dizier, 2012a ; Peldszus et Stede, 2015), les réseaux neuronaux récurrents (Eger *et al.*, 2017 ; Niculae *et al.*, 2017), les champs aléatoires conditionnels (Stab et Gurevych, 2017), l'estimation du maximum de vraisemblance (Levy *et al.*, 2014) et les algorithmes de reconnaissance des inférences textuelles (*textual entailment*) (Cabrio et Villata, 2013). Pour assurer une

uniformité dans la terminologie avec la LATAO, la LATAO et la LATAO, nous proposons de désigner l'analyse d'argument dans le cadre de la LATAO comme « Lecture et Analyse Argumentaire de Texte Assistées par Ordinateur (LAATAO) », ou en anglais « *Computer-Assisted Reading and Argument Analysis of Text (CARAAT)* ».

2.6 Conclusion

En somme, la lecture et l'analyse de texte assistées par ordinateur constituent un ensemble de méthodes qui permettent d'explorer et d'analyser les textes de manière efficace, notamment les textes de haut niveau théorique. Ces approches utilisent des outils informatiques et des techniques de traitement automatique du langage naturel pour extraire des informations, identifier des schémas, effectuer des analyses statistiques et faciliter la compréhension du contenu textuel. Dans le contexte de l'analyse d'arguments, la recherche montre que cette tâche peut être assistée par ordinateur, c.-à-d. qu'elle est décomposable en sous-opérations dont certaines sont traduisibles en fonctions ou en algorithmes exécutables par ordinateur. Cette tâche est applicable à divers types de discours, mais peu de recherche traite d'une telle assistance pour les textes de haut niveau théorique. Notre hypothèse est qu'une telle assistance est possible pour la lecture et l'analyse de texte de haut niveau théorique, notamment les textes courts publiés dans les revues à comité de lecture.

CHAPITRE 3

L'argumentation

Les précédents chapitres traitaient de notions importantes à la fouille d'argument ou LAATAO, c.-à-d. la lecture et l'analyse de texte ainsi que l'assistance informatique de cette tâche. Dans ce chapitre, nous nous intéressons à une seconde notion importante sous-jacente, c.-à-d. l'argumentation. Nous explorons les différentes définitions de l'argumentation, notamment celles du dictionnaire ainsi que celles philosophiques. Plusieurs approches de l'argumentation sont présentées dont l'approche rhétorique (notamment antique, classique et moderne), l'approche épistémologique, l'approche consensualiste et l'approche communicationnelle.

3.1 Introduction

L'argumentation se manifeste dans de nombreuses activités humaines, tant dans les sphères publiques que privées, telles que l'échange d'opinions, le débat, la sollicitation, la caricature, la publicité, le reportage, le témoignage, l'essai, le roman, la fable, et s'étend plus largement aux domaines du droit, de la politique, de la philosophie, de l'éducation, de la science, de l'art, entre autres. Elle se matérialise à travers divers moyens d'expression sémiotiques, en particulier la parole et l'écrit, et se déploie à travers un éventail de genres discursifs, chacun caractérisé par les activités humaines qu'il reflète, comme les exemples mentionnés précédemment. Qu'importe le genre auquel il appartient, le discours est argumentatif lorsque son énonciation constitue un acte d'argumentation. Il peut aussi être partiellement argumentatif si seulement certaines de ses parties ont pour but d'argumenter.

3.2 Les définitions du dictionnaire

Du point de vue des usages en langue française, argumenter, c'est :

- i. Développer une suite d'arguments (TLFi),
- ii. Présenter ou prouver par des arguments (Petit Robert),
- iii. Justifier ou appuyer par des arguments (MDLF), ou
- iv. Présenter un ensemble d'arguments pour soutenir ou étayer une idée (Larousse).

Donc, l'acte d'argumenter est complexe en ce qu'il est composé d'unités – des arguments – qui sont développées et présentées dans le but de soutenir, prouver, justifier, appuyer ou étayer quelque chose.

Autrement dit, les arguments utilisés lors d'une activité où on argumente constituent des moyens par lesquels on atteint une fin, ce qui permet de distinguer ces arguments de ce qui est référé par le mot « argument » en contextes spéciaux comme la logique, les mathématiques, l'informatique ou les humanités. Or, qu'est-ce qu'un argument lors d'un acte d'argumentation? Dans les dictionnaires, un argument est :

- i. Une « affirmation particulière présentée à l'appui d'une démonstration [, un] moyen (autre que le raisonnement, la preuve logique) auquel on a recours pour convaincre, pour influencer la conduite de quelqu'un » (TLFi),
- ii. Un « raisonnement destiné à convaincre, à faire partager son avis par quelqu'un » (MDLF),
- iii. Un « raisonnement destiné à prouver ou réfuter une proposition » (Petit Robert),
- iv. Un « raisonnement [ou une] preuve destinés à appuyer une affirmation [...] moyen auquel on recourt pour convaincre quelqu'un, pour l'amener à modifier sa conduite » (Larousse).

Bref, un argument est un moyen, possiblement une affirmation, une démonstration, une preuve ou un raisonnement, dont la fin est d'influencer :

- a. La croyance d'autrui, possiblement en prouvant (ou en réfutant), en convainquant, ou en persuadant l'autre d'une proposition, d'une affirmation ou d'un avis;
- b. Le comportement d'autrui en modifiant sa conduite.

Dans le langage courant, l'argument semble caractérisé par deux dimensions importantes, c.-à-d. l'une cognitive apparaissant logique, voire formaliste, ou orientée vers la vérité (raisonnement, preuve, réfutation, démonstration, proposition, etc.) et l'autre pragmatique, voire sociale ou politique (moyen de persuader, de convaincre, de manipuler les comportements, les conduites, etc.). Cependant, on doit distinguer les définitions des dictionnaires de celles issues des domaines du savoir où on tente de définir, de décrire et de modéliser l'argumentation.

3.3 Les définitions philosophiques ou scientifiques

On peut distinguer au moins quatre définitions de l'argumentation en philosophie ou en science soit :

- i. Celle rhétorique où elle constitue un moyen de persuader ou de convaincre d'une thèse (Hamblin, 1970 ; Perelman et Olbrechts-Tyteca, 1958 ; Tindale, 2004 ; Toulmin, 1958),
- ii. Celle épistémologique où elle est un moyen de transmettre une croyance vraie justifiée ou d'atteindre collectivement la vérité (Biro et Siegel, 1992 ; Feldman, 1994 ; Goldman, 1999 ; Johnson, 2000 ; Lumer, 1988, 1991, 2005 ; Siegel et Biro, 1997),
- iii. Celle consensualiste où elle s'entend comme un moyen d'atteindre un consensus en résolvant ou en éliminant les différences d'opinions (van Eemeren *et al.*, 1996 ; van Eemeren et Grootendorst, 1984, 2003) et, enfin,
- iv. Celle communicationnelle où elle apparaît comme un processus dynamique entre un émetteur et un récepteur (Sperber et Mercier, 2012).

3.3.1 L'approche rhétorique antique et celle classique

Historiquement, la rhétorique naît à Syracuse au 5^e siècle avant Jésus-Christ lors des procès de propriété suivant le renversement de la tyrannie des Deinoménides où les victimes des déportations ou des expropriations demandent réparation devant de grands jurys populaires (Barthes, 1970). Corax (5^e s. av. J.-C.) est l'un des premiers enseignants de ce nouvel art de l'éloquence permettant de convaincre un jury. Il écrit le premier traité d'art oratoire recueillant un ensemble de techniques, d'exemples et de préceptes à l'usage des accusants et des accusés (Benoît, 1846). L'enseignement de la rhétorique consiste alors à apprendre par cœur des formules applicables à certaines situations, lesquelles forment des arguments types et leur mémorisation se nomme des « lieux » (*topoi*) ou « lieux communs » desquels on peut puiser ces formules permettant de se sortir d'impasse lors d'une plaidoirie (Reboul, 1991). Corax propose une première théorie de l'argumentation qui se déroule en cinq étapes, soit la captation de l'attention (l'exorde), l'exposé des faits, la discussion, la réfutation des arguments opposés et la synthèse (la péroraison) (Breton et Gauthier, 2000). Plusieurs amendements théoriques seront proposés par certains sophistes (Protagoras, Gorgias et Trasymaque par exemple) dont, entre autres, celui d'élargir le champ d'application de la rhétorique pour y inclure les débats d'idées, les joutes de pensées et la philosophie (Robin, 1923). Similairement, Socrate soutient que la rhétorique est applicable à tous les sujets de discussion, mais il s'y oppose parce qu'elle est à la solde du plus offrant (Gorgias) ou sert l'opinion au détriment de sa véritable destinée – la vérité – et, ce faisant, elle est immorale et dénaturée (Breton et Gauthier, 2000). Subséquemment, il propose une nouvelle approche à l'argumentation nommée « dialectique », laquelle comporte de nouvelles techniques comme la division du sujet en ses éléments – l'analyse – et le rassemblement de ces éléments en un tout cohérent – la synthèse (Phèdre). Enfin, Aristote

(2011) répond à ces critiques en arguant que la rhétorique est amoral (en dehors de la morale) et non pas immorale, car si on peut faire le bien ou le mal avec le même outil, c'est plutôt l'agent qui s'en sert qui apparaît comme moral ou immoral (Rhétorique, livre I, 1355b). Ainsi, clarifie le Stagirien, la rhétorique ne s'oppose pas à la dialectique, elle en est une partie et lui ressemble (Rhétorique, livre I, 1356a). Avec Aristote et Quintilien (1er s. apr. J.-C.), la rhétorique se définit comme l'art de la persuasion, puis elle se transforme au 16e siècle en une science de l'élocution où on étudie les figures ou tropes du discours (voir Ramus, Dumarsais, Fontanier).

3.3.2 L'approche rhétorique moderne

On peut considérer Perelman et Olbrechts-Tyteca (1958) ainsi que Toulmin (1958) comme les pères de la rhétorique moderne qu'ils ne conçoivent pas comme art de l'élocution, mais comme moyen de persuasion rationnel. De manière précoce, ces théoriciens modernes vont amorcer un débat qui deviendra important en épistémologie et en sciences cognitives à savoir si la rationalité doit se concevoir en termes de logique classique, ou s'il ne serait pas plus prudent de l'étudier pour ce qu'elle est, empiriquement, en investiguant l'endroit même où elle se déploie, c.-à-d. l'argumentation (Ribeiro, 2009). Dès les premières phrases de la *Nouvelle rhétorique*, Perelman et Olbrechts-Tyteca (1958) rejettent le postulat selon lequel la raison est réductible à la logique et adhèrent plutôt à une rhétorique antique qui, comme on l'a vu, s'appuie sur l'expérience et mène au développement des lieux communs de la raison :

La publication d'un traité consacré à l'argumentation et son rattachement à une vieille tradition, celle de la rhétorique et de la dialectique grecques, constituent une rupture avec une conception de la raison et du raisonnement, issue de Descartes, qui a marqué de son sceau la philosophie occidentale des trois derniers siècles (Perelman et Olbrechts-Tyteca, 1958, p. 1)

Ainsi, une théorie sérieuse de l'argumentation cherche à comprendre ces lieux communs, à savoir comment des arguments provoquent ou accroissent « l'adhésion des esprits aux thèses qu'on présente à leur assentiment » (Perelman et Olbrechts-Tyteca, 1958, p. 5). Parallèlement, dans *The uses of argument*, Toulmin (1958) rejette explicitement une réduction de l'argumentation au formalisme logique puisque, en tant qu'outil (parmi d'autres), celui-ci n'est rien sans un agent pour l'utiliser et lui donner sens – une sémantique – en référant à une chose ou à une autre :

This is where the mathematical logician comes on the scene. For, he can claim, an argument is made up of propositions, and the logician's objects of study are the formal relations between propositions; to ask whether an argument is valid is to ask whether it is of the right

form, and the study of form is best undertaken in a self-consciously mathematical manner; so we must sweep away all references to thinking and rationality and the rest, and bring on the true objects of logical study, the formal relations between different sorts of propositions.... But this is where we came in, and the ensuing paradox is already in sight. We can hardly sweep away all references to thinking without logic losing its original practical application: if this is the price of making logic mathematical, we shall be forced to pose the Kantian-sounding problem, 'Is mathematical logic at all possible?' (Toulmin, 1958, p. 6)

Autrement dit, l'argumentation n'est pas une activité qui se déroule dans le vide, c.-à-d. que ce ne sont pas des arguments ou des propositions qui s'engagent dans cet acte qui est impossible sans une personne réelle qui pense habituellement de manière pragmatique et non pas de façon formelle ou à la manière des mathématiciens.

3.3.3 L'approche épistémologique

L'approche épistémologique, comme on vient de le voir, est critiquée par les pères de la rhétorique moderne. Historiquement, elle naît de la critique de Socrate à la rhétorique naissante, mais depuis Aristote, on peut la concevoir comme une branche de la rhétorique. Selon le Stagire, il y a au moins quatre dimensions rhétoriques, c.-à-d. quatre moyens généraux de persuader, soit l'appel à la logique (*logos*), l'appel à la crédibilité de l'interlocuteur (*ethos*), l'appel à l'émotion (*pathos*), et l'appel à l'action (*kairos*). Dans ce cadre, l'approche épistémologique classique ne s'intéresse qu'à une seule de ces dimensions – l'appel à la logique – et on conçoit l'argumentation en termes de moyen de persuader logiquement (Ehninger, 1954 ; Windes et Kruger, 1961), moyen de transmettre une croyance vraie justifiée (Biro et Siegel, 1992 ; Feldman, 1994 ; Johnson, 2000 ; Lumer, 1991, 2005 ; Siegel et Biro, 1997) ou moyen d'atteindre collectivement la vérité (Goldman, 1999 ; Lumer, 1988). Autrement, Toulmin (1958) s'oppose à une réduction de l'argumentation à la logique formelle parce qu'il considère que celle-ci ne décrit pas l'activité qui a lieu lorsqu'on argumente, mais il ne considère pas l'argumentation comme illogique ou irrationnelle. Subséquemment, certains théoriciens vont développer une approche alternative à la logique classique – laquelle sera dite informelle – au sens d'une logique « naturelle » distincte de la logique formelle auquel Toulmin s'oppose. Analogiquement, si l'usage d'une langue naturelle implique une grammaire, laquelle peut se concevoir comme un ensemble de règles en quelque sorte « informelles » au sens où on doit analyser le discours pour en inférer les règles, alors l'argumentation comporte aussi un ensemble de règles qu'on pourrait dire « informelles » dans la mesure où elles peuvent être inférées en analysant le discours argumentatif et différer des règles de la logique formelle (laquelle pourrait se dire « artificielle » le cas échéant). Par exemple, Blair et Johnson (1987) définissent la logique informelle

comme un ensemble non formel de normes, de critères, et de procédures pour l'analyse, l'interprétation, l'évaluation, et la critique de l'argumentation. Ce faisant, on peut considérer l'approche de Toulmin comme étant rhétorique parce qu'il tient compte des acteurs (et non seulement de la forme des arguments en jeu), mais on peut aussi le considérer comme réformateur de l'approche épistémique dans la mesure où il propose de reformuler les notions de logique et de vérité.

3.3.4 L'approche consensualiste

Tout comme l'approche rhétorique, celle consensualiste postule que la fin de l'argumentation est une croyance, mais celle-ci se démarque par son aspect consensuel, c.-à-d. que l'objectif est le partage d'une même croyance par les acteurs. Elle s'inspire d'une théorie originale développée par Habermas (1981) et Apel (1981) – la pragmatique universelle – dont l'objectif est de dégager les conditions de réalisation de la communication, c.-à-d. d'identifier ce qui rend possible la compréhension mutuelle lors des usages du langage. Il apparaît que la communication est impossible sans une coopération entre les acteurs de sorte que l'activité communicationnelle semble émerger, potentiellement du moins, d'un consensus universel. Comme le souligne Miller (2005), la pragmatique universelle transforme profondément notre compréhension de la notion de rationalité, car si l'acte communicationnel présuppose un idéal intersubjectif, c.-à-d. le potentiel d'un consensus sans condition, alors on ne se trouve pas devant un sujet qui pense, mais devant un sujet qui argumente où chaque acte de langage est un engagement vers une raison commune :

Le consensus intersubjectif inhérent au langage se réalise par l'argumentation et se fonde sur la reconnaissance mutuelle de ces prétentions à la validité élevées par tout acte de parole. Il se base donc sur l'idéal régulateur d'une situation idéale de parole. (Miller, 2005, p. 13)

La théorie pragma-dialectique proposée par Frans van Eemeren et Rob Grootendorst (1984, 2003) est un exemple d'approche consensualiste de l'argumentation. Si le consensus est une condition nécessaire et suffisante à toute communication, alors on peut considérer l'approche consensualiste comme une approche communicationnelle. Or, l'argument transcendantal qui fonde cette approche est insuffisant pour expliquer le tout du phénomène de l'argumentation et il est insuffisant pour fonder un nouveau paradigme à propos de la rationalité. En critiquant plus spécifiquement la pragma-dialectique, Lumer (2010) souligne cette limite de l'approche consensualiste (partagée avec l'approche rhétorique) qui se manifeste par une difficulté à rendre compte des phénomènes de validité, de vérité ou de justification que la tradition épistémologique reconnaît à l'argumentation. Cette limite est d'autant plus problématique

qu'elle ne permet pas le passage du sujet pensant au sujet argumentant puisque seul le sujet cartésien semble garantir du vrai :

It is quite astonishing that even though Van Eemeren and Grootendorst repeat their aim for argumentative discourse, i.e. dispute resolution, countless times, they practically do not justify this most central assumption of their approach. They incidentally justify the need for dispute resolution with the remark that “otherwise we become intellectually isolated and can ultimately even end up in a state of spiritual and mental inertia” [...]. However, “not being intellectually isolated” could be an euphemism for “conformism”. Of course, not being intellectually isolated is good; but it is of secondary importance. It is much more important that one’s beliefs are true (and justifiedly true) and thus can help one orient herself or himself in the world. Intellectual isolation could simply be the price of truth, or more precisely, of justified true beliefs that others are not able or not willing to understand or accept – think of Galileo or Frege. (Lumer, 2010, p. 44-45)

3.3.5 L’approche communicationnelle

Éventuellement, le débat amorcé par Toulmin (1958) ainsi que Perelman et Olbrechts-Tyteca (1958) est approfondi par Cohen (1981) et alimenté par le développement des sciences cognitives et des sciences de l’informatique. En bref, contrairement à la conception traditionnelle en sciences cognitives et en épistémologie selon laquelle la rationalité s’entend en termes de logique classique et de probabilité, celle-ci doit être étudiée pour ce qu’elle est empiriquement et peut se concevoir autrement en termes de normes naturelles, de principes d’adaptation ou de traitement optimal de l’information (Anderson, 1990 ; Cosmides et Tooby, 1996 ; Gigerenzer, 2007 ; Oaksford et Chater, 2007). Dans ce contexte, Sperber et Mercier (2012) proposent une théorie argumentative du raisonnement qui s’inscrit à l’intérieur de ce débat. Similairement à l’approche consensualiste, l’argumentation y est décrite comme une compétence sociale visant à articuler collectivement des solutions, des raisons, des arguments ou des actions de manière à réaliser des buts communs. L’argumentation implique des raisonnements qui sont la mise en œuvre de processus cognitifs par lesquels s’élaborent des solutions, souvent sous forme d’argumentations, dans le but de résoudre des problèmes collectifs, bien que ces mêmes processus puissent être utilisés pour réaliser des objectifs personnels. Les raisonnements sont postulés comme émergeant d’un module métacognitif issu de l’évolution naturelle facilitant la sélection ainsi que la transmission d’informations maximisant un certain apport épistémologique collectif. Fondamentalement, ce qui distingue l’approche communicationnelle de celle consensuelle est une précision de la condition de coopération, c.-à-d. que la coopération est une condition qui avantage l’émetteur, ainsi que l’ajout d’une seconde condition selon laquelle la véracité avantage le récepteur. Autrement dit, l’argumentation doit mettre en œuvre une certaine fonction collective permettant un transfert d’informations véridiques sans quoi le récepteur

n'aurait aucun avantage à s'engager dans cette activité et cette fonction doit permettre d'influencer autrui (en introduisant des faussetés possiblement) sans quoi l'émetteur n'aurait aucun avantage à s'engager dans cette activité :

What makes communications advantageous to receivers is that it provides them with rich information that they could not, or not easily, have obtained on their own. For this, the information they receive has to be genuine information, that is, close enough to the truth. What makes communication adventurous to communicators is that it allows them to achieve some desirable effect in the receivers. For this, the information they emit has to be conducive to this effect, whether it is true or false. (Mercier et Sperber, 2011b, p. 96)

Une seconde distinction entre l'approche communicationnelle et l'approche consensualiste est la visée d'une certaine intégration aux sciences cognitives, laquelle implique que derrière tout acte de langage se trouve un acte cognitif, mais qu'un acte cognitif ne produit pas nécessairement un acte de langage. Par exemple, Sperber et Mercier soutiennent que les raisonnements sous-jacents l'argumentation ne sont pas nécessairement linguistiques, conscients, abstraits, généraux, symboliques (au sens d'une manipulation de symboles), logiques ou probabilistes. Leur théorie s'inscrit à l'intérieur du paradigme de la dualité cognitive, c.-à-d. qu'un raisonnement peut se dérouler de manière générale et systématique en appliquant des règles ou, encore, se dérouler de manière spécifique et heuristique en utilisant d'autres moyens (voir Lareau, 2016). Comme le souligne Poincaré (1902), toutes les combinatoires permises par la logique ne sont pas pertinentes et, pour éviter l'explosion des combinatoires, une astuce semble parfois nécessaire :

Parmi toutes les constructions que l'on peut combiner avec les matériaux fournis par la logique, il faut faire un choix; le vrai géomètre fait ce choix judicieusement parce qu'il est guidé par un sûr instinct, ou par quelque vague conscience de je ne sais pas quelle géométrie plus profonde, et plus cachée, qui seule fait le prix de l'édifice construit. (Poincaré, 1902, p. 129)

Autrement dit, une stratégie analytique systématique parcourant tout l'espace de recherche pour résoudre un problème est rarement optimale et, donc, il est souvent utile et économique de réduire cet espace par une heuristique ou une autre. Ainsi, Sperber et Mercier (2012) conçoivent deux manières de raisonner. La première – disons de type 1 – met en jeu des procédures spécifiques et quelque chose qui n'est pas une proposition (disons une assumption) et la deuxième – disons de type 2 – met en jeu des procédures générales et des propositions :

Inferences in many specific domains could be more efficient if tailored to take advantage of the specificity of the domain and to employ dedicated procedures rather than have that

specificity be represented in propositional form and used by general procedures such as conditional inferences. (Sperber et Mercier, 2012, p. 373)

Les raisonnements de type 1 se conçoivent en termes de modularité massive dans la mesure où les diverses procédures sous-jacentes sont implémentées par une multitude de modules cognitifs spécialisés – appris ou innés – dont les domaines d’application sont spécifiques contrairement aux procédures générales des raisonnements de type 2 :

[...] inferences may be more effectively carried by specialized domain-specific mental devices or “modules” than by a general ability drawing on a single huge database [...], those inferences that become modularized either in evolution or in cognitive development are those that are the most likely to be relevant and to be performed spontaneously (Sperber et Mercier, 2012, p. 374)

Les raisonnements de type 1 – largement inconscients ou fondés par l’intuition – et les raisonnements de type 2 – conscients ou fondés par la raison – peuvent s’articuler pour former des argumentations qui sont ainsi des productions hybrides, mais qui peuvent elles-mêmes faire l’objet de raisonnement ou d’argumentation de niveau supérieur :

[...] when one consciously reasons and sees oneself as involved in a series of personal epistemic decisions, one is mistaking the visible tip of a mental iceberg for its largely invisible structure. [Otherwise,] one can engage in higher-order reasoning, that is, reason about reasons, rather than just accept them or reject them as intuitively strong or weak. (Sperber et Mercier, 2012, p. 375-6)

La recherche en sciences cognitives semble soutenir la théorie argumentative du raisonnement puisque l’argumentation est une compétence acquise très tôt (dès l’âge de trois ans) lors du développement cognitif humain (Eisenberg et Garvey, 1981 ; Stein et Albro, 2001 ; Stein et Bernas, 1999 ; Stein et Miller, 1993), mais elle montre que l’être humain raisonne souvent de manière suboptimale (Ragni *et al.*, 2017 ; Wason, 1966) de sorte qu’on est en droit de se demander comment les processus de raisonnement réussissent à générer de la connaissance? Les auteurs expliquent que le succès épistémologique provient du fait que l’argumentation est une entreprise éminemment collective, c.-à-d. que ce qui apparaît comme inefficace ou irrationnel derrière les processus individuels de raisonnement apparaît autrement efficace et rationnel (évolutivement du moins) lorsqu’ils sont envisagés collectivement. En somme, l’argumentation est un produit collectif émanant de raisonnements qui sont exécutés par des modules cognitifs appartenant à des individus et dont la fonction est l’évaluation ou la production de raisons sous forme d’assumptions ou de propositions. Dans la mesure où les raisonnements qui constituent

l'argumentation se situent en contextes dialogiques, l'argumentation se trouve en porte à faux entre le besoin de persuasion de l'émetteur et le besoin de vérité du récepteur et, ce faisant, les assomptions ou les propositions sont à la fois performatives et relatives à la vérité ou à la vraisemblance (Mercier et Sperber, 2011b, 2011a).

3.4 Conclusion

En somme, nous avons présenté les quatre principales théories de l'argumentation, soit l'approche rhétorique, l'approche épistémologique, l'approche consensualiste et l'approche communicationnelle. L'approche rhétorique de l'argumentation met l'accent sur les techniques de persuasion utilisées dans les discours. Elle remonte à l'Antiquité grecque, où les orateurs cherchaient à convaincre leur auditoire en utilisant des dispositifs rhétoriques tels que l'utilisation d'émotions, de figures de style et de stratégies de manipulation verbale. Selon cette approche, l'argumentation est principalement axée sur la forme du discours et les techniques de persuasion employées pour influencer l'opinion. L'approche épistémologique de l'argumentation se concentre sur la recherche de la vérité et la justification des croyances. Elle vise à évaluer la validité et la solidité des arguments en se basant sur des critères logiques et épistémiques. Selon cette approche, l'argumentation repose sur des raisonnements logiques, des preuves empiriques et des normes de justification. Elle cherche à établir la validité d'une affirmation en se fondant sur des principes rationnels et des règles d'inférence. L'approche consensualiste de l'argumentation met l'accent sur la recherche d'un consensus ou d'un accord entre les parties en conflit. Elle considère que l'argumentation est un processus de négociation où les participants cherchent à atteindre un compromis acceptable pour tous. Selon cette approche, l'argumentation est une forme de communication coopérative visant à résoudre les différends et à parvenir à un accord mutuel. L'approche communicationnelle de l'argumentation se concentre sur le processus de communication lui-même. Elle met l'accent sur la façon dont les arguments sont construits, interprétés et échangés dans un contexte de communication. Selon cette approche, l'argumentation est un processus interactif où les participants s'engagent dans des échanges de points de vue, des négociations et des dialogues pour construire et partager des significations. Elle met en avant l'importance de la coopération, de la compréhension mutuelle et de la recherche de solutions communes. Ces différentes approches offrent des perspectives complémentaires pour comprendre les mécanismes et les enjeux de la persuasion et de la communication. Chacune met l'accent sur des aspects spécifiques de l'argumentation, tels que les techniques rhétoriques, la recherche de vérité, la recherche d'un consensus ou la dynamique de la communication.

CHAPITRE 4

Les modèles computationnels de l'argumentation

Dans ce chapitre, nous présentons les différentes modélisations des approches rhétorique, épistémologique, consensualiste et communicationnelle. Soulignons que certaines de ces modélisations ont été implémentées dans des applications informatiques. Lorsque c'est le cas, nous examinons ces applications ainsi que les données sur lesquelles elles ont été testées. Bien que ce chapitre ne se concentre pas sur les descriptions formelles ou computationnelles, nous fournirons une vue d'ensemble conceptuelle de ces modèles. Ces derniers, lorsqu'ils sont mis en œuvre de manière computationnelle, peuvent également être décrits de manière formelle et computationnelle.

4.1 Introduction

Les quatre approches mentionnées au chapitre précédent forment des cadres théoriques paradigmatiques dont certains font l'objet d'une modélisation. Parmi les modèles développés, certains sont propices à une traduction computationnelle, c.-à-d. que certains modèles peuvent être exprimés en un langage informatique permettant un traitement computationnel des données textuelles. Conformément à nos objectifs, nous nous intéressons plus particulièrement aux modèles faisant l'objet d'une traduction computationnelle pour chacun des cadres théoriques paradigmatiques, c.-à-d. les modèles rhétoriques, les modèles épistémologiques, les modèles consensualistes et les modèles communicationnels.

4.2 Les modèles rhétoriques

De nombreux théoriciens de l'argumentation, tels qu'Aristote, Cicéron, et La Boétie, ont établi une classification des arguments dans des catégories abstraites nommées *topoi* ou *loci*. Ces lieux communs permettent de regrouper les arguments selon des principes généraux, facilitant ainsi leur mémorisation et leur utilisation en situation concrète. Ces hauts lieux de l'Argumentation jouent un rôle pratique pour la mémorisation et l'enseignement, ainsi qu'un rôle épistémologique pour l'explication ou la description. Rudolph Agricola, dans son œuvre *De Inventione Dialectica*, a synthétisé les théories de la dialectique et de la rhétorique d'Aristote, Cicéron, La Boétie et d'autres. Agricola a contribué à systématiser et à clarifier les lieux communs en les intégrant dans un cadre dialectique cohérent. Comme le souligne McNally (1967), l'importance d'Agricola réside dans le fait qu'il a assimilé l'art de la dialectique à celui de la rhétorique qui,

selon lui, est un domaine de recherche permettant de révéler les fondements ultimes du discours, du langage, du raisonnement et de l'homme :

Agricola is attempting to reveal the ultimate foundations of discourse in the nature of language, reasoning, and man, and to relate all the functions and forms of language in a single comprehensive theory. (McNally, 1967, p. 394)

Le Père Gérard Pelletier, dans le contexte de l'éducation jésuite et de la rhétorique de la Renaissance, a aussi apporté une contribution significative à l'approche rhétorique. Son œuvre principale, *Palatium Reginae Eloquentiae*, parue en 1641, est un traité sur l'art de la rhétorique. Cette œuvre vise à enseigner les principes de l'art oratoire, la construction d'arguments et les techniques de persuasion (Pelletier, 2021). Parmi les catégories étudiées, on retrouve des arguments basés sur la définition (*locus a definitione*), la similitude (*locus a simili*), l'opposition ou le contraste (*locus ex contrario*), la causalité (*locus a causa*), l'autorité (*locus ex autoritate*), etc. Éventuellement, les théoriciens de l'approche rhétorique vont s'orienter vers une science de l'argumentation, délaissant l'art oratoire et la notion de lieu commun au profit de celle de schème ou de type argumentatif. Par exemple, Hastings (1962) a distingué différents schèmes argumentatifs permettant de classer les arguments en fonction des processus de raisonnement liant les prémisses aux conclusions (voir le tableau 4.1 suivant).

Tableau 4.1 Schèmes argumentatifs d'Hastings

Schémes argumentatifs d'Hastings

De l'exemple à la généralisation
Du critère à la classification verbale
De la définition aux caractéristiques
Du signe à l'événement inobservé
De la cause à l'effet
De la circonstance à l'hypothèse
Par comparaison
Par analogie
Par témoignage

Dans le modèle d'Hasting, le schème « de l'exemple à la généralisation » s'appuie sur l'utilisation d'exemples spécifiques pour formuler une règle ou une conclusion de portée générale. Dans le schème

« du critère à la classification verbale », on se base sur certains critères pour inférer une catégorie ou une définition. Le schème « de la définition aux caractéristiques » infère certains attributs à partir d'une définition. Le schème « du signe à l'événement inobservé » crée une corrélation entre un signe observable et un événement ou un état qui n'est pas directement observable. Quant au schème « de la cause à l'effet », il repose sur une relation causale pour associer une cause donnée à son effet potentiel. Le schème « de la circonstance à l'hypothèse » utilise les circonstances ou les contextes spécifiques pour élaborer des hypothèses. Le schème « par comparaison » forge des arguments en mettant en parallèle des éléments ou des situations, mettant en évidence leurs similitudes ou leurs différences. De manière similaire, le schème « par analogie » construit des arguments en établissant des liens de similarité ou de parallélisme entre des éléments ou situations distincts. Enfin, le schème « par témoignage » s'appuie sur des témoignages ou des déclarations d'autorités reconnues pour renforcer la validité d'un argument.

Perelman et Olbrechts-Tyteca (1958) proposent d'autres schèmes s'organisant de manière hiérarchique dont les niveaux les plus abstraits sont subsumés par des principes généraux comme l'association (jouant le rôle de support à une thèse), la division (qui joue un rôle d'opposition à une thèse), le lien à la logique (ou jouant un rôle inférentiel quasi logique), et le lien à la réalité (permettant la distinction entre le descriptif et le normatif). Kienpointner (1992) distingue 21 schèmes variant selon 4 propriétés générales, soit le type d'inférence, la nature épistémique des prémisses, la fonction dialectique (permettant de distinguer ce qui supporte et ce qui s'oppose à une thèse) et la fonction pragmatique (permettant la distinction entre le descriptif et le normatif). Éventuellement, plusieurs théoriciens vont proposer différents jeux de schèmes argumentatifs. Notamment, Walton (1996) postule une trentaine de schèmes argumentatifs, Grennan (1997) suggère une cinquantaine de schèmes argumentatifs, Katzav et Reed (2004) présente une centaine de schèmes argumentatifs, Walton, Reed et Macagno (2008) en distinguent une soixantaine et introduisent des schèmes de questionnement critique, Rigotti et Morasso (2010) proposent une arborescence compatible avec la théorie pragma-dialectique et impliquant 3 schèmes de haut niveau, 8 schèmes de niveau intermédiaire et un nombre indéterminé de schèmes de bas niveau.

Ainsi, bien qu'il y ait de nombreuses théories sur les schémas argumentatifs, chacune avec ses propres principes et caractéristiques distincts, Walton et Macagno (2016) soulignent que les schémas suggérés sont souvent difficiles à distinguer pour les experts, se différenciant parfois uniquement par de subtiles ressemblances ou airs de famille. Parmi ces différentes théories, certaines possèdent un petit nombre de schèmes, d'autres un grand nombre de schèmes, ce qui peut s'expliquer par le fait que la recherche est

motivée par différents objectifs dont, notamment, le besoin d'une théorie plus ou moins générale ou spécifique, ou encore le besoin d'une théorie plus ou moins normative ou descriptive. D'une part, le nombre de classes d'arguments nécessaire dépend du niveau de généralité escompté et, d'autre part, ce nombre dépend également du besoin de décrire en tout ou en partie le phénomène de l'argumentation. À cet effet, une théorie de l'argumentation est descriptive lorsqu'elle décrit les arguments performatifs (ce qui correspond aux arguments qui sont convaincants pour ce qui est de l'approche rhétorique) tandis qu'une théorie est normative lorsqu'on se restreint aux arguments rationnels (ou ceux rationnellement convaincants pour l'approche rhétorique). Walton et coll. (2008), par exemple, acceptent des schèmes classiques qui sont rationnellement convaincants, mais qui peuvent faire l'objet de critique à l'aide d'autres schèmes, et ils acceptent également des schèmes sophistiques non rationnellement convaincants comme l'argument de biais, l'argument par ignorance, l'argument de la pente glissante, etc. À notre connaissance, il existe relativement peu de modèles hiérarchiques descriptifs de l'argumentation sous forme de schèmes. Cette situation pourrait être due à un écart entre les modèles schématiques, qui tendent à être simplifiés, logiques et linéaires, et la nature réelle de l'argumentation, qui se caractérise souvent par sa complexité, sa sophistication et ses processus parallèles. Toutefois, de nombreux chercheurs s'efforcent de développer des réseaux de schémas argumentatifs qui devraient théoriquement permettre une meilleure compréhension du phénomène de l'argumentation, comme l'indiquent Macagno et coll. (2017).

4.2.1 Les applications computationnelles des modèles rhétoriques

Dans le domaine de la fouille d'argument, plusieurs chercheurs souscrivent à l'approche rhétorique. Les modèles les plus utilisés sont ceux d'Aristote (*logos, ethos, pathos, et kairos*), de Walton (1996) et Walton, Reed et Macagno (2008). Par exemple, le modèle aristotélicien est appliqué sur un corpus de débats parlementaires par Duthie et coll. (2016) ainsi que des essais persuasifs par Stab et Gurevych (2014a). Les modèles de Walton (1996) ou de Walton, Reed et Macagno (2008) sont appliqués sur des débats moraux par Lawrence et Reed (2016), des articles de journaux, des jugements et des débats par Reed (2006), des discussions sur Wikipédia par Schneider et coll. (2013), des débats politiques par Walton (2012), la base de données Araucaria qui est composée d'articles de journaux, de jugements et de débats par Feng et Hirst (2011), et des articles de blogue par Anand et coll. (2011).

Les schèmes argumentatifs sont aussi utilisés pour l'extraction d'arguments dans le domaine du droit, notamment pour classifier automatiquement des arguments dans les textes juridiques en utilisant des indicateurs de schème (Mochales Palau et Moens, 2009, 2011 ; Moens *et al.*, 2007).

4.3 Les modèles épistémologiques

Un avantage de l'approche épistémologique est de profiter des développements de la logique qui, même si elle fait l'objet de critiques quant à sa capacité à décrire adéquatement tous les phénomènes de l'argumentation, permet une modélisation informatique, partiellement du moins, et même une modélisation d'agents capables d'argumenter. Plusieurs modèles épistémologiques utilisent certaines propriétés de graphes pour analyser ou manipuler des arguments.

4.3.1 Les modèles épistémologiques à base de graphes

Historiquement, Whately (1854) est un des premiers penseurs de l'argumentation à proposer une modélisation sous forme de graphe afin de faciliter l'évaluation de la validité des arguments :

Many students probably will find it a very clear and convenient mode of exhibiting the logical analysis of the course of argument, to draw it out in the form of a Tree (Whately, 1854, p. 422).

Ensuite, Peirce (1909) propose une approche de la logique à l'aide de graphes (nommés graphes existentiels) dont le processus inférentiel est plus long et difficile comparativement à celui de la logique de premier ordre, mais qui favorise la compréhension et permet de mieux disséquer ou guider l'attention vers les différentes étapes du raisonnement :

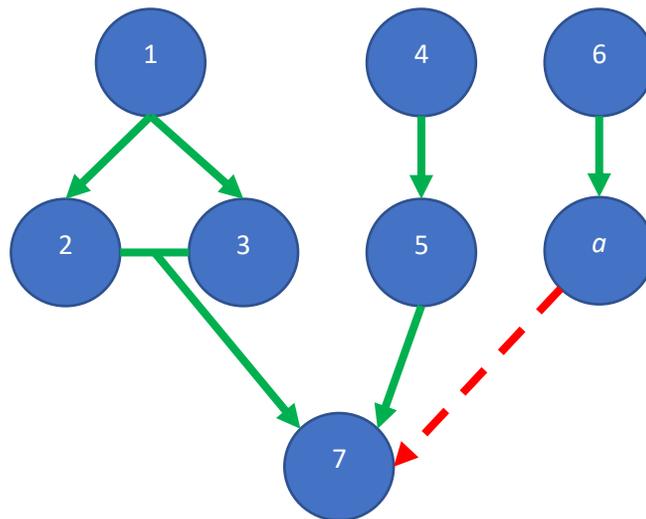
[...] I wish to draw your attention, in the most emphatic way possible, to the purpose this Syntax [of Existential Graphs] is intended to subserve: since anybody who did not pay attention to that statement would be all but sure, not merely to mistake the intention of this syntax, but to think that intention as contrary to what is as well he could. Namely he would suppose the object was to reach the conclusion from given premises with the utmost facility and speed, while the real purpose is to dissect the reasoning into the greatest possible number of distinct steps and so to force attention to every requisite of the reasoning. The supposed purpose would be of little consequence, and it is the fussiness of the mathematicians to furnish inventions to attain it; but the real purpose is to supply a real and crying need, although logicians are so stupid as not to recognize it and to put obstacles in the way of meeting it. (Peirce, 1909, p. 20-1)

Selon Peirce (1909), ces graphes sont non seulement plus compréhensibles et interprétables, leurs propriétés graphiques sont exploitables dans la mesure où le parcours inférentiel consistant à insérer, relier, annoter ou effacer des items dans le graphe semble faire émerger des conclusions nécessaires de manière intuitive, évidente ou automatique :

I invented several different systems of signs to deal with relations. One of them is called the general algebra of relations, and another the algebra of dyadic relations. I was finally led to prefer what I call a diagrammatic syntax. It is a way of setting down on paper any assertion, however intricate, and if one so sets down any premises, and then (guided by 3 simple rules) makes erasures and insertions, he will read before his eyes a necessary conclusion from premises. (Peirce, 1909, p. 11)

Subséquentement, Wigmore (1913) propose de représenter les arguments dans des graphes à l'aide de nombres (tel qu'exemplifié dans la figure 4.1 suivante).

Figure 4.1 Modèle rhétorique de Wigmore



Beardsley (1950) suggère de distinguer les convergences (dans la figure, les arguments #2, #3 et #5 convergent vers l'argument #7), les divergences (#1 diverge vers #2 et #3) et les séries (#4, #5 et #7 forment une série inférentielle). Thomas (1973) propose des liaisons horizontales (#2 et #3 partagent une relation horizontale). Scriven (1976) représente les prémisses implicites par une lettre (*a* correspond à une prémisses implicites) et il introduit les contre-arguments (*a* réfute #7). Enfin, Engelbart (1962) suggère l'implémentation informatique des graphes d'arguments :

To help us get better comprehension of the structure of an argument we can also call forth a schematic or graphical display. Once the antecedent-consequent links have been established the computer can automatically construct [a] network of lines and dots that looked something like tree (Engelbart, 1962, p. 87-8)

Éventuellement, le débat opposant l'approche à l'aide de graphes et celle de la logique classique réapparaît dans le domaine de l'informatique. D'une part, on développe des systèmes experts où les connaissances sont implémentées par un nombre limité de règles de forme si-alors, lesquelles permettent d'opérer des raisonnements et de tirer des conclusions à l'intérieur d'un domaine bien défini. Par exemple, on développe à l'université de Stanford un système expert par chaînage arrière (MYCIN) contenant environ 600 règles et permettant d'identifier certaines bactéries causant des infections ainsi que, le cas échéant, de suggérer un traitement approprié (Buchanan et Shortliffe, 1984). Or, contrairement aux systèmes fondés par des graphes, les systèmes fondés par la règle occultent la structure de connaissance sous-jacente et, ce faisant, peuvent rendre difficile pour la machine et le néophyte de distinguer ce qu'on peut inférer de ce qu'on devrait inférer :

These systems were, however, criticized for [...] that neither the system nor the non-expert human reader had any way to distinguish whether the encoded information was meaningful or not, so the entire search space had to be considered when reasoning (Knorr et Hitzler, 2014, p. 229).

D'autre part, on reproche aux systèmes fondés par des graphes d'être ambigus, notamment en ce qui a trait à la relation entre les concepts (Woods, 1975). Subséquemment, Hayes (1979) propose une approche utilisant la logique de premier ordre comme sémantique formelle de manière à désambiguïser les relations dans les graphes, notamment en définissant les relations entre classes conceptuelles à partir des relations binaires entre les individus qui les composent. Par rapport à la seule logique de premier ordre, il devient alors possible d'éviter les problèmes d'indécidabilité en utilisant des fragments décidables de la logique de premier ordre (Brachman et Levesque, 1985). Comme le font remarquer Knorr et Hitzler (2014), cette approche fondée sur la sémantique formelle mène au développement du système KL-ONE qui introduit les notions fondamentales des logiques de description (LD) dont, notamment, celles de concept (relation unaire ou classe), de rôle (relation binaire ou propriété), de TBox (*terminological box*), et de ABox (*assertional box*). Une LD de base nommée *Attributive Concept Language with Complements* (ALC) est proposée par Schmidt-Schauß et Smolka (1991), laquelle définit trois ensembles d'entités (individus, concepts et rôles) ainsi que des opérateurs logiques permettant de former les axiomes d'une ABox et d'une TBox. Les auteurs montrent que les raisonnements dans ALC sont décidables, mais parfois insolubles

en temps polynomial même lorsque les concepts formant les axiomes de la TBox sont acycliques (ne portent pas sur eux-mêmes). Parmi les différentes combinaisons possibles, certaines LD sont recommandées par le *World Wide Web Consortium* (W3C) dont, notamment, SHIF(D), SHOIN(D), SROIQ(D), lesquelles sont standardisées pour former des langages ontologiques compatibles avec d'autres langages Web (comme le *Resource Description Framework* ou RDF). SHIF(D), dont la complexité faible garantit la solution d'un raisonnement en un temps raisonnable, fonde une version légère du *Ontology Web Langage* (OWL-Lite). SHOIN(D) est une version plus complexe et décidable qui fonde le standard OWL-DL. SROIQ(D) fonde le standard OWL 2. Enfin, il existe une version complète indécidable du *Ontology Web Langage* – nommé OWL-Full – qui vise une plus grande compatibilité avec le standard RDF et son extension *RDF Schema* (RDFS). Plus précisément, le standard RDF vise à décrire les ressources Web et à permettre la réutilisation, l'échange et le traitement automatique de contenus. Dans ce formalisme, les connaissances sont représentées par des triplets constitués d'un sujet, d'un prédicat et d'un objet (exprimé par des nœuds, des *Internationalized Resource Identifier* (IRI) ou des littéraux), lesquels peuvent former des ensembles qu'on nomme graphes RDF. Par exemple, DBpedia est un graphe RDF organisant les connaissances collectées sur Wikipédia, laquelle est utilisée par plusieurs systèmes dont notamment IBM Watson (Chen *et al.*, 2020).

Bref, les premières modélisations épistémologiques de l'argumentation se présentent sous la forme de logique classique ou de logique graphique permettant des cartographies d'arguments où les nœuds peuvent correspondre à des arguments et les arêtes à des relations inférentielles logiquement valides (ou indéterminées). D'un point de vue computationnel, ces modèles peuvent s'implémenter sous forme de logique de programmation, d'ontologie ou de logique de description. D'ailleurs, plusieurs systèmes informatiques fondés par des graphes ou la logique de description exploitent avec succès les caractéristiques théoriques des graphes lors du processus de raisonnement de sorte que ces technologies sont maintenant adoptées par une majorité de grands joueurs comme Google (Knowledge Graph), Yahoo (Spark), Microsoft (Satori), Facebook (Entities Graph), etc. (Mika *et al.*, 2014) En d'autres termes, la recherche montre que ces modèles computationnels sont efficaces pour produire des raisonnements à partir d'ensemble de connaissances et, donc, qu'ils peuvent produire des arguments. Subséquemment, la recherche utilise ce type de modèle épistémologique à base de graphe pour la fouille d'argument, notamment pour produire des inférences pouvant être utilisées afin de guider la recherche textuelle. Cette intégration des connaissances ontologiques peut augmenter les performances, la précision et les taux de réussite dans diverses applications (Wotzlaw, 2010). Entre autres, le processus d'implication textuelle

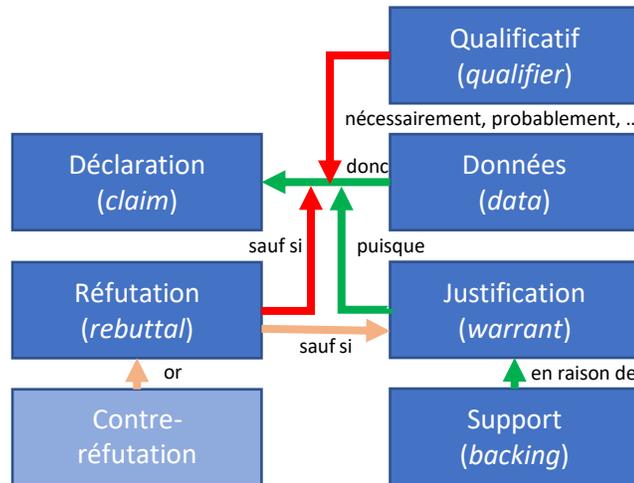
(*textual entailment*) peut être guidé par des hypothèses extraites d'une ontologie du domaine. Cette approche vérifie si la vérité d'une hypothèse découle d'un texte donné, ce qui est particulièrement utile dans des applications telles que l'évaluation automatisée des essais. Par exemple, dans Groza et Szabo (2015), l'implication textuelle est utilisée pour comparer la réponse d'un étudiant à une réponse modèle obtenue à partir d'une ontologie, améliorant ainsi le processus d'évaluation. Dans le domaine juridique, les outils de traitement automatique du langage naturel (TALN) sont utilisés pour extraire des informations telles que des entités nommées et des triplets syntaxiques pour construire des ontologies liées aux concepts et aux relations. Dans ce contexte, des éléments ontologiques sont associés à des règles juridiques pour tirer des conclusions sur des tâches telles que répondre à des questions (Fawei *et al.*, 2018). De plus, il est possible d'informer les ontologies existantes de nouvelles connaissances contenues dans des textes du domaine. Par exemple, dans l'analyse des débats sur le changement climatique, des mécanismes d'implication textuelle combinés à des ontologies sont utilisés pour identifier les meilleures réponses aux questions. Cette approche implique de traduire les axiomes de la logique de description à partir d'ontologies pour améliorer le processus de reconnaissance des implications (Szabo et Groza, 2017). Enfin, Rahwan et coll. (2007) ont développé une ontologie basée sur OWL qui montre comment les techniques d'inférence en logique descriptive peuvent être utilisées pour raisonner sur la classification automatique des arguments. Leur méthode d'identification des arguments a été implémentée dans un système web nommé Avicenna. En résumé, les ontologies jouent un rôle important dans divers domaines en fournissant des connaissances structurées, en guidant la formulation d'hypothèses et en améliorant la capacité des outils de TALN à tirer des inférences à partir du texte. Cette synergie entre les ontologies et diverses méthodes TALN est essentielle pour faire progresser les applications dans des domaines tels que l'analyse juridique, la notation automatisée et les technologies de l'information.

4.3.2 Les modèles épistémologiques informels

Très différents des modèles épistémologiques précédemment présentés, on peut considérer le modèle proposé par Toulmin (1958) comme un modèle épistémologique si on accepte un certain relativisme épistémologique. Exemplifié dans la figure suivante 4.2, on voit dans ce modèle que la relation inférentielle (représentée par une flèche verte) entre les données (*data*) et la déclaration (*claim*) est caractérisée par un qualificatif (*qualifier*) de manière à distinguer la nécessité, la probabilité, ou tout autre type de relation

épistémique (la relation entre le qualificatif et l'inférence est représentée en rouge, car le qualificatif peut restreindre la force de l'inférence).

Figure 4.2 Modèle épistémologique de Toulmin



Dans ce modèle, les données sont comme des prémisses qui fondent l'inférence et la déclaration (*claim*) correspond à la conclusion. Ce qu'on nomme justification soutient l'inférence de manière souvent implicite et correspond approximativement à ce que l'on nomme schème (ou lieu commun) dans l'approche rhétorique. Notons que la justification peut elle-même être justifiée par ce qui est nommé support (*backing*), lequel correspond à un certain arrière-plan épistémologique. Toulmin modélise aussi la réfutation (*rebuttal*) qui infirme la relation inférentielle. À titre d'exemple, la proposition « Socrate est un homme » est une donnée, « Socrate est mortel » est une déclaration, « tous les hommes sont mortels » est ce qui justifie l'inférence de l'un à l'autre, « personne n'a observé d'homme immortel » est ce qui supporte la justification et on peut ajouter le qualificatif « probablement » à la déclaration. De plus, on peut limiter la force de la déclaration en stipulant « sauf si Socrate est un personnage de fiction », ce qui constitue une réfutation. Éventuellement, Freeman (1991) apporte des amendements à ce modèle dont, notamment, la possibilité pour une réfutation d'infirmer la justification ainsi que la possibilité de contrer une réfutation par une contre-réfutation (*counter-rebuttal*). Enfin, Cohen (2001) propose de distinguer les inférences de premier ordre de celles de second ordre, c.-à-d. les méta-arguments. Cependant, notons que le modèle Toulmin-Freeman est ambigu dans la mesure où il peut être difficile de différencier les données et la justification. Par exemple, contre l'exemplification précédente, on peut concevoir que « tous les hommes sont mortels » est une connaissance comme « Socrate est un homme », et que ce qui justifie

la déclaration « Socrate est mortel » est le schéma inférentiel général « $P \rightarrow Q, P, \text{ donc } Q$ ». Enfin, le modèle épistémologique à base de graphe semble s'opposer à celui informel dans la mesure où le premier apparaît plus normatif et formel (voire syntaxique) comparativement au second qui apparaît plus descriptif et informel (voire sémantique).

4.3.3 Les applications computationnelles des modèles épistémologiques

Plusieurs chercheurs en fouille d'argument souscrivent à l'approche épistémologique. Les modèles utilisés sont ceux de Toulmin et de Toulmin-Freeman. Par exemple, le modèle de Toulmin est appliqué sur des commentaires d'utilisateurs, des interventions sur des forums de discussion et d'articles de journaux par Habernal et Gurevych (2017), des jugements de la Cour suprême des États-Unis par Newman et Marshall (1991), ainsi que des commentaires Web des lecteurs du New York Times par Habernal et coll. (2018). Le modèle de Toulmin-Freeman est appliqué sur des documents artificiels par Peldszus et Stede (2013b), des jugements de la Cour européenne des droits de l'homme par Mochales et Moens (2011), des essais persuasifs par Stab et Gurevych (2014b), des textes courts par Peldszus et Stede (2015) ainsi que Peldszus (2014).

4.4 Les modèles consensualistes

Généralement, un modèle est considéré comme monologique lorsqu'on s'intéresse plus spécifiquement à la structure interne des arguments ou lorsqu'on analyse séparément les différentes postures d'une argumentation. Ce faisant, les modèles monologiques ne permettent pas l'étude de la manière dont les arguments peuvent former des macrostructures plus complexes et, donc, ils ne modélisent que partiellement l'argumentation et, plus précisément, ils modélisent des arguments plutôt que des argumentations. Autrement, les modèles dialogiques s'intéressent aux relations entre les arguments ainsi qu'à la manière dont les arguments peuvent former des macrostructures. Par exemple, Dung (1995) ne s'intéresse pas à la structure interne des arguments entendue comme chaîne de raisonnement, mais à la relation entre les arguments, laquelle permet de les distinguer comme tel, ce qui permet de les définir :

Our theory of argumentation is based on a notion of argumentation framework defined as a pair of a set of arguments, and a binary relation representing the attack relationship between arguments. Here, an argument is an abstract entity whose role is solely determined by its relations to other arguments. No special attention is paid to the internal structure of the arguments (Dung, 1995, p. 325-6)

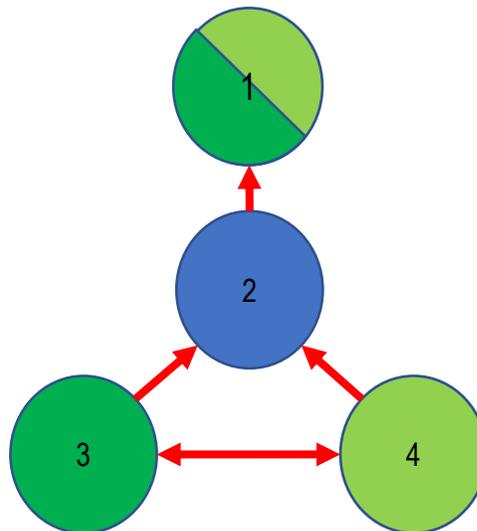
Autrement dit, ce modèle est fondé par les relations binaires qu'entretiennent les arguments qui sont des entités abstraites et dont le rôle est déterminé par ces relations et non pas leur structure interne ou leur origine. Généralement, l'ensemble des arguments et des relations binaires appartenant à un cadre d'argumentation (*argumentation framework*) est fini et, dépendamment des objectifs, on peut définir différentes sémantiques formelles fondées sur la notion de défense des arguments et de sous-ensemble sans conflit (un ensemble d'arguments qui n'attaquent pas d'autres arguments) :

Définition 1 (cadre d'argumentation). Un cadre d'argumentation est un couple (A, \rightarrow) où A est un ensemble fini (d'arguments) et \rightarrow est une relation binaire (d'attaque) définie entre A et A .

Définition 2 (défense). L'ensemble $S \subseteq A$ défend a si $\forall b \in A$ tel que $b \rightarrow a, \exists c \in S$ tel que $c \rightarrow b$.

Définition 3 (sans conflit). L'ensemble $S \subseteq A$ est sans conflit si et seulement s'il n'y a aucun $a, b \in S$ tel que $a \rightarrow b$.

Figure 4.3 Cadre d'argumentation



Dans la figure 4.3, on voit un cadre d'argumentation composé d'un ensemble d'arguments représentés par des nœuds et des relations d'attaque représentées par des arêtes orientées. Tel que défini par la sémantique formelle de Dung, l'ensemble $\{3,4\}$ défend l'ensemble $\{1\}$ et les ensembles $\{1,3\}$ et $\{1,4\}$ sont sans conflit, mais le modèle ne nous dit pas quelle extension privilégiée entre les deux.

Subséquentement, Amgoud et Vesic (2014) proposent un modèle d'argumentation basée sur la préférence où ils accordent différents poids aux arguments. Dans la figure précédente par exemple, si l'argument #3 était privilégié par rapport à l'argument #4 (en vertu de certains critères), alors l'ensemble {1,3} serait privilégié à l'ensemble {1,4}. Éventuellement, Bondarenko et coll. (1997) proposent un modèle d'argumentation basée sur l'assomption où les arguments possèdent une structure interne composée de prémisses, de règles, et d'une conclusion. Dans ce modèle, certaines prémisses sont des assomptions, c.-à-d. qu'elles ne proviennent pas d'un processus inférentiel logiquement valide et, donc, ce modèle est réductible à des assomptions, des règles et des relations d'attaque. De plus, il est possible d'utiliser différents systèmes logiques comme la programmation logique (Green, 1969), la logique par défaut (Reiter, 1980) ou la logique autoépistémique (Moore, 1985). Enfin, les modèles consensualistes d'inspiration dunguienne font face à quelques problématiques. D'abord, ils ne permettent pas de représenter les arguments s'attaquant eux-mêmes. Ensuite, plusieurs développements de la sémantique dunguienne fondés par les extensions formelles permettent d'envisager l'identification de sous-ensembles particulièrement intéressants parmi un cadre d'argumentation donné, mais la recherche montre que cette identification constitue un problème NP-complet (voir Dunne, 2008). Aussi, les modèles dunguiens qui n'explicitent pas les relations de support souffrent d'un manque similaire à certains des modèles épistémiques vus précédemment, c.-à-d. que certaines relations sont omises et, donc, que ces théories ne peuvent pas rendre compte du phénomène de l'argumentation dans toute sa complexité. Enfin, on remarque qu'il est possible de définir les arguments à l'aide de caractéristiques qui leur sont internes (formelles ou sémantiques) ou externes (formelles ou performatives).

4.4.1 Les applications computationnelles des modèles consensualistes

Les modèles dialogiques basés sur Dung permettent l'analyse, l'évaluation et la résolution de débats. Par exemple, le modèle de Dung est appliqué sur des débats issus de Debatepedia par Cabrio et Villata (2012a) ainsi que sur des débats issus de médias sociaux non spécifiés par Schneider et coll. (2012).

4.5 Les modèles communicationnels

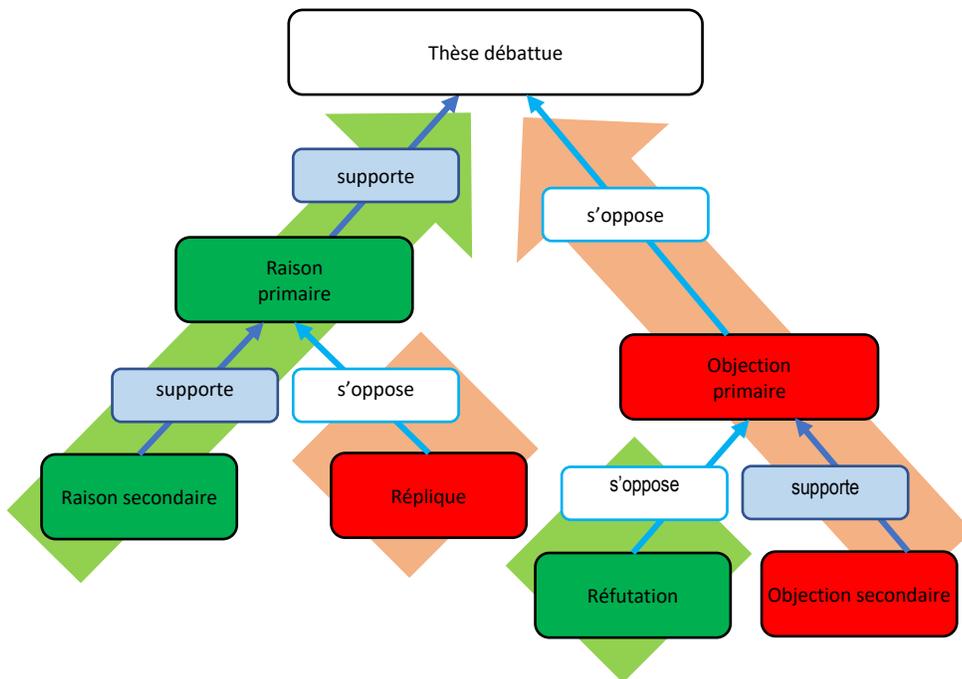
4.5.1 Les modèles communicationnels en sciences cognitives

L'hypothèse communicationnelle soutient que l'argumentation est un phénomène émergeant de la dynamique entre un émetteur et un récepteur. Du côté de l'émetteur, le motif sous-jacent l'acte communicationnel est la persuasion et, du côté du récepteur, le motif est épistémique au sens où on s'attend à un certain apport en information ou en connaissance. Subséquemment, le rôle du récepteur

diverge de celui de l'émetteur en ce que le premier vise et tend à développer des raisons de croire en quelque chose (une action, une déclaration, une conclusion, une thèse, etc.) et le second vise ou tend à vérifier la valeur épistémique de ladite chose, notamment en développant des raisons de ne pas y croire. Autrement dit, les arguments émanent du besoin concret de persuasion, les contre-arguments du besoin concret de vérité et on peut concevoir le consensus, tel que le soutiennent Habermas et Apel, comme l'idéal communicationnel *a priori* de l'argumentation, lequel se réalise parfois, ce qui d'un point de vue évolutionniste est suffisant pour avantager l'espèce.

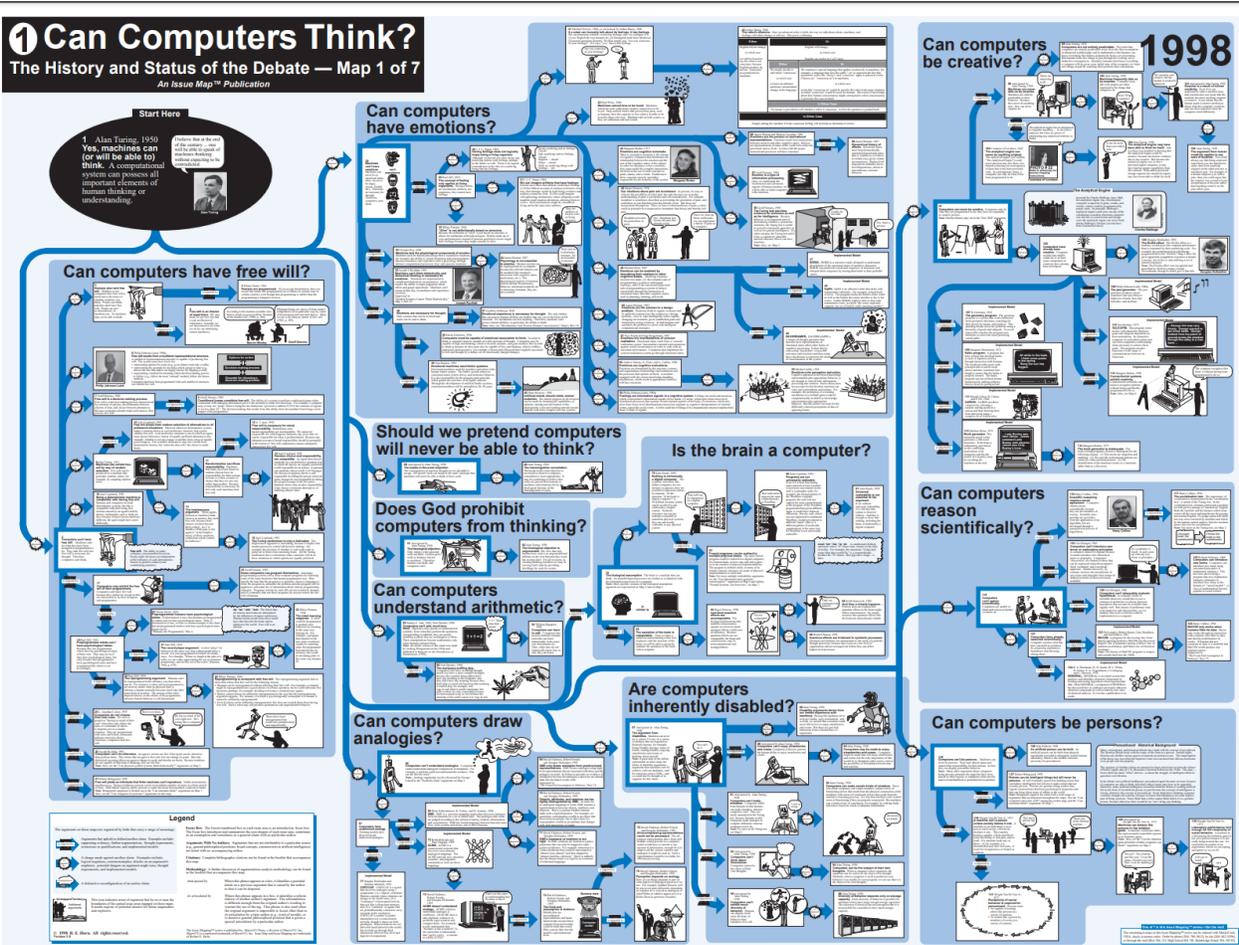
Dans la figure 4.4 suivante, on voit représentés dans des boites vertes au-dessus de zones vertes les éléments d'une argumentation émergeant du besoin rhétorique (les arguments ou, plus précisément, une raison primaire, une raison secondaire et une réfutation) et en rouge au-dessus de zones rouges les éléments émergeant du besoin épistémique (les contre-arguments ou, plus précisément, une objection primaire, une objection secondaire et une réplique).

Figure 4.4 Modèle communicationnel



Dans cet exemple, notons que seulement deux niveaux argumentaires sont représentés et que, en principe, il n'y a pas de limite au nombre de niveaux, c.-à-d. qu'il aurait été possible de représenter des raisons ou des objections tertiaires, quaternaires et ainsi de suite de manière incrémentale. Parallèlement, il serait possible de compléter de répliques ou de réfutations ces raisons ou ces objections de niveau supérieur à la manière d'une structure fractale où, à l'image des poupées russes, chaque niveau peut en principe en cacher un autre. Notons que les termes de « raison primaire », « raison secondaire », « objection primaire », « objection secondaire », « réfutation » et « réplique » réfèrent à des raisons qui sont en relation de support ou d'opposition à d'autres raisons et, ultimement, à une thèse principale. Comme le montrent Mercier et Sperber, concevoir les éléments de l'argumentation comme des raisons (plutôt que des arguments) permet d'explicitier le lien étroit (autrement implicite) entre une théorie de l'argumentation et une théorie du raisonnement. Autrement dit, une théorie de l'argumentation, même si elle est identifiée à une approche formelle (en tant que manipulation symbolique) ou linguistique (en tant qu'acte de langage), doit ultimement se fonder sur du cognitif. Enfin, spécifions que les modèles dunguiens augmentés des relations de support (comme celui de Bondarenko et coll., (1997)) peuvent s'interpréter comme des modèles communicationnels. La figure 4.5 suivante exemplifie de manière graphique la modélisation communicationnelle d'une argumentation à plusieurs niveaux de profondeur à propos d'une thèse originalement soutenue par Turing à savoir que les ordinateurs peuvent penser. Cet extrait produit par Horn (1998) montre à quoi pourrait ressembler la représentation des résultats d'une fouille d'arguments ou d'une LAATAO dans un corpus composé de textes de haut niveau théorique à propos d'une question donnée.

Figure 4.5 Modèle de Horn

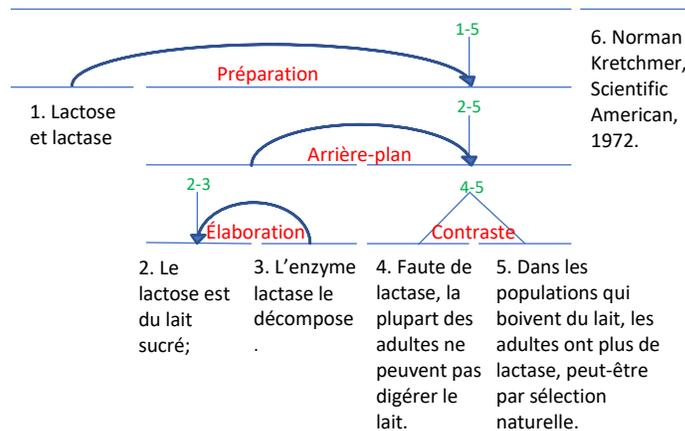


Source : Horn (1998)

4.5.2 Les modèles communicationnels en linguistique

Une modélisation communicationnelle est proposée par Mann et Thompson (1988), lesquels postulent que la grande majorité des discours sont surdéterminés par des fonctions communicationnelles.¹⁶ Inspiré des travaux de Grimes (1975), la théorie des structures rhétoriques (de l'anglais *Rhetorical Structure Theory* ou RST) postule que la plupart des textes sont composés d'unités élémentaires (UEs) fonctionnellement reliées pour former des étendues discursives nommées « segments », lesquelles se relient à d'autres UEs (ou segments). On postule deux types d'UEs ou de segments soit le nucléus caractérisé par son aspect central et indépendant ainsi que les satellites caractérisés par leur aspect périphérique et interdépendant. Les auteurs recensent 23 types de relation entre satellite et nucléus (ou nucléus-nucléus) qu'ils définissent en termes de contraintes et d'effets dont, par exemple, l'élaboration (*elaboration*) qui ajoute de l'information à une information plus fondamentale, le contraste (*contrast*) qui indique une alternative, l'évidence (*evidence*) qui agit comme support envers une déclaration, la préparation (*preparation*) qui annonce le thème de la communication, l'arrière-plan (*background*) énonçant les définitions ou les informations nécessaires à la compréhension du texte.¹⁷

Figure 4.6 Modèle RST



¹⁶ Comme le soulignent les auteurs, presque tous les textes peuvent faire l'objet d'une analyse au moyen de la RST, à l'exception de certains comme les textes de loi, les contrats, les rapports d'événement et certains types artistiques comme la poésie (Mann et Thompson, 1988, p. 259).

¹⁷ Les 23 relations identifiées par la RST sont : *circumstance, solutionhood, elaboration, background, enablement, motivation, evidence, justify, volitional cause, non-volitional cause, volitional result, non-volitional result, purpose, antithesis, concession, condition, otherwise, interpretation, evaluation, restatement, summary, sequence et contrast.*

Dans la figure 4.6, on voit un exemple prototypique de modélisation RST appliqué sur le titre ainsi qu'un extrait d'un article de Norman Kretchmer (1972, notre trad.). L'exemple analysé par Bill Mann est composé de six (6) UE, certaines reliées par une flèche montrant une relation orientée du satellite vers le nucléus et d'autres reliées par des lignes non orientées montrant une relation d'équivalence fonctionnelle entre un nucléus et un autre nucléus. On voit aussi que les relations peuvent se produire entre des groupes d'UE qui forment ainsi des segments prenant le rôle de nucléus ou de satellite. Parce qu'elles ne sont jamais en relation satellitaire, on peut considérer que les UEs #4 et #5 sont absolument centrales et indépendantes ou, pour la même raison, que le segment « 4-5 » composé des UEs #4 et #5 est absolument central et indépendant. Parallèlement, on voit que l'UE #3 est un satellite élaborant l'UE #2 qui est un nucléus central et indépendant, mais relativement dans la mesure où elles forment le segment « 2-3 » qui constitue l'arrière-plan de « 4-5 ». Ce faisant, cette relation d'arrière-plan relie les segments « 2-3 » et « 4-5 » pour former le segment « 2-5 », lequel constitue un nucléus relativement à l'UE #1 qui, en tant que titre de l'article, prépare la communication. Enfin, on pourrait considérer que le segment « 1-5 » constitue un nucléus relativement à l'UE #6 qui indique que le discours est attribuable à Norman Kretchmer et qu'il a été publié en 1972 dans le journal *Scientific American*. Éventuellement, Teufel, Carletta et Moens (1999) vont introduire des relations globales au modèle RST. De plus, la recherche montre que les types de relation RST sont corrélés à des marqueurs de surface (Corston-Oliver, 1998 ; Taboada et Mann, 2006) et que les UEs et segments entretiennent des relations globales pouvant correspondre à des thématiques (Skoufaki, 2020).

Un problème important soulevé par la communauté scientifique à propos de la RST est que l'ensemble des 23 relations initialement identifiées semble insuffisant pour rendre compte de toutes les structures rhétoriques importantes (voir Nicholas, 1994). Par exemple, le modèle original omet certaines relations comme l'attribution dont, notamment, l'attribution d'un texte à son auteur ou l'attribution d'une UE ou d'un segment particulier à quelqu'un d'autre ou à un autre texte (Potter, 2019). Stede et coll. (2017) suggèrent une version révisée de la RST dans laquelle trois types de modifications sont apportées. D'abord, certaines relations jugées redondantes sont éliminées.¹⁸ Ensuite, certaines relations ne sont plus envisagées à sens unique, c.-à-d. que si les UEs ou segments qui constituent le nucléus et ceux qui constituent le satellite sont toujours les mêmes dans une relation à sens unique, les rôles peuvent être inversés lorsque cette relation n'est pas à sens unique. Par exemple, la relation d'évaluation (*evaluation*)

¹⁸ Les relations *volitional cause* et *non-volitional cause* sont remplacées par une seule relation nommée *cause*. Les relations de *volitional result* et *non-volitional result* sont remplacées par une seule relation nommée *result*.

originellement théorisée prédit que l'évaluation est toujours un satellite de ce qui est évalué. Or, soutiennent Stede et coll. (2017), l'évaluation peut parfois constituer le nucléus de ce type de relation et, le cas échéant, ce qui est évalué est satellitaire.¹⁹ Enfin, les théoriciens postulent huit relations en plus de celles fournies par le modèle original.²⁰ Nous produisons en annexe A toutes les définitions des relations rhétoriques du modèle RST révisé proposé par Stede et coll. (2017) auquel on adjoint les définitions rendant compte des huit types de relation d'attribution identifiés par Potter (2019). Spécifions que les relations sont regroupées en cinq espèces, c.-à-d. les relations pragmatiques, les relations sémantiques, les relations textuelles, les relations multinucléées et les relations d'attribution. Les relations pragmatiques visent à changer l'attitude des lecteurs. Les relations sémantiques servent à décrire le monde. Les relations textuelles sont utiles dans l'organisation du texte et sa compréhension. Les relations multinucléées ont la particularité de contenir plusieurs nucléus et, selon le contexte, leur fonction peut être pragmatique, sémantique ou textuelle (à l'exception de la relation séquentielle qui est une fonction multinucléée proprement sémantique). Enfin, lorsque l'effet principal d'une relation d'attribution est effectivement l'attribution, alors sa fonction est textuelle. Lorsque l'effet principal de l'attribution est identique à celui défini par certaines relations pragmatiques ou sémantiques, alors la relation d'attribution est pragmatique ou sémantique.

4.5.3 Les applications computationnelles des modèles communicationnels

La modélisation communicationnelle inspirée de Sperber et Mercier est utilisée en analyse d'argument (Kiesel *et al.*, 2015) ainsi que dans le domaine de la cartographie des arguments, car elle est reconnue comme facilitant l'interprétation et l'apprentissage (Kirschner *et al.*, 2002 ; van Gelder, 2001). Elle est utilisée pour modéliser des argumentations complexes de haut niveau théorique et offre un modèle de représentation facilitant l'interprétation, mais elle est peu utilisée en fouille d'argument. Cette approche offre un cadre théorique compatible avec plusieurs des autres approches et elle est ancrée par les sciences cognitives dans la mesure où elle postule que le but premier du raisonnement est l'argumentation. Parallèlement, la modélisation communicationnelle inspirée de Mann et Thompson fait le lien entre le discours et l'argumentation en postulant que, fondamentalement, le discours est argumentation et, donc, que toutes les constituantes du discours sont des constituantes de l'argumentation. Cependant, la recherche montre qu'il est difficile d'identifier les types de relation définie par la RST, notamment l'effet

¹⁹ Une autre relation permet l'interchangeabilité des rôles de satellite et de nucléus. Elle se nomme *reason*.

²⁰ Les nouvelles relations se nomment *reason, unless, e-elaboration, means, preparation, list, conjunction* et *joint*.

(ou l'intention) d'une partie du discours. Il est également difficile d'identifier les unités élémentaires, c.-à-d. qu'on n'a pas de procédure claire indiquant où s'arrêtent les divisions du discours. Enfin, une autre limite à la RST s'exprime par une tension entre la modélisation du discours comme objet général et le discours comme objet spécifique. Elle se manifeste notamment par la difficulté des chercheurs à appliquer la RST à un domaine discursif particulier, lesquels doivent parfois adapter le modèle à leur domaine.

En somme, la théorie communicationnelle de l'argumentation encadre souvent de manière plus ou moins explicite les modélisations et applications issues des autres approches. Ce faisant, il est possible de considérer toutes les applications citées précédemment comme modélisant, partiellement du moins, l'argumentation comme phénomène de communication. Quant à eux, les modèles communicationnels d'inspiration dunguienne sont peu utilisés en fouille d'argument, car l'ajout de la relation de support au modèle original se traduit par une augmentation de la complexité computationnelle. Enfin, certains modèles communicationnels mettant en jeu des schèmes de niveau discursif sont utilisés en fouille d'argument, et peuvent être appliqués sur des *corpora* de revues d'hôtels, de revues de produits et de débats (García Villalba et Saint-Dizier, 2012b), sur des commentaires de journaux allemands (Peldszus et Stede, 2013a), et sur des articles scientifiques (Teufel *et al.*, 1999). Notons que ces exemples d'application en fouille d'argument montrent que ces modèles permettent de réaliser d'autres objectifs connexes comme la fouille d'opinion, la cartographie d'arguments et le résumé. Parallèlement, Hougbo et Mercer (2014) montrent comment des marqueurs discursifs permettent d'identifier certains schèmes argumentaires dans les articles scientifiques biomédicaux tandis que Lawrence et Reed (2016) montrent comment des structures argumentatives de haut niveau sont corrélées à des thèmes dans un corpus de débats moraux de la BBC (BBC Moral Maze).

4.6 Les *corpora* sur lesquels sont appliqués les différents modèles

Les applications des modèles d'arguments à la fouille d'arguments impliquent deux tâches importantes, soit l'extraction de composantes argumentatives (ou arguments) et l'identification des relations entre ces composantes. Selon le modèle théorique sous-jacent, soit le discours est entièrement considéré comme argumentatif et tout partitionnement raisonnable produit des composantes argumentatives qu'on peut ensuite classer en vertu de certains types suggérés par le modèle sélectionné, soit le discours est partiellement considéré comme argumentatif de sorte qu'une sous-tâche d'identification des composantes proprement argumentatives est nécessaire avant même de pouvoir les classer selon leur type respectif. L'extraction de composantes argumentatives (ou arguments) implique donc le

partitionnement du discours ainsi que la classification des partitions discursives. Selon le modèle théorique sélectionné, l'identification des relations entre les composantes argumentatives (ou arguments) peut impliquer la détection de relations hétérogènes comme celle de support et d'opposition, de relations hiérarchiques comme celle entre les prémisses et la conclusion, de relations internes comme celle qualifiant le lien inférentiel ou de relations externes comme celle méta-argumentative (Stab et Gurevych, 2017). De manière classique en informatique, les techniques computationnelles de partitionnement et de classification développées et utilisées en fouille d'argument sont testées et évaluées à l'aide de corpus annotés. Un premier problème à cette entreprise est la difficulté d'annoter des *corpora* et, *a fortiori*, celle de trouver des *corpora* annotés soit dans la langue ou le domaine discursif qui nous intéresse, ce qui mène à des outils spécialisés de fouille, soit d'un niveau de généralité suffisant, ce qui mène à des outils généraux de fouille. Un second problème est l'absence de consensus à propos du modèle théorique à implémenter, laquelle se manifeste par une pluralité de modèles ainsi que de protocoles d'annotation et, subséquemment, une incompatibilité entre certains *corpora* annotés et certains modèles théoriques. Comme vu précédemment, les *corpora* sur lesquels les modèles rhétoriques sont évalués sont majoritairement constitués d'articles de journaux, de jugements, de débats, de discussions (blogue, talks, etc.) et d'essais persuasifs; ceux sur lesquels les modèles épistémologiques sont évalués sont majoritairement constitués de commentaires, de jugements, d'essais, de débats, d'articles de journaux et de discussions; ceux sur lesquels les modèles consensualistes sont évalués sont majoritairement constitués de débats (médias sociaux); et ceux sur lesquels les modèles communicationnels sont évalués sont majoritairement constitués de revues (évaluation de service ou de produit), de commentaires, de débats et d'articles scientifiques. Spécifions également que la majorité des corpus sont en langue anglaise.

4.7 Conclusion

En somme, la majorité des modèles postulent des unités élémentaires, mais il n'y a pas de consensus sur leur étendue, leurs limites, ni leur fonction argumentative (ou non). En général, les unités élémentaires peuvent se relier pour former des structures complexes et, habituellement, on admet la distinction entre des unités qui supportent (raison, prémisse, preuve, évidence, etc.) et des unités qui sont supportées (thèse, conclusion, déclaration, etc.) ou, inversement, des unités qui s'opposent et des unités qui sont opposées. Les structures complexes peuvent se distinguer selon certains critères (*topoi*, *loci*, schèmes, types, thèmes, etc.) et peuvent se relier entre elles, mais il n'y a pas de consensus sur ces critères ou ces relations. De plus, entre la théorie et la pratique, la recherche se confronte à des besoins plus ou moins compatibles tels l'objectif d'un modèle général et la réalité des domaines spécifiques (et des langues

particulières), le besoin d'un modèle normatif et celui d'une description de la réalité, l'utilité d'un modèle formel (syntaxique ou de surface) et la réalité possiblement plus informelle (sémantique ou conceptuelle), le besoin d'un modèle simple computable en temps raisonnable et la complexité du phénomène réel de l'argumentation.

CHAPITRE 5

La modélisation computationnelle de la tâche d'analyse d'argument

Dans ce chapitre, nous présentons une modélisation computationnelle générale de la tâche d'analyse d'argument au moyen d'un certain formalisme informatique. La tâche d'analyse d'argument est décomposée en une suite d'opérations dont certaines sont exécutables par ordinateur (en plus d'être exécutables cognitivement ou « à la main », bien que cela pourrait prendre beaucoup de temps) et d'autres sont exécutables cognitivement (et non pas par ordinateur à notre connaissance).

5.1 Introduction

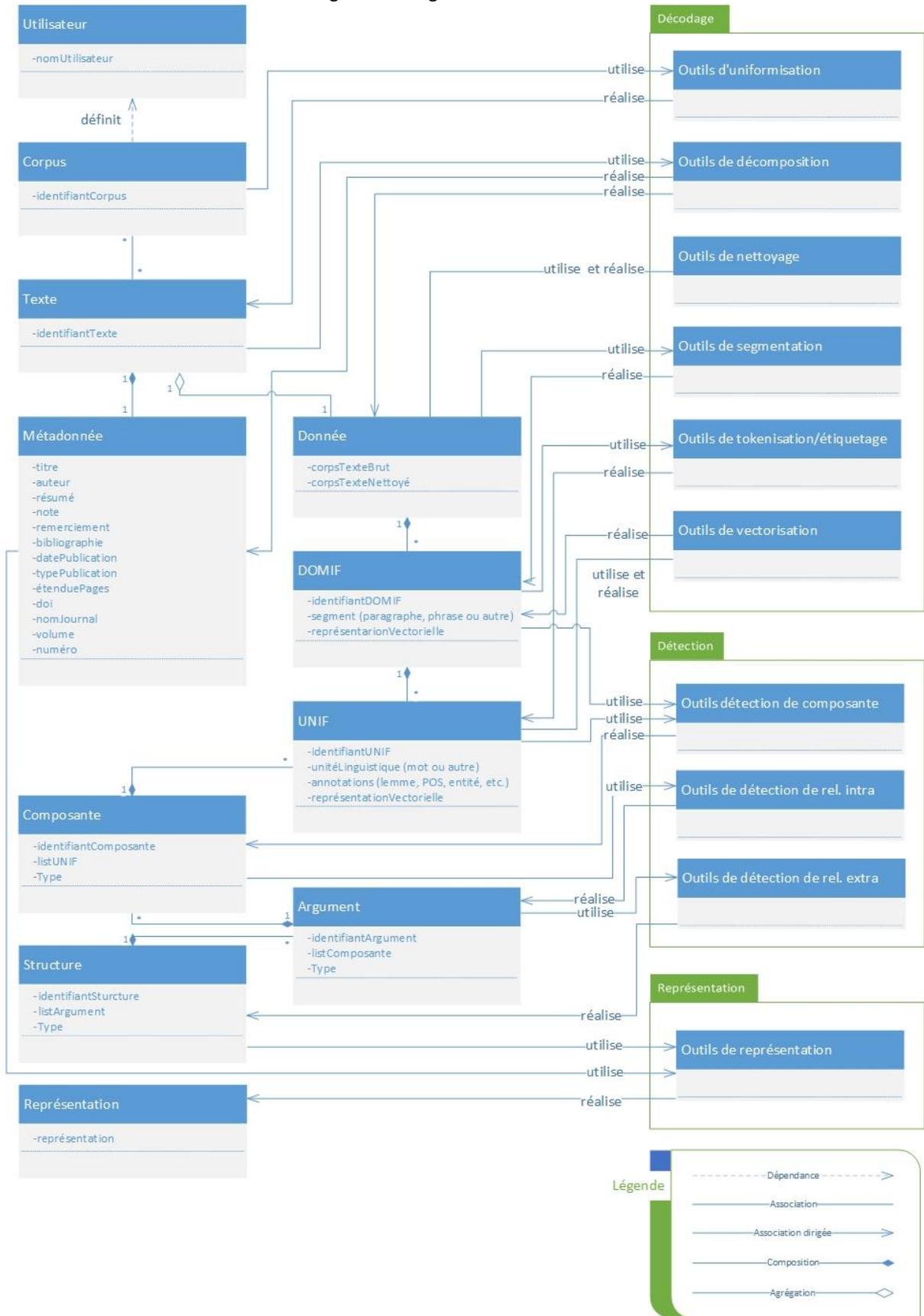
L'analyse d'arguments est une activité exécutée par des systèmes complexes, qu'ils soient biologiques ou informatiques. Selon Marr (1982), Pylyshyn (1984) et Meunier (2017, 2019), de tels systèmes peuvent être examinés selon une approche fonctionnelle comportant au moins trois dimensions de description ou niveaux de modélisation. Un premier permet de déterminer la fonction générale du système, son objectif et les concepts en jeu. Un deuxième se concentre sur la manière dont le système accomplit sa fonction, comment elle peut être instanciée par des opérations mathématiques ou algorithmiques, ou par quel traitement de l'information on peut produire certains résultats. Un troisième permet de décrire les structures physiques ou biologiques qui permettent au système d'effectuer sa fonction. Dans le cas de l'analyse d'argument, le système complexe effectuant cette tâche vise à découvrir les arguments qui sont exprimés à travers un ensemble de textes qui composent un corpus. Une description computationnelle détaille les opérations pour cette tâche, et une description physique explique comment ces opérations sont réalisées physiquement. Cependant, si différents médiums peuvent réaliser les mêmes fonctions, il n'est pas nécessaire de décrire le système au niveau physique. Nous adoptons une approche générale et systématique pour modéliser cette tâche d'analyse, inspirée des travaux de Lareau (2022) et Kaiser et coll. (à paraître), utilisant le formalisme UML (de l'anglais : *Unified Modeling Language*) pour une description de deuxième niveau cohérente avec diverses théories de l'argumentation. Dans le cadre du développement logiciel et de la programmation orientée objet, le langage de modélisation unifié UML aide à structurer et décomposer un logiciel en classes et visualiser les objets impliqués. Habituellement, UML utilise des représentations distinctes pour les objets et les processus, mais dans notre cas, nous utilisons un diagramme de classe pour représenter à la fois les objets et les processus, une approche compatible avec la programmation orientée objet. La raison est que notre système est composé d'outils empilés (*stacked*), où la sortie de l'un constitue l'entrée de l'autre, permettant ainsi de représenter à la fois les

objets et les processus. Les outils performant différentes sous-tâches sur les objets, les modifient ou gèrent de nouveaux objets qui sont conceptuellement liés. Les différents outils sont regroupés en paquets (*packages*) lorsqu'ils performent conjointement une opération plus générale. Bref, la description UML proposée représente visuellement et de manière quasi formelle un système d'analyse d'argument facilitant, en principe, la compréhension des objets et des processus en jeu. Similairement à Fayyad et coll. (1996), nous décomposons la tâche d'analyse d'arguments en cinq (5) opérations générales elles-mêmes décomposables en sous-opérations. La première opération consiste à définir le corpus. La seconde vise à décoder le corpus de manière à structurer les textes étudiés. La troisième vise la détection des structures argumentatives. La quatrième vise la représentation des résultats. La dernière vise l'interprétation et l'évaluation des résultats.

5.2 La décomposition de la tâche d'analyse d'argument

Nous présentons à la figure 5.1 suivante un diagramme de classes d'un système de fouille d'argument ou LAATAO. Ces classes sont représentées par des encadrés de couleur bleue. Trois types d'objets sont représentés par ces encadrés de couleur bleue, soit l'utilisateur qui représente l'expert utilisant le système de fouille d'argument ou LAATAO, les classes d'objets manipulés par le système et les classes d'objets qui sont des outils permettant de manipuler ces dernières. Ces outils représentent les différents algorithmes en jeu et sont regroupés dans des encadrés de couleur verte pour former des *packages* qui correspondent à trois (3) des cinq (5) étapes de la tâche d'analyse d'argument, soit le décodage du corpus, la détection des structures argumentatives et la représentation des résultats.

Figure 5.1 Diagramme de classes



5.2.1 La définition du corpus

Dans le diagramme de classes, la première étape – la définition du corpus – est représentée par une relation hachurée et orientée (--->) de l'objet 'Corpus' vers l'objet 'Utilisateur' indiquant que le premier est en relation de dépendance par rapport au second, c.-à-d. que l'objet 'Corpus' est dépendant de sa définition préalable par 'Utilisateur'. Autrement dit, l'exécution du système de fouille d'argument ou de LAATAO sur un jeu de données textuelles nécessite préalablement le choix d'un ensemble de documents. En apparence très simple dans cette modélisation de l'analyse d'argument, la définition du corpus est une opération complexe et cruciale en linguistique (Biber *et al.*, 1998), FT (Ignatow et Mihalcea, 2016), AM (Pustejovsky et Stubbs, 2012), TALN (Manning et Schütze, 1999) et HN (Hockey, 2001). Spécifions que l'étape de définition du corpus est dépendante des objectifs de la recherche et que ces objectifs peuvent être très différents d'un domaine de recherche à l'autre.²¹ De plus, d'autres considérations pragmatiques peuvent entrer en jeu lors de la définition du corpus, notamment l'accessibilité des documents, le droit d'utilisation pour la recherche, la composition statistique, la pertinence, la cohérence, etc. Enfin, notons que la collecte des documents est une sous-étape de la définition du corpus qui n'est pas modélisée par notre système de fouille d'argument ou LAATAO et cette sous-opération peut faire l'objet d'une automatisation.²²

5.2.2 Le décodage du corpus

Lors de la deuxième étape – le 'Décodage' – le 'Corpus' référant à un certain nombre de documents est utilisé par 'Outils d'uniformisation' afin de réaliser des 'Textes' uniformes.²³ Chaque 'Texte' est ensuite

²¹ En général, différents domaines de recherche tendent à étudier différents objets. À titre d'exemple, en linguistique, on peut viser l'étude de certains usages linguistiques; en FT, on peut chercher à extraire des connaissances; en TALN, on peut viser le développement de certaines applications; en HN, on peut chercher une meilleure compréhension d'un corpus donné. Subséquemment, on comprend que le corpus étudié est parfois un objet d'étude en soi, parfois un moyen d'étudier autre chose comme une méthode, une application ou un phénomène plus large dont le corpus n'est qu'un échantillon.

²² Notons que le corpus est ici défini comme un ensemble de documents, et que n'importe lequel de ces documents pourrait en principe être membre d'un autre corpus. De plus, plusieurs techniques de moissonnage permettent l'automatisation de la collecte de documents pour former un corpus. Par exemple, ces techniques ont été utilisées lors de la collecte automatisée des documents d'un large corpus de philosophie des sciences (Malaterre *et al.*, 2021 ; Malaterre et Lareau, 2022).

²³ Certains documents peuvent être de formats manuscrits ou informatiques. Les premiers nécessitent une opération de reconnaissance des caractères de manière à transformer le manuscrit en un format permettant un traitement informatique. Or, un manuscrit peut être représenté par différents encodages ou formats informatiques. Par exemple, un même caractère manuscrit peut être représenté par différents caractères

utilisé par 'Outils de décomposition des textes' pour extraire la 'Donnée' et la 'Métadonnée' associées à chaque 'Texte'. La 'Donnée' représente l'objet textuel investigué par le système, typiquement le corps du texte de chaque 'Texte', tandis que la 'Métadonnée' concerne les informations relatives à la 'Donnée', souvent qualifiées de « a priori » (Lebart et Salem, 1994), telles que le titre, le(s) auteur(s), le résumé, les notes, les remerciements, les références, la date de publication, etc.²⁴ La 'Donnée' de chaque 'Texte' est ensuite utilisée par 'Outils de nettoyage' pour réaliser une donnée nettoyée où l'objet textuel investigué est exempt de toute information périphérique comme, à titre d'exemple, la répétition du titre à chaque page, la bibliographie, les remerciements, etc. Chaque 'Donnée' nettoyée est ensuite utilisée par 'Outils de décomposition des textes' pour réaliser des DOMaines d'InFormation – 'DOMIF' – c.-à-d. des segments de texte correspondant à des paragraphes, des phrases ou toutes autres segmentations possibles.²⁵ Chaque 'DOMIF' est utilisé par 'Outils d'étiquetage' pour décomposer les 'DOMIFs' en Unités d'InFormation – 'UNIF' – habituellement des mots, et réaliser l'étiquetage de chaque 'UNIF' selon sa fonction dans la phrase (sa catégorie grammaticale), ce qui permet aussi d'annoter chaque 'UNIF' de son lemme qui correspond à l'entrée du mot dans le dictionnaire. De plus, l'outil de vectorisation utilise les 'UNIFs' afin de réaliser une représentation vectorielle des 'DOMIFs' et des 'UNIFs'. Plus précisément, le 'Corpus' est composé de 'Textes' dont chacun est associé à une 'Métadonnée' et une 'Donnée', laquelle est composée d'unités nommées 'DOMIFs' qui sont représentées par des vecteurs, elles-mêmes composées d'unités nommées 'UNIFs' également représentées par des vecteurs. Spécifions que cette étape 'Décodage' est partagée avec d'autres opérations de lecture et d'analyse de texte assistées par ordinateur comme l'analyse thématique et l'analyse conceptuelle.

5.2.3 La détection des structures argumentatives

Lors de la troisième étape – la 'Détection' – les 'DOMIFs' et/ou les 'UNIFs' sont utilisées par 'Outils de détection de composante' afin d'identifier les 'Composantes' argumentatives, lesquelles sont utilisées par

informatiques et un même manuscrit peut être encodé au moyen de différents formats informatiques comme PDF, DOCX, XML, HTML, etc.

²⁴ Typiquement, les éléments d'un document qui ne font pas partie intégrante du contenu principal sont classés comme des métadonnées. Par exemple, le corps du texte est souvent vu comme une donnée essentielle, tandis que d'autres parties du document, telles que le titre ou le résumé, peuvent être considérées comme des métadonnées. Cependant, il existe des situations où cela est inversée, et des éléments comme le titre ou le résumé sont traités comme des données clés, alors que le corps du texte n'est pas considéré comme tel. Cette approche peut être motivée par des raisons théoriques ou des considérations pratiques.

²⁵ La segmentation du texte est optionnelle et motivée par des considérations théoriques et/ou pragmatiques. Dans le cas où la segmentation du texte n'est pas nécessaire, la donnée nettoyée correspond au DOMIF.

‘Outils de détection de relation intra’ afin d’identifier des relations entre ‘Composantes’ argumentatives formant ainsi des ‘Arguments’ unitaires, lesquels sont utilisés par ‘Outils de détection de relation extra’ afin d’identifier les relations entre ‘Arguments’ formant ainsi des ‘Structures’ argumentatives. Ainsi, on comprend qu’une ‘Structure’ argumentative est composée d’un ensemble de ‘Arguments’ et qu’un ‘Argument’ est composé d’un ensemble de ‘Composantes’ argumentatives. Selon le modèle d’argumentation, notons qu’une composante argumentative peut correspondre ou non à une phrase et qu’il est possible de considérer tout le discours comme étant argumentatif.

5.2.4 La représentation des résultats

Lors de la quatrième étape – la ‘Représentation’ – les ‘Structures’ argumentatives ainsi que les ‘Métadonnées’ sont utilisées par ‘Outils de représentation’ afin de réaliser une ‘Représentation’ des résultats. La représentation des résultats est effectuée à l’aide d’outils de représentation qui rendent possible l’exploitation des résultats. Ces outils permettent de créer des visualisations synthétiques des résultats de la recherche.

5.2.5 L’interprétation et évaluation des résultats

La cinquième et dernière étape – l’interprétation et l’évaluation des résultats – n’est pas une étape exécutée par le système de fouille d’argument ou LAATAO. Cette étape est exécutée cognitivement par le chercheur. L’interprétation des résultats fait référence au processus d’analyse et de compréhension des données obtenues à partir d’expériences ou d’études de recherche. Elle implique l’examen des données collectées, l’identification des régularités (*patterns*) et des inférences significatives sur la base des preuves recueillies. L’interprétation des résultats est une étape cruciale, car elle permet aux chercheurs de donner un sens à leurs découvertes et de comprendre les implications de leurs travaux. Cela implique d’appliquer une analyse statistique, d’évaluer la signification des résultats et de les comparer aux connaissances ou aux hypothèses existantes. Au cours du processus d’interprétation, les scientifiques peuvent rechercher des corrélations, des causes ou des relations entre les variables et envisager des explications alternatives. Ils peuvent également discuter des limites de leur étude, aborder tout résultat inattendu ou contradictoire et proposer de futures directions de recherche basées sur leurs découvertes.

5.3 Conclusion

La décomposition de la tâche d’analyse d’argument en une suite d’étapes afin de modéliser un système automatique de fouille d’argument ou LAATAO montre qu’un tel système est en fait un outil informatique

d'assistance à l'analyse d'argument. La raison principale est que des étapes cruciales, en amont et en aval de la fouille d'argument ou LAATAO, ne sont pas modélisées par un tel système, notamment l'étape de définition du corpus et l'étape d'interprétation et de validation des résultats. Une autre limite de notre modèle est son incapacité à tenir compte de tous les objectifs possibles de la recherche. Néanmoins, notre modélisation de la tâche de fouille d'argument ou LAATAO peut être adaptée aux différents objectifs de recherche, notamment en ajoutant des étapes ou des sous-étapes nécessaires à l'atteinte de ces objectifs. Par exemple, notre modèle ne tient pas compte du travail considérable qui peut être nécessaire au traitement de la métadonnée lorsque celle-ci est bruitée, partielle ou absente. Lorsque les objectifs de la recherche impliquent l'usage d'une métadonnée non bruitée, des étapes de collecte et/ou de nettoyage peuvent être nécessaires (voir Kaiser *et al.*, à paraître). De plus, les objectifs de la recherche peuvent nécessiter une métadonnée structurée et, le cas échéant, des étapes de structuration de la métadonnée peuvent être nécessaires. Enfin, notre modèle définit la donnée comme étant le corps du texte, mais cette définition est matière à débat. Par exemple, la recherche montre qu'il est possible d'appliquer la fouille d'argument sur des résumés et, ce faisant, le résumé peut constituer une donnée.

CHAPITRE 6

Implémentation de la tâche de détection des arguments

Au chapitre précédent, nous avons présenté un système de fouille d'argument ou LAATAO dans lequel trois (3) étapes importantes sont modélisées. Dans ce chapitre, nous nous intéressons plus spécifiquement à l'étape de détection des structures argumentatives, c.-à-d. la tâche de détection des unités élémentaires et des relations propres à l'argumentation. D'abord, le matériel sur lequel peut s'effectuer cette détection est présenté, c.-à-d. les représentations vectorielles classiques et par plongement lexical, notamment word2vec, doc2vec, BERT et ChatGPT. Ensuite sont présentées différentes approches de détection et de représentation des structures argumentatives. Enfin, nous présentons une approche qualitative et une autre quantitative permettant l'évaluation de la détection des structures argumentatives ainsi que le matériel nécessaire pour effectuer cette tâche.

6.1 Introduction

L'étape de détection des structures argumentatives implique la détection des composantes argumentatives et des relations argumentatives (CARA) qui permettent de former des arguments et des structures argumentatives. La recherche en fouille d'argument montre que plusieurs outils informatiques permettent la détection des CARA à l'aide, notamment, d'approches inductives comme l'apprentissage machine, d'approches déductives comme celles à base de règles et d'approches hybrides. Puisque cette tâche suit celle du décodage du corpus, elle est appliquée sur une représentation du corpus se présentant habituellement sous forme vectorielle et permettant le traitement computationnel, mais il existe plusieurs représentations possibles du corpus.

6.2 Les différentes représentations vectorielles du texte

Notre travail de recherche explore une gamme étendue de représentations vectorielles du texte, allant des plus classiques aux plus sophistiquées. Cette diversité de représentations est importante pour comprendre certaines nuances dans le traitement et l'analyse des textes. Voici une présentation détaillée de ces représentations et du processus permettant leur création :

6.2.1 Les représentations vectorielles classiques

En FT, la représentation classique des textes est le sac de mots ou BOW (de l'anglais : *bag of words*) où chaque texte ou segment de texte (section, paragraphe, phrase ou toute autre segmentation) est représenté par un vecteur. Dans ce modèle proposé par Salton (Salton, 1971 ; Salton et Lesk, 1965), chaque composante du vecteur correspond à une unité linguistique atomique qui est dénombrée, c.-à-d. que l'on compte le nombre d'occurrences de chacune d'entre elles dans chaque texte ou segment de texte du corpus. Spécifions que si l'unité linguistique atomique correspond habituellement à un mot, celui-ci peut être autrement représenté par son entrée dans un dictionnaire (un lemme), sa racine linguistique (un stemm), sa fonction (une catégorie grammaticale ou son statut d'entité nommée) ou tout autre schème d'annotation. De plus, il est possible de définir des critères restreignant l'adhésion d'une unité linguistique atomique à l'ensemble des unités qui serviront à la représentation vectorielle des textes. Généralement, ces critères visent à identifier les unités lexicales les plus susceptibles de fournir des informations à propos du problème pour lequel une représentation vectorielle des textes est nécessaire. Par exemple, il est habituellement admis que les mots très rares ainsi que ceux très fréquents sont moins susceptibles d'informer le lecteur à propos du problème investigué. À cet effet, un domaine de recherche en traitement du langage naturel est l'identification de mots vides (peu informatif) comme les déterminants « le », « la », « les », etc. Puisque ce ne sont pas toutes les unités linguistiques atomiques ni même les mots ou les termes qui constituent les unités représentationnelles, nous nommons ces dernières « unité d'information » (UNIF). Subséquemment, nous nommons « domaines d'information » (DOMIF) les textes ou segments de textes faisant l'objet d'une représentation vectorielle. Bref, on peut regrouper les vecteurs dénombrant les UNIFs dans les DOMIFs dans une matrice DOMIFs x UNIFs. Il existe plusieurs variantes à cette approche, dont les représentations :

- i. Booléenne où on note la présence ou l'absence de chaque UNIF dans chaque DOMIF,
- ii. Normalisée où les valeurs des UNIFs sont divisées par la norme de chaque DOMIF,
- iii. Pondérée comme le TfIdf (de l'anglais : *Term frequency Inverse document frequency*) où, pour chaque UNIF de chaque DOMIF, Tf (le nombre d'occurrences de l'UNIF dans le DOMIF) est multiplié par IdF (le logarithme du nombre total de DOMIFs divisé par le nombre de DOMIFs contenant l'UNIF).

Or, BOW ou ses variantes ne tiennent pas compte de l'ordre dans lequel se présentent les mots. Une représentation alternative qui tient compte de cet ordre, partiellement du moins, est le sac de n-grammes

ou BONG (de l'anglais : *bag of n-gram*) où chaque DOMIF est représenté par un vecteur dont chacune des composantes correspond à la fréquence de chaque suite de n termes dans les textes. Ainsi, on comprend que les UNIFs de BONG sont constituées de n -grammes de sorte que BOW n'est qu'un cas particulier de BONG, lequel est le seul qui ne tient pas compte de l'ordre puisque la valeur de n est un (1).

6.2.2 Les représentations vectorielles par plongement lexical

Il existe deux techniques principales pour générer des représentations vectorielles basées sur des plongements lexicaux. La première consiste à réduire les dimensions vectorielles traditionnelles, qui représentent les UNIFs, en un nombre plus restreint de dimensions continues, souvent réalisé par une méthode telle que la décomposition en valeurs singulières. La seconde technique, plus élaborée, emploie des réseaux de neurones artificiels pour créer ces représentations vectorielles, permettant une immersion lexicale plus sophistiquée. Généralement, ce type de représentation est constitué de vecteurs dont les composantes sont issues d'une couche neuronale cachée dans un réseau de neurones artificiel préalablement entraîné. Cet entraînement implique généralement la prédiction d'une ou plusieurs unités lexicales étant donné la présence d'une ou plusieurs autres unités lexicales. Celui-ci peut s'effectuer sur le corpus d'étude (d'une étendue habituellement limitée), sur un autre corpus (habituellement très vaste) ou sur les deux lorsque, par exemple, le modèle de représentation entraîné sur un vaste corpus est ensuite affiné sur un corpus d'étude plus limité. Spécifions que l'entraînement d'un réseau de neurones artificiel dans le but de générer une représentation par plongement peut en principe s'effectuer à l'aide de toutes les unités lexicales ou d'une partie restreinte de celles-ci. Par exemple, un critère de sélection peut être l'appartenance des unités à certaines catégories grammaticales ou encore correspondre à la forme lemmatisée de ces unités. Cependant, lorsque le modèle de représentation par plongement est affiné, il est d'usage d'utiliser le même type d'unités lexicales utilisé lors du premier entraînement. La recherche à propos du plongement lexical donne lieu à une multitude de modèles de représentation linguistique et, à ce jour, il n'y a pas de consensus à propos d'un modèle optimal applicable à toutes les applications possibles du traitement du langage naturel (voir Muennighoff *et al.*, 2023).

6.2.2.1 Les représentations vectorielles *word2vec* et *doc2vec*

L'une de ces méthodes sophistiquées de représentation par plongement est présentée par Mikolov et coll. (2013) où les dimensions vectorielles constituent des plongements de types *word2vec* ou *doc2vec*. La méthode utilisée afin de générer la représentation *word2vec* se trouve en deux variantes, soit *continuous bag of words* (CBOW) et *skip-gram* (SG). CBOW est un réseau de neurones artificiels munis d'une couche

cachée dont l'entraînement vise à prédire un terme à partir d'une fenêtre de termes adjacents, c.-à-d. son contexte. On utilise ensuite ce modèle de prédiction du mot à partir des contextes pour extraire la représentation *word2vec*. Celle-ci est composée de chaque vecteur de valeurs générées par la couche cachée pour chacun des termes du vocabulaire de sorte que l'étendue de ces vecteurs correspond au nombre de neurones de la couche cachée (de 50 à 300 habituellement). SG fonctionne de manière similaire, mais inversée, c.-à-d. qu'à partir d'un terme, l'entraînement vise à prédire le contexte. L'équivalent de CBOW et de SG pour *doc2vec* se nomment respectivement *distributed memory* (DM) et *distributed bag of words* (DBOW). Le procédé pour générer la représentation *doc2vec* ne se distingue que par l'ajout d'un identifiant aux vecteurs d'entrée et de sortie correspondant aux DOMIFs. *Doc2vec* permet ainsi de représenter à la fois les termes et les documents par des vecteurs de caractéristiques (*features*) issues d'une couche neuronale cachée. Ces caractéristiques encapsulent de manière implicite des propriétés syntaxiques ou sémantiques comme la synonymie, l'antonymie, l'analogie, etc. L'exemple classique montrant la performance de ce type de représentation est celui où l'addition du vecteur correspondant au terme *king* à celui de *woman* tout en y soustrayant celui de *man* correspond au vecteur de *queen*. Or, si l'apprentissage séquentiel de cette approche explique des performances spectaculaires, cette séquentialité en constitue aussi la limite puisque l'apprentissage ne s'effectue que de gauche à droite (ou de droite à gauche). Afin de dépasser cette limitation, de nouvelles techniques permettent la lecture et l'apprentissage bidirectionnel.

6.2.2.2 Les représentations vectorielles de type BERT

BERT (de l'anglais : *Bidirectional Encoder Representations from Transformers*) est une architecture neuronale profonde développée par Google en 2018 permettant des améliorations notables sur un large éventail de tâches de traitement du langage naturel, y compris la réponse aux questions, l'analyse des sentiments, et la traduction linguistique. Cette architecture sophistiquée permet de capturer les relations complexes entre les mots et de produire des représentations très précises des mots et des phrases. L'entraînement de BERT vise deux tâches distinctes, soit la prédiction d'un terme étant donné son contexte ainsi que la prédiction de la phrase suivante. Plus précisément, le modèle doit prédire quels sont les termes masqués (15% des termes seront masqués lors de l'entraînement) et déterminer si deux phrases se suivent séquentiellement (50% des phrases se suivront séquentiellement lors de l'entraînement). Pour réaliser cette tâche, le modèle utilise un mécanisme nommé *Transformer* introduit par Vaswani *et al.* (2017) et devenu populaire pour les tâches en traitement du langage naturel en raison de sa capacité à rendre compte de dépendances entre les mots. Ce mécanisme incorpore deux principes, soit la bidirectionnalité

et l'autoattention (*self-attention*). La bidirectionnalité provient du fait qu'on traite toute séquence d'entrée dans les deux sens (de gauche à droite et de droite à gauche) pendant l'entraînement. L'autoattention se base sur l'hypothèse que la génération de la représentation d'un mot particulier peut être améliorée en étant attentif à tous les autres mots d'une séquence linguistique. Plus précisément, voici les principales étapes de fonctionnement du *Transformer* :

- i. Une séquence de mots en entrée passe par une couche neuronale d'intégration, laquelle convertit chaque mot d'une séquence en une représentation vectorielle (similaire à *word2vec*). Notons que cette représentation tient compte du positionnement de chaque mot dans la séquence, qu'elle est apprise pendant l'entraînement et qu'elle peut être ajustée pour une tâche spécifique en aval.
- ii. Ensuite, pour chaque mot de la séquence, la couche d'autoattention calcule trois ensembles de vecteurs (nommés vecteurs de requête, de clés et de valeur) en multipliant les représentations vectorielles précédemment énoncées par trois matrices de poids apprises lors de l'entraînement.
- iii. La couche d'autoattention calcule ensuite la similarité du produit scalaire entre les vecteurs de requête et les vecteurs clés pour chaque mot de la séquence. Il en résulte un score pour chaque mot de la séquence qui représente sa pertinence par rapport aux autres mots de la séquence.
- iv. Les scores sont ensuite normalisés à l'aide de la fonction *softmax* pour obtenir un ensemble de poids d'attention pour chaque mot de la séquence. Ces poids d'attention représentent à quel point chaque mot doit être pris en compte par les autres mots de la séquence.
- v. Enfin, les poids d'attention sont utilisés pour calculer une somme pondérée des vecteurs de valeur pour chaque mot de la séquence. Il en résulte un nouvel ensemble de représentations vectorielles qui prennent en compte les informations contextuelles des autres mots de la séquence.
- vi. Ces nouvelles représentations sont transmises à travers un réseau neuronal d'anticipation qui applique une fonction d'activation non linéaire (telle que *ReLU*²⁶) pour produire la sortie du *Transformer*.

BERT étant composé de plusieurs couches de *Transformers* (6 pour la version légère et 24 pour la version large), ce processus est répété pour donner, à terme, une séquence de représentations contextualisées capturant les relations entre tous les mots de la séquence d'entrée. En bref, BERT-large est un modèle de

²⁶ La fonction d'unité linéaire rectifiée (*rectified linear unit* ou *ReLU*) est définie comme $f(x) = \max(0, x)$, ce qui signifie que cette fonction renvoie la valeur d'entrée si elle est positive et zéro si elle est négative. Cette fonction est très populaire, car elle est efficace en termes de calcul et il a été démontré qu'elle fonctionne bien dans de nombreuses applications telles que la classification d'images et le traitement du langage naturel (Eger *et al.*, 2018).

langage²⁷ constitué de 340 millions de paramètres dont l'entraînement original est effectué sur un corpus massivement composé de textes issus de Wikipédia (~2,5 milliards de mots) et de Google's BooksCorpus (~ 800 millions de mots). De plus, une fois entraîné, BERT peut être affiné sur des tâches spécifiques du traitement du langage naturel en ajoutant une couche neuronale en aval du modèle et en l'entraînant sur un ensemble de données étiquetées (habituellement beaucoup plus petit que le corpus d'entraînement). Ce processus permet à BERT de s'adapter aux exigences spécifiques de la tâche et d'atteindre des performances de pointe sur un large éventail de tâches.

6.2.2.3 Les représentations vectorielles inspirées des modèles génératifs comme GPT ou ChatGPT

GPT (de l'anglais : *Generative Pre-trained Transformer*) et ChatGPT (de l'anglais : *Chat Generative Pre-trained Transformer*) sont des architectures neuronales profondes développées par OpenAI conçues pour générer des réponses en langage naturel à diverses amorces conversationnelles, y compris des questions ou des déclarations. Comme BERT, ces architectures sont basées sur le mécanisme du *Transformer*, mais ces modèles se distinguent par le nombre massif de données d'entraînement (390 Go de texte jusqu'en 2021 pour ChatGPT), et un plus grand nombre de paramètres²⁸. Étant des modèles génératifs, ces architectures impliquent des *Transformers* unidirectionnels dont l'entraînement principal vise à prédire le prochain mot d'une séquence donnée. Une fois préentraînés, ces architectures peuvent être affinées sur des tâches spécifiques, telles que la réponse aux questions, la complétion de texte ou la génération de conversation. Parallèlement, ces modèles génératifs inspirent la création de nouveaux modèles de représentation vectorielle mettant à profit certaines de leurs caractéristiques (OpenAI, 2022). Dans le cadre de notre recherche, on s'intéressera plus particulièrement aux modèles de représentation vectorielle tirant parti de l'architecture de modèles génératifs, notamment celui relativement récent nommé « text-embedding-ada-002 » utilisant entre autres le BPE (de l'anglais : *Byte-Pair Encoding*) pour

²⁷ Il est courant de dire que les modèles neuronaux massifs comme *BERT* sont des modèles de langage. Or, si *BERT* modélise ou représente effectivement le corpus massif sur lequel il a été entraîné, il n'est pas évident qu'il modélise par la même occasion une langue ou le langage lui-même. Ce n'est que dans la mesure où les échantillons du corpus d'apprentissage sont suffisamment représentatifs d'une langue (ou du langage) qu'une telle modélisation est possible.

²⁸ La version originale nommée *GPT*, publiée en 2018, avait 117 millions de paramètres. Depuis, de nouvelles versions de *GPT* ont été publiées avec un nombre croissant de paramètres. Par exemple, *GPT-2*, qui a été publié en 2019, avait 1,5 milliard de paramètres, tandis que *GPT-3*, qui a été publié en 2020, avait un nombre de 175 milliards de paramètres. Open-AI n'a pas communiqué le nombre de paramètres de son modèle *ChatGPT*, mais on peut s'attendre à plus de 175 milliards de paramètres.

convertir un texte en unités élémentaires (tokens)²⁹, l'apprentissage contrastif (Neelakantan *et al.*, 2022), et certaines techniques de réduction dimensionnelle (voir Kusupati *et al.*, 2024).

6.3 Les différentes approches de détection des structures argumentatives

Trois approches permettent la détection des CARA. La première est déductive, la seconde est inductive et la dernière est hybride.

6.3.1 Les approches déductives de la détection des structures argumentatives

Les approches déductives de la détection des structures argumentatives nécessitent la connaissance *a priori* de ce qui constitue une CARA ainsi qu'un moyen déductif permettant leur identification. Par exemple, si un expert est généralement capable de détecter des CARA et, si les moyens par lesquels celui-ci réussit cette identification sont traduisibles en des règles, alors un outil informatique composé de ces règles peut automatiquement les identifier. Cette stratégie est possible et utilisée en fouille d'argument lorsque, notamment, certaines CARA sont corrélées à des marqueurs linguistiques (Eckle-Kohler *et al.*, 2015 ; Mizrahi et Dickinson, 2020 ; Stevenson, 1944 ; Tseronis, 2011).

6.3.2 Les approches inductives de la détection des structures argumentatives

Les approches inductives de la détection des structures argumentatives ne nécessitent pas de connaissance *a priori* dans la mesure où elles permettent d'inférer les CARA directement à partir des données, ou encore indirectement, à l'aide d'exemples de CARA et d'exemples de non CARA. La recherche en fouille d'argument montre que plusieurs outils inductifs permettent la détection des CARA comme les machines à support vectoriel (Bar-Haim *et al.*, 2017 ; Duthie *et al.*, 2016 ; Habernal et Gurevych, 2017 ; Lippi et Torroni, 2016a, 2016c ; Menini *et al.*, 2018 ; Mochales Palau et Moens, 2011 ; Naderi et Hirst, 2016 ; Niculae *et al.*, 2017 ; Park et Cardie, 2014 ; Stab et Gurevych, 2017), la régression logistique (Dusmanu *et al.*, 2017 ; Ibeke *et al.*, 2017 ; Levy *et al.*, 2014 ; Nguyen et Litman, 2018 ; Rinott *et al.*, 2015), les algorithmes d'analyse syntaxique (Eger *et al.*, 2017 ; García Villalba et Saint-Dizier, 2012a ; Peldszus et Stede, 2015), les réseaux neuronaux récurrents (Eger *et al.*, 2017 ; Niculae *et al.*, 2017). Les classifieurs naïfs bayésiens (Biran et Rambow, 2011 ; Duthie *et al.*, 2016), les modèles d'entropie maximale (Duthie *et al.*, 2016 ; Mochales Palau et Moens, 2009, 2011 ; Moens *et al.*, 2007), les arbres de décisions et les forêts

²⁹ BPE fusionne itérativement les paires d'octets les plus fréquentes pour créer un vocabulaire fixe d'unités élémentaires (tokens). Par exemple, un mot comme « encoding » sera divisé en « encod » et « ing » parce qu'ils sont statistiquement courants en anglais.

aléatoires (Dusmanu *et al.*, 2017 ; Eckle-Kohler *et al.*, 2015 ; Stab et Gurevych, 2014b), les champs aléatoires conditionnels (Stab et Gurevych, 2017), et l'estimation du maximum de vraisemblance (Levy *et al.*, 2014) sont aussi utilisés pour la détection des composantes argumentatives. Enfin, les algorithmes de reconnaissance des inférences textuelles (*textual entailment*) (Cabrio et Villata, 2013) sont utilisés afin de détecter des relations argumentatives.

6.3.3 Les approches hybrides de la détection des structures argumentatives

Les approches hybrides de détection des structures argumentatives nécessitent un sous-ensemble de cas obtenu déductivement à l'aide de règles ainsi qu'une stratégie inductive permettant l'apprentissage à partir de ce sous-ensemble. Lorsque ce sous-ensemble est exclusivement constitué de cas positifs (des CARA en l'occurrence), alors le problème en est un de classification à l'aide de cas positifs et de cas indéterminés. Nigam et coll. (2000) et Lareau (2022) suggèrent une approche en deux étapes pour ce type de problème. Dans un premier temps, on détermine un sous-ensemble de cas négatifs à partir des cas positifs. Dans un deuxième temps, on identifie l'appartenance des cas indéterminés à l'aide des cas positifs et des cas négatifs. Bref, il s'agit d'entraîner un classifieur sur des cas positifs et négatifs afin de classer les cas indéterminés comme étant positifs ou négatifs. Cette approche est ainsi un cas particulier d'apprentissage semi-supervisé (Claesen *et al.*, 2015 ; Denis *et al.*, 2005 ; Elkan et Noto, 2008 ; Liu *et al.*, 2003 ; Mordelet et Vert, 2011 ; Yu *et al.*, 2004). D'une part, plusieurs algorithmes peuvent être utilisés afin d'identifier les cas négatifs comme, entre autres, la technique de l'espion (Liu, 2011 ; Liu *et al.*, 2002), de Rocchio (Cooley *et al.*, 1999 ; Li et Liu, 2003), de 1-DNF (Yu *et al.*, 2002, 2004), de 1-DNFII (Peng *et al.*, 2008), de NB (Chuang et Chien, 2003), de *Positive examples and Negative examples Labeling Heuristic* (PNLH) (Fung *et al.*, 2006), de *Positive Enlargement* (PE) (Vapnik, 1995), de PPGU (He *et al.*, 2020), de *k-means* (Chaudhari et Shevade, 2012), de *k-nearest neighbors* (KNN) (Zhang et Zuo, 2009), de *Clustering-based method for Collecting Reliable Negative Examples* (C-CRNE) (Liu et Peng, 2014), de *Reliable negatives are selected based on a trainable distance measure Distance Learning for Categorical Attributes* (DILCA) (Ienco *et al.*, 2012), de *Generative Positive-Unlabeled* (GPU) (Basile *et al.*, 2018), de l'augmentation des cas négatifs (Li et Liu, 2005), de l'unique négatif (Li *et al.*, 2007) ou de propagation d'affinité (Luo *et al.*, 2009). D'autre part, différents algorithmes peuvent être utilisés afin de classer les cas indéterminés comme l'espérance-maximisation (Liu *et al.*, 2002), le PEBL (Yu *et al.*, 2002), les SVM (Li et Liu, 2003 ; Liu *et al.*, 2003 ; Manevitz et Yousef, 2001 ; Peng *et al.*, 2008 ; Xiao *et al.*, 2011), la régression logistique pondérée (Lee et Liu, 2003), les graphes (Yu et Li, 2007) ou le TFIPNDF (Liu et Peng, 2014). Enfin, n'importe

quel technique d'apprentissage supervisé peut en principe être utilisée afin de classer les cas indéterminés une fois qu'un nombre suffisant de cas positifs et négatifs sont identifiés (Bekker et Davis, 2020).

6.4 Les représentations visuelles des structures argumentatives

La détection des CARA implique une théorie de l'argumentation dans laquelle sont définies les CARA. Une fois les CARA détectées, il est possible d'inférer une structure graphique représentant l'organisation structurelle des arguments. Au moins deux approches permettent de telles représentations au sein du cadre théorique communicationnel de l'argumentation, soit les arbres d'argumentation et les modèles de séquence. Les arbres d'argumentation sont des structures hiérarchiques qui représentent les relations entre les différentes unités argumentatives d'un texte. Ces unités peuvent inclure des affirmations, des prémices, des conclusions, des contre-arguments, etc. Les arbres d'argumentation sont construits en reliant les différentes unités par des relations telles que l'attaque ou le support. Ces arbres permettent de visualiser et d'analyser la structure globale de l'argumentation, en mettant en évidence les relations entre les différentes parties du texte. Les modèles de séquence sont des approches qui traitent l'argumentation comme une séquence ordonnée d'unités. Dans ce contexte, les modèles d'apprentissage machine peuvent prédire les relations entre les différentes unités argumentatives. Les modèles de séquence peuvent prendre en compte le contexte local des unités, ainsi que des informations globales sur l'argumentation, pour détecter les relations entre les différentes parties du texte.

6.5 L'évaluation des outils de détection

6.5.1 L'évaluation qualitative

Un moyen d'évaluer un système d'analyse d'argument est de l'appliquer à un cas d'étude. L'évaluation est alors comparable à n'importe quelle analyse traditionnellement faite à la main où l'expert expose sa méthodologie, procède à l'analyse, interprète ses résultats, puis la démarche fait l'objet de critique par les pairs qui évaluent ainsi la qualité de l'analyse ainsi que la méthode employée. Comme le soutiennent Barrett (2011) ainsi que Trent et Cho (2014), la démarche scientifique incluant l'interprétation des résultats suit un processus d'argumentation scientifique standard où :

- a. Certaines positions épistémiques apparaissent meilleures que d'autres;
- b. Certaines peuvent différer, se contredire ou être compétitives;
- c. Certaines sont au mieux plus raisonnables, convaincantes ou informatives; mais

- d. Toutes sont assujetties au jugement, lequel peut impliquer des critères consensuels tels que la cohérence, la correspondance (aux données) ou l'exhaustivité et
- e. Le processus global est ultimement une entreprise collective intersubjective.

6.5.2 L'évaluation quantitative

Outre la décomposition du texte en un certain nombre d'entités (phrases ou segments textuels) d'intérêt, la tâche de détection des CARA en est une de catégorisation où la performance est mesurable en termes de précision et de rappel. La précision (*precision*) correspond aux items correctement attribués à la catégorie cible par rapport à l'ensemble des items attribués à la catégorie cible, c.-à-d. le nombre de vrais positifs divisé par le nombre de vrais positifs et de faux positifs. Le rappel (*recall*) correspond aux items correctement attribués à la catégorie cible par rapport à l'ensemble des items appartenant effectivement à la catégorie cible, c.-à-d. le nombre de vrais positifs divisé par le nombre de vrais positifs et de faux négatifs. Une mesure combinatoire est la moyenne harmonique (F-mesure) où :

$$F\text{-mesure} = 2 * (\text{précision} * \text{rappel}) / (\text{précision} + \text{rappel})$$

Une autre mesure tenant compte des vrais négatifs est l'exactitude (*accuracy*) :

$$\text{Exactitude} = (\text{vrais positifs} + \text{vrais négatifs}) / (\text{vrais positifs} + \text{vrais négatifs} + \text{faux positifs} + \text{faux négatifs})$$

Bien que l'exactitude soit une mesure utile, elle peut être trompeuse dans le cas de classes déséquilibrées où une classe prédomine. Pour pallier ce problème, on peut utiliser la macro-moyenne, qui calcule la performance pour chaque classe et en fait la moyenne, accordant un poids égal à chaque classe indépendamment de leur taille. D'autre part, la moyenne pondérée, qui tient compte du nombre d'instances dans chaque classe, mesure également la performance par classe, mais pondère ces mesures en fonction de la taille de chaque classe avant de faire la moyenne. Ces méthodes sont particulièrement pertinentes dans les ensembles de données déséquilibrés pour assurer que la performance du modèle sur les classes moins représentées ne soit pas éclipsée par celle sur les classes majoritaires. De plus, les macro-moyennes et les moyennes pondérées sont d'autant plus utiles lorsque les CARA sont répartis en plusieurs classes. Dans ce contexte, il est possible de mesurer la précision, le rappel ou la F-mesure pour chaque classe individuellement par rapport aux autres, et ensuite agréger ces performances en macro-moyennes et moyennes pondérées pour une évaluation plus globale et équilibrée de la performance du modèle.

6.5.3 Les *corpora* annotés

Selon Cabrio et Villata (2018), les corpus annotés sont développés conformément aux considérations des domaines spécifiques desquels émerge le besoin d'une fouille d'argument. Subséquemment, on peut classer les corpus annotés par rapport à leur domaine d'origine dont, notamment :

- i. En éducation où on s'intéresse entre autres aux essais (persuasifs) des étudiants (Eger *et al.*, 2017 ; Nguyen et Litman, 2018 ; Peldszus et Stede, 2015 ; Stab et Gurevych, 2017) ainsi que les articles scientifiques (Teufel *et al.*, 2009);
- ii. En marketing où on s'intéresse à la revue de services ou de produits et où on utilise entre autres des techniques de détection de sentiment afin d'identifier des relations de support ou d'opposition (García Villalba et Saint-Dizier, 2012a);
- iii. En Droit et en politique où on s'intéresse aux documents légaux (Grabmair *et al.*, 2015 ; Mochales Palau et Moens, 2011 ; Teruel *et al.*, 2018); iv) et aux débats parlementaires ainsi que les débats citoyens à propos de thèmes politiques (Duthie *et al.*, 2016 ; Lippi et Torroni, 2016a ; Menini *et al.*, 2018 ; Naderi et Hirst, 2016 ; Walker *et al.*, 2012);
- iv. En journalisme où on s'intéresse aux articles de journaux (Kiesel *et al.*, 2015) et
- v. En informatique où on s'intéresse aux données volumineuses, comme chez IBM où on étudie la fouille d'argument à l'aide d'un corpus extrait de Wikipédia (Bar-Haim *et al.*, 2017 ; Levy *et al.*, 2014 ; Rinott *et al.*, 2015) ou encore d'autres corpus extraits de plateformes de débats, de blogues ou de tweets (Al-Khatib *et al.*, 2016 ; Cabrio et Villata, 2013 ; Dusmanu *et al.*, 2017 ; Habernal et Gurevych, 2017 ; Lippi et Torroni, 2016c ; Niculae *et al.*, 2017 ; Stab et Gurevych, 2017).

De plus, certains vont générer des corpus de textes artificiels (Peldszus et Stede, 2013b) et d'autres vont fusionner les genres pour générer des corpus hétérogènes, pour une plus grande généralisation de leur modèle, comme le corpus Araucaria (<http://corpora.aifdb.org/araucaria>) qui est annoté selon le modèle rhétorique de Walton et qui est composé d'articles de journaux, de débats parlementaires, de jugements et de discussion (Hua et Wang, 2017 ; Stab *et al.*, 2018).

6.6 Conclusion

En somme, la détection des structures argumentatives est une tâche cruciale dans l'analyse argumentaire. Dans ce chapitre, nous avons examiné différentes approches et techniques utilisées pour détecter ces structures, en mettant l'accent sur les unités élémentaires et leurs relations. Nous avons d'abord exploré

le matériel sur lequel peut s'effectuer la détection des structures argumentatives, en examinant des représentations vectorielles telles que BOW, *doc2vec*, BERT et celles inspirées des modèles génératifs. Ces méthodes fournissent des représentations puissantes pour capturer les informations sémantiques et contextuelles des textes, facilitant ainsi la détection des structures argumentatives. Ensuite, nous avons abordé différentes approches de détection des structures argumentatives. Certaines approches se concentrent sur la détection des unités argumentatives telles que les prémices, les contre-arguments et les conclusions, tandis que d'autres se concentrent sur l'identification des relations entre ces unités. Des méthodes déductives, inductives et hybrides ont été présentées comme des outils efficaces pour cette tâche. Enfin, nous avons discuté des approches d'évaluation qualitative et quantitative de la détection des structures argumentatives. L'évaluation qualitative implique une analyse manuelle des résultats pour évaluer la précision et la cohérence de la détection, tandis que l'évaluation quantitative repose sur des mesures telles que la précision, le rappel, la F-mesure et l'exactitude. Nous avons également souligné l'importance de disposer de jeux de données annotées et de corpus spécifiques pour l'évaluation précise de la détection des structures argumentatives. Subséquemment, nous proposons d'entraîner et d'évaluer différents outils de détection des CARA de manière à identifier leur performance. Par exemple, Sklearn (Pedregosa *et al.*, 2011) est un package implémentant dans le langage informatique Python plusieurs outils permettant la catégorisation comme les méthodes SVM, les méthodes bayésiennes, les méthodes arborescentes, les méthodes neuronales, etc.

CHAPITRE 7

Les hypothèses de recherche

Les deux chapitres précédents présentaient une modélisation computationnelle de la tâche d'analyse d'argument ainsi qu'un approfondissement de l'étape de détection des arguments. Cette modélisation se voulait suffisamment abstraite pour être compatible avec les différentes théories de l'argumentation présentées préalablement. Dans ce chapitre, nous présentons un cadre théorique novateur synthétisant certaines théories du discours et de l'argumentation. Ce cadre théorique permet de mieux comprendre comment les arguments s'organisent dans les textes de haut niveau théorique tout en mettant en lumière certaines composantes argumentatives cruciales pour l'analyse d'argument. Subséquemment, certains objectifs de recherche et certaines hypothèses sont présentés.

7.1 Introduction

La présente thèse doctorale explore une théorie structurelle novatrice de l'argumentation dans le discours de haut niveau théorique. Cette théorie puise ses fondements dans trois cadres conceptuels : la *Rhetorical Structure Theory* (RST), la théorie IMRAD et la théorie de l'intertextualité. L'objectif est de comprendre la manière dont les arguments sont organisés et présentés dans les textes de haut niveau théorique, notamment scientifiques ou philosophiques, permettant ainsi d'approfondir notre compréhension des raisonnements avancés dans ces domaines. La théorie structurelle de l'argumentation postulée propose une approche en trois niveaux distincts : les microstructures argumentaires, les macrostructures argumentaires et les structures argumentaires étendues. Chacun de ces niveaux offre un éclairage précis sur la façon dont les arguments sont agencés et reliés entre eux, contribuant ainsi à dégager une vue d'ensemble des discours de haut niveau théorique. L'objectif de ce cadre théorique est le développement de la lecture et de l'analyse d'un ensemble de textes de haut niveau théorique en vue d'une analyse argumentaire. Plus précisément, on vise l'identification des thèses principales, leurs relations à des arguments et contre-arguments, de manière autant intratextuelle qu'extratextuelle, puisque ces composantes argumentatives se relient non seulement à l'intérieur des textes, mais entre les textes au moyen de composantes textuelles permettant de relier un texte citant à un texte cité, c.-à-d. des contextes de citation. Pour ce faire, nous avançons plusieurs hypothèses concernant l'identification des contextes de citation et des thèses principales dans ces textes spécialisés. Les méthodes de détection envisagées incluent l'exploitation des macrostructures de type IMRAD, la localisation des segments textuels exprimant les thèses principales en fonction de l'emplacement, l'usage de connecteurs de discours ou de certaines

propriétés structurelles comme le titre, le résumé, l'introduction ou la discussion. La combinaison de ces éléments, issus des théories RST, IMRAD et de l'intertextualité, offre une approche rigoureuse pour l'analyse des discours de haut niveau théorique. Cette recherche vise à fournir aux chercheurs, aux analystes et aux lecteurs un outil puissant pour mieux saisir la structure argumentative de ces textes spécialisés.

7.2 Le contexte de citation

En général, on peut comprendre la notion de « contexte de citation » comme l'entourage, l'environnement ou le cadre textuel dans lequel une citation est placée. Cela inclut le texte qui entoure directement la citation. Les délimitations de ce cadre textuel sont sujettes à discussion, à la fois d'un point de vue théorique et pratique, car elles peuvent varier selon les objectifs spécifiques de la recherche. Par ailleurs, une citation est une référence à une source d'information ou à un travail intellectuel, généralement utilisée dans des contextes académiques ou de recherche. Son principal objectif est de reconnaître le travail d'autrui, de fournir des preuves d'affirmations factuelles et de permettre aux lecteurs de retracer l'origine des informations ou des idées. Les citations varient en style et en format, en fonction de la discipline académique et du guide de style suivi (par exemple, APA, MLA, Chicago, Harvard, etc.). Dans les textes de haut niveau théorique, les citations peuvent se trouver dans le corps du texte, dans une note en bas de page ou dans une liste de référence ou bibliographie. Les citations intégrées au texte offrent des références succinctes, mentionnant typiquement le nom de l'auteur et l'année de publication. Les notes de bas de page et de fin, reliées au texte principal par des symboles (comme des chiffres), communiquent des informations supplémentaires sur les sources citées en bas de page (note de bas de page) ou à la fin du document (note de fin). La liste de références ou la bibliographie regroupe toutes les sources mentionnées dans un document, et ces sources peuvent également être signalées dans le texte par des symboles. Dans le texte lui-même, la citation peut signaler la reprise mot pour mot des propos d'une source, la reformulation des idées d'une source avec des mots différents (paraphrase), ou la référence au travail d'autrui pour soutenir un argument ou fournir une information contextuelle. En résumé, les marqueurs textuels qui identifient une citation, et par extension son contexte, sont soit des références

directes dans le texte, soit des renvois à des notes de bas de page, des notes de fin, ou à un élément d'une liste de référence ou bibliographie.³⁰

7.3 La structure argumentaire des textes de haut niveau théorique

Comme vu au premier chapitre, le texte de haut niveau théorique est un genre discursif appartenant généralement aux domaines scientifiques ou philosophiques. Sa forme privilégiée, relativement courte, est celle de l'article publié dans une revue à comité de lecture. Nous nous concentrerons donc sur cette forme particulière qu'est le TEXTE COURT de Haut Niveau Théorique publié dans une REVUE à Comité de lecture (TECHNETIUM). Le principal objectif du TECHNETIUM est de résoudre une question de recherche. Subséquemment, ce type de texte tend à supporter une certaine résolution qu'on nomme Thèse Principale (TP). Cette résolution peut s'effectuer à l'aide de quatre sous-objectifs (eux-mêmes décomposables en sous-objectifs) et, ce faisant, le TECHNETIUM peut généralement se décomposer en au moins quatre structures, parfois titrées « introduction », « méthodes », « résultats » et « discussion », visant à résoudre ces sous-objectifs.

7.3.1 Les microstructures argumentaires

Si on accepte le cadre théorique révisé de la RST, tel que vu au chapitre 4, alors certaines composantes textuelles élémentaires (nommées « UEs » dans la RST) se relient à d'autres UEs en fonction de leur rôle rhétorique pour former des microstructures argumentaires (nommées « segments » dans la RST), lesquelles peuvent se relier à d'autres UEs ou segments pour former d'autres microstructures plus larges. De plus, si l'objectif *a priori* du TECHNETIUM est de résoudre une question de recherche et que son objectif *a posteriori* est de soutenir une TP (qui est la résolution de cette question de recherche), alors, telle que le soutiennent Mann et Thompson (1988), la question de recherche est une unité satellitaire entretenant un lien fonctionnel de type solution (*solutionhood*) envers la solution privilégiée par les auteurs, c.-à-d. la TP. Par conséquent, on peut considérer que la TP constitue l'unité élémentaire la plus centrale du texte au sens où celle-ci est un nucléus sans relation satellitaire par rapport à d'autres UEs ou segments (*most-nuclear* selon Mann et Thompson).

³⁰ En plus des citations directes et des références formelles, il existe d'autres indices de citation dans les textes, qui peuvent être plus subtils ou implicites. Ces indices incluent divers marqueurs discursifs et techniques littéraires qui signalent une allusion ou une référence à d'autres œuvres, idées, ou discours. Dans le cadre de cette recherche, nous nous concentrerons sur les citations directes et les références formelles.

7.3.2 Les macrostructures argumentaires

Si on accepte le cadre théorique des structures argumentaires IMRAD, alors on accepte que l'objectif principal d'un texte structuré IMRAD est décomposable en au moins quatre sous-objectifs de haut niveau. Dans ces textes possédant une structure IMRAD, les microstructures précédemment énoncées se relient les unes aux autres pour former des structures plus larges et éventuellement constituer des macrostructures de type IMRAD répondant à ces sous-objectifs. Ces macrostructures de type IMRAD entretiennent des fonctions rhétoriques entre elles et, ultimement, envers la TP. On peut concevoir les relations fonctionnelles entre les macrostructures de type IMRAD à l'aide des relations prédéfinies par la RST entre un satellite et un nucléus (ou entre des nucléus pour les relations multinucléées). À cet effet, Mann et Thompson (1988) soutiennent que la catégorisation nucléus-satellite est justifiée lorsque la partie nucléaire du texte fait sens sans la partie satellitaire, mais pas inversement, c.-à-d. que la partie satellitaire du texte a peu de sens sans la partie nucléaire:

One way to recognize the functional distinctiveness of nuclei and satellites is to examine the effects of perturbing texts [...] We predict that if a particular nucleus is removed, then the significance of the material in its satellite(s) will not be apparent. Very clear examples of this arise when the 'most-nuclear' unit of a text (a single unit identified by tracing down through the text structure to the nucleus at each level) is removed. (Mann et Thompson, 1988, p. 266-7)

Au niveau des macrostructures IMRAD, une lecture exclusive de la section « discussion » peut sembler insatisfaisante, mais elle permet de comprendre où l'auteur veut en venir, et elle fait sens. Sans cette dernière, l'introduction, la méthodologie et les résultats semblent autrement ne mener nulle part.

7.3.2.1 La macrostructure « discussion »

La macrostructure nommée « discussion » semble constituer un nucléus par rapport aux autres macrostructures qui apparaissent satellitaires. Cette macrostructure constitue un lieu de discussion de laquelle résulte l'adhésion des auteurs (et possiblement des lecteurs) à la TP. Dans la discussion, certains UEs ou segments peuvent constituer des interprétations, notamment des résultats, et une relation de type Interprétation (*Interpretation*) peut les relier. Parallèlement, certains résultats ou certaines interprétations peuvent servir de justification, notamment envers la TP, et une relation de type justifier (*justify*) peut les relier. Bref, les UEs et les segments de la section « discussion » se relient entre eux pour former une macrostructure (un large segment selon la RST) dont l'élément le plus nucléaire est la TP.

7.3.2.2 La macrostructure « introduction »

La macrostructure nommée « introduction » constitue un lieu d'introduction, notamment de la question de recherche pour laquelle des hypothèses peuvent être proposées, des antithèses relevées et un arrière-plan conceptuel associé. Ce faisant, certains UEs ou segments de l'introduction peuvent entretenir des relations de types condition (*condition*), antithèse (*antithesis*) ou arrière-plan (*background*). Le lieu de l'introduction de la question de recherche constitue une unité satellitaire entretenant une relation fonctionnelle de type solution (*solutionhood*) envers la discussion et, ultimement, envers la TP. Bref, les UEs et segments de la section « introduction » forment une macrostructure pour laquelle la question de recherche est l'élément nucléaire.

7.3.2.3 La macrostructure « méthodes »

La macrostructure nommée « méthodes » est le lieu de présentation de la méthodologie par laquelle on entend répondre à la question de recherche en produisant certains résultats ou évidences en faveur d'une solution. Elle constitue un élément satellitaire entretenant une relation de type moyen (*means*) envers les résultats ou évidences recherchés. Les UEs et segments de la section « méthodes » forment une macrostructure dont l'élément nucléaire est la méthodologie de recherche en son sens abstrait. Si le nucléus abstrait de la macrostructure méthodologique est habituellement élaboré plus en détail, alors on peut s'attendre à la présence de relations de type élaboration (*elaboration*) ou de microstructures multinucléées pouvant correspondre à des séquences ou des étapes méthodologiques et, le cas échéant, à ce que plusieurs UEs ou segments entretiennent des relations de types séquence (*sequence*), liste (*list*), conjonction (*conjunction*) ou articulation (*joint*). Par exemple, le matériel de recherche utilisé peut être explicité dans la méthodologie et, le cas échéant, ce matériel constitue une UE ou un segment de la méthodologie qui peut entretenir une relation multinucléée, possiblement de type articulation (*joint*), envers les autres UEs ou segments principaux de la méthodologie.

7.3.2.4 La macrostructure « résultats »

La macrostructure nommée « résultats » est le lieu de présentation des résultats et des évidences obtenus au moyen de la méthodologie. Elle constitue une macrostructure satellitaire entretenant des relations, possiblement de type évidence (*evidence*), interprétation (*interpretation*) ou justification (*justify*) envers certains UEs ou segments de la discussion, notamment la TP. Les UEs ou les segments de la section « résultats » forment une macrostructure pour laquelle le nucléus constitue les résultats de recherche en un sens abstrait. Subséquemment, certains UEs ou segments de cette section peuvent être multinucléés

et entretenir des relations de types séquence (*sequence*), liste (*list*), conjonction (*conjonction*) ou articulation (*joint*). De plus, on peut s'attendre à des relations de type élaboration (*elaboration*) lorsque certains UEs ou segments sont détaillés.

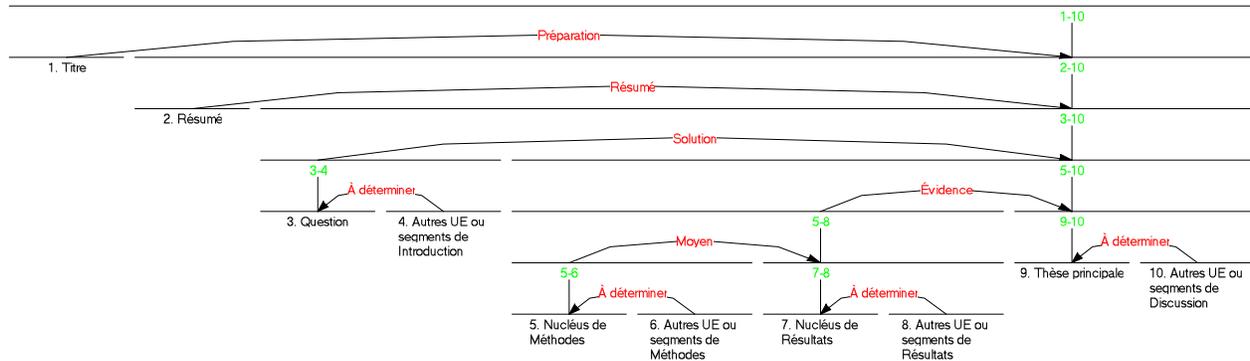
7.3.2.5 La structure « corps du texte » et autres structures adjacentes

Les macrostructures IMRAD forment ensemble une structure plus large que nous nommons « corps du texte ». Le corps du texte répond à l'objectif principal du texte. Il est le lieu d'expression du contenu rhétorique principal et il constitue un nucléus par rapport aux structures textuelles adjacentes comme le titre, le résumé, les mots-clés, les remerciements, les annexes, la bibliographie, etc.³¹ Ce faisant, le corps du texte entretient des relations fonctionnelles envers les autres structures textuelles reconnues par la théorie IMRAD. Bien qu'il existe une infinité de configurations possibles dans le cadre de la RST, la figure 7.1 suivante montre de manière très générale comment peuvent s'organiser les macrostructures « introduction », « méthodes », résultats » et « discussion » pour former le corps du texte, et comment celui-ci entre en relation avec le titre et le résumé. Dans cette figure, les relations entre les éléments constitutifs de chaque macrostructure IMRAD sont dites « à déterminer » dans la mesure où l'exemple constitue une abstraction textuelle sans détail. Spécifions que les relations « à déterminer » sont présentées dans la figure comme des relations entre un nucléus et un(des) satellite(s), mais celles-ci pourraient autrement être multinucléées. De telles relations se trouvent entre la question de recherche et les autres EU ou segments de la macrostructure « introduction » et forment ensemble le segment « 3-4 ». De telles relations se trouvent aussi entre les EU ou segments de la macrostructure « méthodes » pour former le segment « 5-6 » ainsi qu'entre les UEs ou segments de la macrostructure « résultats » pour former le segment « 7-8 ». Enfin, des relations « à déterminer » se trouvent entre la TP et les UEs ou segments de la macrostructure « discussion » pour former le segment « 9-10 ». Le segment « 3-4 » est en relation de type solution (*solutionhood*) avec le segment « 9-10 », le segment « 5-6 » est en relation de type moyen (*mean*) avec le segment « 7-8 » et ce dernier est en relation de type évidence (*evidence*) avec le segment « 9-10 ». Enfin, ces relations forment ensemble le segment « 3-10 » – la structure du corps du texte – à laquelle se relie le résumé pour former le segment « 2-10 », lequel segment se relie au titre pour former le segment « 1-10 ». La RST reconnaît les relations fonctionnelles de type préparation (*preparation*)

³¹ Il peut être matière à débat si les mots-clés font l'objet d'une relation rhétorique décrite dans le cadre de la RST. Leur effet pragmatique est de susciter l'intérêt du lecteur au moyen, notamment, de la suggestion du texte par des moteurs de recherche. Ce faisant, la relation entre le texte et les mots-clés est très similaire à celle de type préparation (*preparation*).

et de type résumé (*summary*) respectivement applicables au titre et au résumé qui constituent ainsi des UEs ou segments satellitaires du corps du texte.

Figure 7.1 Organisation des macrostructures IMRAD



7.3.3 Les structures argumentaires étendues

Les structures argumentaires étendues émergent du phénomène de l'intertextualité. La notion d'intertextualité est originalement élaborée par le groupe de théoriciens Tel Quel (R. Barthes, J.-L. Baudry, J. Derrida, M. Foucault, J.-J. Goux, J.-L. Houdebine, J. Kristeva, M. Pleyne, J. Ricardou, J. Risset, D. Roche, P. Rottenberg, Ph. Sollers, J. Thibaut) et met en jeu la relation entre différents textes, où un texte fait référence à un autre texte ou incorpore des éléments d'autres textes. Les théoriciens de l'intertextualité reconnaissent que les textes ne sont pas isolés, mais qu'ils sont en constante interaction les uns avec les autres. Lorsqu'un auteur crée un texte, il puise souvent dans un réservoir de connaissances et de références culturelles déjà existantes dans d'autres textes :

[...] tout texte se construit comme mosaïque de citations, tout texte est absorption et transformation d'un autre texte. (Kristeva, 1969, p. 85)

La notion d'intertextualité a fait l'objet de multiples débats, mais la définition fournie par Gérard Genette semble maintenant faire l'objet d'un relatif consensus (voir Roux-Faucard, 2006). Genette (1982) développe une approche structurée de l'intertextualité qui est décrite comme l'une des cinq catégories de la transtextualité. La transtextualité est tout ce qui met le texte « en relation, manifeste ou secrète, avec d'autres textes » (Genette, 1982, p. 7). La transtextualité peut être explicite, par le biais de citations ou de références directes à d'autres textes, ou implicite, par le biais de plagiats, d'allusions ou de similitudes thématiques (voir Wagner, 2006). L'intertextualité chez Genette se réfère spécifiquement aux relations explicites de co-présence entre deux ou plusieurs textes, c.-à-d. les citations, les allusions, et les références

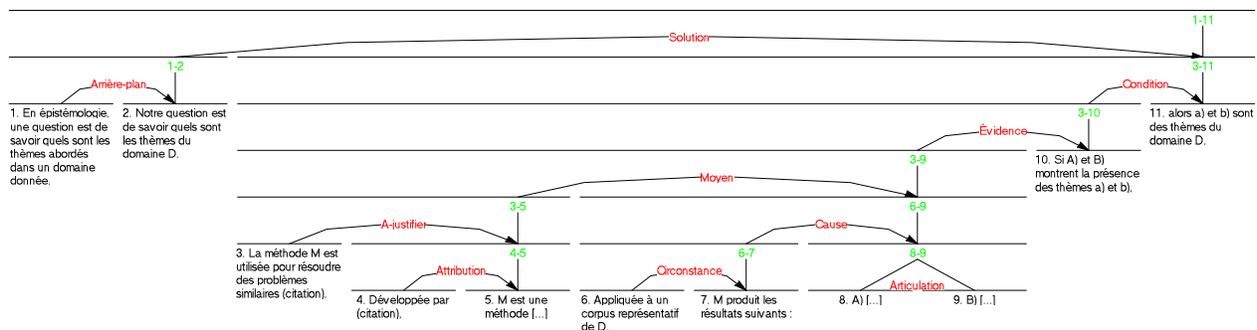
directes, soit « une relation de coprésence entre deux ou plusieurs textes c'est-à-dire, eidétiquement et le plus souvent, [...] la présence effective d'un texte dans un autre » (Genette, 1982, p. 8). Habituellement, on utilise les expressions « texte citant » pour référer au texte qui cite et « texte cité » pour référer au texte préexistant, lequel est cité. Le texte formé de macrostructures IMRAD implique généralement des procédures de citations et de référencement à d'autres textes de sorte qu'il tend à citer explicitement selon un format bien défini. Dans le cadre de la RST (révisée), certains UEs ou segments entretiennent des relations d'attribution envers d'autres UEs ou segment. Lorsque les UEs ou les segments reliés par une relation d'attribution contiennent une citation, alors ceux-ci se nomment « contexte de citation » (*citation contexts*). Autrement dit, les contextes de citation sont des UEs ou des segments où on réfère à un ou d'autres textes. Comme le suggère Potter (2019), une attribution n'est parfois pas autre chose qu'une attribution lorsqu'on ne vise qu'à informer le lecteur sur l'identité de l'auteur ou la provenance d'une idée, d'une UE ou d'un segment. Autrement, la relation attributive n'a pas toujours comme principal objectif l'attribution, car bien souvent la relation attributive joue un rôle prédéfini par le modèle classique de la RST comme la justification (*justify*), l'évaluation (*evaluate*), l'interprétation (*intepretation*), la cause (*cause*), l'élaboration (*elaboration*) ou l'évidence (*evidence*). Une relation d'attribution très utilisée dans les textes structurés IMRAD est celle fonctionnant comme justification – A-justification (*Attribution as Justify*) définie en annexe – laquelle instancie habituellement un appel à l'autorité. Or, la justification ne se trouve pas nécessairement dans l'autorité de l'auteur cité. Elle peut entre autres provenir d'une information, d'un contexte ou d'un argument développé plus en détail dans le texte cité. Subséquemment, il est possible que la compréhension satisfaisante d'un lecteur critique implique la lecture du texte cité. Selon Laurent Jenny (1976), l'intertextualité est un phénomène qui implique des effets importants lors de l'expérience de lecture, notamment de possibles sauts paradigmatiques brisant la linéarité du texte :

Le propre de l'intertextualité est d'introduire à un nouveau mode de lecture qui fait éclater la linéarité du texte. Chaque référence intertextuelle est le lieu d'une alternative : ou bien poursuivre la lecture en ne voyant là qu'un fragment comme un autre, qui fait partie intégrante de la syntagmatique du texte – ou bien retourner vers le texte-origine en opérant une sorte d'anamnèse intellectuelle où la référence intertextuelle apparaît comme un élément paradigmatique « déplacé » et issu d'une syntagmatique oubliée. (Jenny, 1976, p. 281)

Suivant Jenny, la lecture du texte cité est optionnelle parce qu'elle n'est habituellement pas nécessaire à la compréhension du texte citant. Par conséquent, les UEs ou segments cités sont en relation satellitaire par rapport aux UEs ou segments citant. Si ce qui est cité est une UE ou un segment dans le texte cité, alors on peut relier ce qui est cité au contexte de citation dans le texte citant pour former une structure

argumentative étendue. Une telle structure étendue résulte de la relation fonctionnelle rhétorique entre la microstructure intratextuelle qui cite – le contexte de citation – et son extension extratextuelle qui est citée. Puisque la lecture de cette structure argumentative étendue est optionnelle au sens où la lecture de l’extension structurale n’est pas nécessaire à la compréhension du texte citant, alors l’extension structurale est satellitaire par rapport au contexte de citation. La figure 7.2 suivante montre comment sont reliées les UEs d’un pseudotexte structuré IMRAD à l’aide de relations RST dont certaines sont des relations d’attribution. Dans cet exemple, l’UE #3 cite un ou plusieurs textes dans lesquels une méthode M est utilisée pour résoudre un problème similaire au texte citant. Cette UE est en relation de type A-justification (*Attribution as Justify*) avec le segment « 4-5 » qui explicite la méthode M. Autrement dit, l’usage de la méthode M est justifié, car cette dernière est reconnue capable de résoudre un même genre de problème par la communauté scientifique. Cependant, soit le lecteur fait confiance en l’auteur et accepte cette justification, soit celui-ci ne l’accepte pas sans vérifier préalablement si, effectivement, la méthode M est employée pour résoudre des problèmes similaires. Dans la deuxième éventualité, il serait possible d’étendre la microstructure argumentative du texte citant à l’aide de la microstructure argumentative du texte cité, lesquelles formeraient ainsi une structure argumentative étendue. Cette dernière pourrait montrer que la méthode M est effectivement utilisée afin de résoudre le même genre de question et, le cas échéant, l’extension de la structure argumentative étendue serait en relation de type Évidence (*Evidence*) envers l’UE #3.

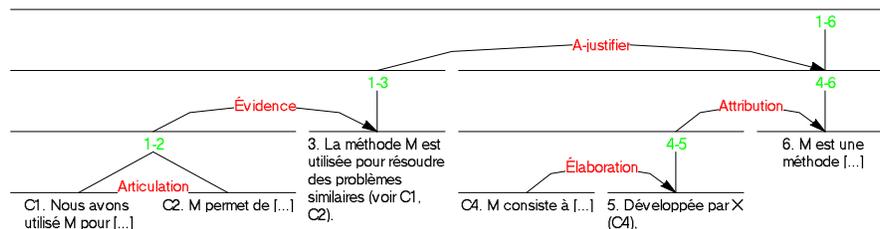
Figure 7.2 Exemple de relation d’un pseudotexte structuré IMRAD



Dans la même figure 7.2, on voit que l’UE #4 est en relation d’attribution avec l’UE #5 et, dans la mesure où cette dernière exprime suffisamment en quoi consiste la méthode M, il est possible de considérer l’attribution comme seulement une attribution. Autrement, si une bonne partie de ce en quoi consiste la méthode M n’est pas explicite dans le texte citant, alors le rôle de l’UE #4 n’est pas seulement l’attribution, mais aussi d’indiquer où se trouvent les détails de la méthode M. Le cas échéant, on peut soutenir que la

relation d'attribution entre #4 et #5 permet de relier la microstructure citée qui élabore la méthode M au contexte de citation à l'aide d'une relation conséquente, une relation de type *Élaboration* (*Elaboration*) par exemple. La figure 7.3 suivante exemplifie cela en reprenant l'exemple fictif de la figure précédente, mais en introduisant les structures étendues précédemment décrites. Les extraits « C1 », « C2 » et « C4 » sont des UEs provenant de textes cités. Les UEs #3, #5 et #6 proviennent du texte citant. C1 et C2 sont reliés par une relation de type *Articulation* (*Articulation*) et forment la microstructure « 1-2 », laquelle est reliée à l'UE #3 par une relation de type *Évidence* (*Evidence*). C4 est relié à l'UE #5 par une relation de type *Élaboration* (*Elaboration*), laquelle est justifiée par le double rôle de la relation d'Attribution (*Attribution*) entre #5 et #6 (en supposant que la méthode n'est pas détaillée dans le texte citant et que C4 est effectivement une élaboration de la méthode M qui est développée dans le texte cité).

Figure 7.3 Exemple de structures étendues



Somme toute, les textes structurés IMRAD comportent au moins trois niveaux structurels d'argumentation. Un premier niveau structurel est composé d'UE et de segments intratextuels formant des microstructures à l'aide de relations définies par la RST (révisée). Un second niveau est composé de structures étendues qui émergent du phénomène de l'intertextualité: d'abord d'une manière implicite en cela que la compréhension d'un texte peut nécessiter celle d'autres textes et, ensuite, de manière explicite en cela que le genre discursif formant une structure IMRAD implique des procédures de citations et de référencement à d'autres textes. Il s'ensuit que la structure étendue peut être hétérogène, car elle n'est pas nécessairement composée de structures provenant de textes de haut niveau théorique – un texte pouvant faire appel à n'importe quel texte – et ce qui est cité peut être n'importe quelle structure du texte (microstructure, macrostructure, structure de haut niveau, tout le texte). Enfin, un dernier niveau structurel est composé de macrostructures dont les composantes essentielles répondent à la formule IMRAD. La première macrostructure – l'introduction – vise à soumettre une question de recherche (ou une hypothétique réponse à cette question sous la forme d'hypothèses de recherche) et à la positionner face à d'autres textes; la seconde macrostructure – les méthodes – vise à présenter les matériaux et les méthodes permettant de répondre à la question de recherche; la troisième macrostructure – les résultats

– vise à présenter l’aboutissement de l’application des méthodes et la dernière macrostructure – la discussion – vise une résolution de la question de recherche à la lumière de l’investigation effectuée, ainsi qu’une évaluation de cette entreprise en relevant ses limites et les implications des solutions trouvées. Subséquemment, on peut concevoir toutes les composantes textuelles comme jouant un rôle argumentatif. Enfin, si on accepte l’hypothèse selon laquelle la structure argumentative du texte structuré IMRAD s’organise en trois niveaux distincts, alors une tâche importante de l’analyse d’argument est d’identifier le rôle des composantes argumentatives ainsi que leurs relations par rapport aux microstructures argumentaires, macrostructures argumentaires et aux structures argumentaires étendues.

7.4 Les hypothèses

7.4.1 L’identification de l’appartenance d’un contexte de citation à la macrostructure IMRAD

Dans le contexte du texte dont la structure comporte trois niveaux argumentatifs, certaines composantes argumentaires jouent un rôle important d’articulation entre la microstructure, la macrostructure et la structure étendue. Parmi celles-ci se trouvent les contextes de citation, lesquels sont des UEs ou des segments où on réfère à un ou d’autres textes. Les contextes de citation peuvent se relier à d’autres UEs ou segments à l’intérieur du texte pour former des microstructures argumentatives, ou à l’extérieur du texte pour former des structures argumentatives étendues. De plus, les microstructures et les structures étendues se positionnent habituellement à l’intérieur d’une macrostructure IMRAD. Par exemple, si un auteur qui explicite sa TP dans la section discussion spécifie que celle-ci est énoncée dans un autre texte, alors une microstructure argumentative est composée entre la TP et son attribution à l’intérieur du texte. Parallèlement, l’UE ou le segment textuel qui est cité peut lui-même se relier à la microstructure d’attribution du texte citant pour former une structure étendue. Enfin, cette dernière se positionne relationnellement à l’intérieur de la macrostructure de discussion. Notre première question de recherche est de savoir s’il est possible d’identifier les macrostructures dans lesquelles occurrent les contextes de citation dans les TECHNETIUM. Notre intuition est que cette identification est possible à l’aide de différentes techniques d’apprentissage machine. Plus spécifiquement, nous cherchons à savoir quels types de représentations textuelles et quels types de classifieur sont les plus performants pour cette tâche. Notre première hypothèse (H1) est que l’identification de l’appartenance des contextes de citation aux structures et macrostructures IMRAD est possible via certaines représentations vectorielles du texte et certains classifieurs. Comme vu au chapitre 6, différentes techniques de représentation de la donnée textuelle sont possibles. Notre première hypothèse subordonnée (à H1) est que les techniques de représentation impliquant un grand nombre de paramètres, inspirées de modèles génératifs comme

ChatGPT (qui est composé de ~175 milliards de paramètres), sont plus performantes que les autres (H1.1). Ensuite, différentes méthodes de classification sont possibles, notamment des méthodes fondées par la classification catégorielle, la classification instanciée, la classification linéaire, la classification arborescente et la classification neuronale. Ainsi, notre seconde hypothèse subordonnée (à H1) est que la classification linéaire est performante pour cette tâche (H1.2).

7.4.2 L'identification de la thèse principale

Une autre composante argumentative particulièrement importante lors de l'analyse d'argument en contexte de TECHNETIUM est la TP dans la mesure où celle-ci est centrale puisqu'elle n'est pas le satellite d'une autre composante argumentative. Notre seconde question de recherche est de savoir s'il est possible d'identifier la TP dans les TECHNETIUM. Notre seconde hypothèse (H2) est que la détection des énoncés de TP est possible via certains indices matériels. Plus précisément, notre intuition est que cette identification est possible à l'aide de l'emplacement, certains connecteurs de discours ainsi que certaines propriétés syntaxiques ou sémantiques propres aux structures et macrostructures IMRAD.

7.4.2.1 L'identification de la thèse principale à l'aide de l'emplacement

En contexte de TECHNETIUM, la TP semble plus susceptible d'être présente à des emplacements particuliers du texte, notamment la discussion et, dans une moindre mesure, l'introduction en raison de la relation privilégiée de type solution (*solutionhood*) entre la question de recherche habituellement exprimée en introduction et la solution généralement exposée dans la discussion. En introduction, la question de recherche peut donner lieu à une tentative de solution sous la forme d'une hypothèse en fonction du schème épistémologique adopté par l'auteur. Les macrostructures « méthodes » et « résultats » semblent des emplacements où la TP est moins susceptible d'être exprimée, sous forme de question ou de réponse, car la première vise à expliciter par quel moyen on compte obtenir certains résultats et la seconde vise à exposer ces résultats qui permettront de soutenir la TP lors de la discussion. Subséquemment, l'emplacement semble constituer un indicateur important pour l'identification des segments de textes énonçant la TP. D'une part, l'emplacement peut être considéré de manière structurale, c.-à-d. que certaines macrostructures propices à l'énonciation de la TP tendent à se trouver à des emplacements spécifiques du texte comme cela semble être le cas pour les macrostructures de type IMRAD (Pontille, 2007). D'autre part, un emplacement plus précis dans le texte est possiblement privilégié pour l'énonciation de la TP, notamment à l'intérieur des macrostructures de type IMRAD. À cet effet, nous postulons une hypothèse subordonnée (à H2) selon laquelle l'identification des TP est possible via

l'emplacement (H2.1). Cette dernière donne lieu à deux hypothèses subordonnées soit i) que la macrostructure de la discussion est un emplacement probable de l'énonciation des TP (H2.1.1) et ii) ces dernières se positionnent à des emplacements spécifiques à l'intérieur des structures ou macrostructures dans lesquelles elles sont situées (H2.1.2).

7.4.2.2 L'identification de la thèse principale à l'aide de connecteurs de discours

Les connecteurs de discours ont été observés depuis longtemps et divers chercheurs ont apporté des contributions significatives à leur compréhension. Parmi ceux-ci se trouve le linguiste Émile Benveniste qui a entre autres exploré la notion de « connecteur logique » et a discuté leur rôle dans la structuration du discours et l'établissement de relations entre différents éléments textuels (Benveniste, 1966). Une autre contribution vient de William Labov qui a mené des recherches sur le discours narratif et a identifié des connecteurs de discours – nommés dispositifs narratifs (*narratives devices*) – qui servent à relier les différents éléments de la narration et à assurer leur cohérence (Labov, 1972 ; Labov et Waletzky, 1967). Également connu sous les noms de particule pragmatique (Ducrot, 1980 ; Schourup, 1985), marqueur discursif (Portolés, 1998 ; Schiffrin, 1987) ou opérateur discursif (Anscombe *et al.*, 2013 ; Redeker, 1991), le connecteur de discours est un élément ou un mot qui sert à signaler une relation entre différentes parties discursives et qui aide à établir la cohérence ainsi que l'organisation du discours. Comme le soulignent Anscombe et Delahaie (2014), les connecteurs de discours se distinguent des noms et des verbes en cela que leur rôle n'est pas de représenter des objets, des propriétés, des états, des procès ou des événements, mais plutôt d'articuler différentes parties du discours :

À côté des entités comme les verbes qui renvoient à des états ou encore à des procès spécifiques, ou les noms, qui évoquent soit des objets « matériels » soit des propriétés intangibles, soit même des procès ou des événements, ont été étudiées d'autres entités et notamment celles qui [ont] une fonction de *relateur*. Le champ d'analyse devait par la suite se diversifier et apparaissaient ensuite des recherches autour des particules pragmatiques [,] des marqueurs discursifs [et] opérateur discursif [...] (Anscombe et Delahaie, 2014, p. 161)

Les connecteurs de discours jouent un rôle important dans l'orientation du flux d'informations et la formation de la structure globale d'un texte. Ils fournissent des indices au lecteur sur la façon dont différentes idées, phrases ou clauses sont liées et comment elles contribuent au sens général. Par exemple, les connecteurs de discours permettent de communiquer des relations telles que la causalité, le contraste, la comparaison, la concession, l'explication, la séquence temporelle, etc. Ils peuvent être explicites ou s'appuyer sur le contexte général et les indices linguistiques pour établir un lien entre les différentes

parties du discours. Ils contribuent à la cohésion et à la cohérence d'un texte, facilitant la circulation fluide et cohérente de l'information pour une communication efficace.³² Notre hypothèse subordonnée (à H2) est que certains connecteurs de discours peuvent être utilisés pour l'identification des TP (H2.2).

7.4.2.3 L'identification de la thèse principale à l'aide de propriétés syntaxiques ou sémantiques.

Nous cherchons à savoir quelles structures ou macrostructures argumentaires sont susceptibles de permettre l'identification des TP. Notre hypothèse subordonnée (à H2) est que certaines propriétés syntaxiques ou sémantiques propres aux structures ou macrostructures IMRAD permettent l'identification des TP (H2.3). À cet effet, quatre (4) structures argumentatives semblent contenir des informations utiles à l'identification des TP, soit le titre, le résumé, l'introduction et la discussion.

7.4.2.3.1 L'identification de la thèse principale à l'aide du titre

Le titre est une unité linguistique indépendante apparaissant au début d'un texte qui résume le texte de manière très concise et qui renseigne sur son contenu thématique (Shilo et Ragonis, 2023). Certains chercheurs postulent une relation étroite entre le thème principal et la TP d'un texte (Levy *et al.*, 2014 ; Trautmann *et al.*, 2020). En analyse du discours, tel que vu au premier chapitre, le thème est une composante en relation à une autre – le rhème – qui forment ensemble une structure linguistique informative. La première exprime de quoi il s'agit et la seconde exprime quelque chose à son propos. La relation thème-rhème permet entre autres de décrire la structure classique sujet-prédicat, mais en rendant compte de certains effets pragmatiques. Par exemple, elle permet de décrire l'effet produit par la juxtaposition du thème et du rhème, respectivement en termes d'informations déjà connues sur le thème (implicitement) et d'informations exprimées à propos du thème (explicitement).³³ Traditionnellement, la structure sujet-prédicat (de laquelle s'inspire la structure thème-rhème) est connue comme étant un jugement (Kneale et Kneale, 1962). Or, selon Brentano (et son élève Marty), le

³² Spécifions que différentes typologies peuvent être utilisées pour nommer les éléments linguistiques ayant pour fonction de relier entre elles certaines parties du discours. Selon Lewis (2000, p. 51), par exemple, le connecteur de discours est un cas particulier de marqueur discursif dont la spécificité est de signaler une relation rhétorique entre au moins deux segments discursifs. Cette relation rhétorique entre segments discursifs est comprise dans un contexte plus large de structuration de l'information du discours ainsi que de l'objectif global du discours (Lewis, 2000, p. 98).

³³ Bien qu'il existe une forte corrélation entre le thème et le sujet, spécifions qu'un thème n'est pas nécessairement un sujet et un sujet n'est pas nécessairement un thème. D'une part, une proposition peut être à propos d'un référent qui n'est pas explicite dans son expression. D'autre part, une expression référentielle à l'intérieur d'un énoncé propositionnel peut ne pas constituer ce à propos de quoi porte l'énoncé (Lambrecht, 1994).

jugement catégoriel de forme sujet-prédicat n'est pas le seul type de jugement possible puisqu'un jugement peut ne pas posséder de sujet (voir Kuroda, 1972). Le cas échéant, Lambrecht (1994) suggère de distinguer la référence au thème de son expression linguistique, car l'absence d'une expression référentielle n'implique pas nécessairement l'absence d'un référent thématique. Comme le souligne Abbott (2004), un jugement en apparence sans sujet décrit généralement un certain état du monde. Par exemple, l'énoncé « il pleut » ne semble pas référer explicitement à un thème et, pourtant, un certain événement en l'occurrence pluvieux est désigné. Similairement, l'énoncé « quelque chose brûle » ne semble pas référer explicitement à un thème, mais une certaine entité, qui en l'occurrence brûle, est désignée. Selon Lambrecht, ce type d'énoncé, qualifié de « thétique », a pour fonction d'introduire un événement ou une entité de sorte que celui-ci joue un rôle thématique :

[...] the sentence expressing the thematic proposition introduces a new element into the discourse [and] the newly introduced element is an ENTITY (a discourse referent) [or] an EVENT, which necessarily involves an entity [...] It is not the absence of any topic relation that characterizes thematic sentences but the absence of a topic relation between the proposition and that argument which functions as the topic in the categorical counterpart. (Lambrecht, 1994, p. 144)

Par conséquent, si le thème d'un texte entretient une relation envers la TP, et que le titre explicite d'une manière ou d'une autre le thème d'un texte, alors il est plausible que le titre constitue une structure argumentative permettant l'identification de la TP dans le texte. De plus, la recherche montre qu'un nombre croissant d'articles scientifiques expriment une question de recherche dans leur titre (Ball, 2009). Or, les questions et les thèmes sont étroitement liés, car les questions sont généralement formulées dans le contexte d'un thème particulier. Lorsqu'une personne pose une question, elle recherche généralement des informations liées à un sujet spécifique. Dans la mesure où la TP répond à une question de recherche, alors un titre exprimant une question de recherche permet aussi l'identification de la TP dans le texte. Néanmoins, cette intuition d'une relation entre le titre et la TP est à vérifier dans la mesure où le titre peut jouer un rôle de médiation entre le texte et le lecteur. En effet, la forme concise du titre ne permet pas d'explicitement toute l'information pertinente du texte de sorte que le titre doit faire appel à des procédés pragmatiques transtextuels mettant en scène la culture du lecteur (Dor, 2003). Ce faisant, on pourrait aussi s'attendre à ce que le titre et la TP ne soient pas liés de manière immédiate, mais au moyen de processus de pertinence et de construction du sens. Autrement dit, si certains contenus importants ne se trouvent pas explicites dans le titre, alors celui-ci pourrait entretenir une faible similarité syntaxico-sémantique avec la TP. Comme le souligne Dor, le titre n'est pas un résumé, mais « un dispositif de communication dont la

fonction est de produire le niveau optimal d'affinité entre le contenu et le contexte d'interprétation du lecteur, afin de rendre le texte pertinent pour le lecteur » (Dor, 2003, p. 720, notre trad.). Bref, notre hypothèse subordonnée (à H2.3) est que le titre est une structure contenant des informations favorisant la détection des TP (H2.3.1).

7.4.2.3.2 L'identification de la thèse principale à l'aide du résumé

Le résumé est une unité linguistique indépendante qui apparaît au début d'un texte et qui exprime une version condensée de ce texte. Finkelstein (2009, p. 217-232) distingue trois types de résumés, soit i) le résumé descriptif qui exprime sous forme de paragraphes la structure du texte (la table des matières par exemple), ii) celui informatif dont la forme plus libre présente brièvement le contenu ou les thèmes importants et iii) celui exécutif qui reprend la structure du texte en plus de présenter les thèmes ou les contenus principaux. Parallèlement, Nundy, Kakar et Bhutta (2022a) reconnaissent cinq types de résumés. En plus des résumés descriptifs, informatifs et exécutifs, les auteurs reconnaissent l'existence de résumés critiques qui sont orientés vers le jugement et la validation ainsi que les résumés rhétoriques (nommés *highlighting*) qui servent à attirer l'attention du lecteur et qui sont très peu utilisés en science. Contrairement aux résumés descriptifs, informatifs, critiques ou rhétoriques, on peut considérer le résumé exécutif ou structuré comme un texte autonome exprimant de manière condensée les structures importantes et les contenus principaux d'un texte plus long et plus détaillé. Généralement, les résumés des TECHNETIUM répondent aux normes édictées par un comité éditorial et peuvent prendre une forme descriptive, informative, critique, rhétorique ou structurée. Le résumé le plus utilisé pour une majorité de revues à comité de lecture est celui structuré. Habituellement, la structure textuelle du résumé est limitée à un certain nombre de mots et les sections correspondent minimalement aux structures IMRAD (Andrade, 2011 ; Nundy *et al.*, 2022a).³⁴ Enfin, dans la mesure où un texte soutient une TP, il semble fort probable que celle-ci soit exprimée dans le résumé, notamment lorsque celui-ci est structuré. Subséquemment, notre hypothèse subordonnée (à H2.3) est que le résumé est une structure contenant des informations pertinentes pour la détection des TP (H2.3.2).

³⁴ Rappelons que les sections introduction, méthodologie, résultats et discussion peuvent être titrées (ou non), se nommer autrement et certaines sections peuvent s'ajouter.

7.4.2.3.3 L'identification de la thèse principale à l'aide de l'introduction

Dans les textes structurés IMRAD, l'introduction est une macrostructure textuelle jouant un rôle argumentaire où est exprimée et contextualisée une question de recherche en relation aux recherches qui la précède ainsi que les intuitions ou indices qui sont exprimés sous forme d'hypothèse de recherche. Selon Armağan (2013), le but sous-jacent l'introduction est de « transmettre des informations de base aux lecteurs sans les obliger à enquêter sur les publications précédentes et de fournir des indices sur les résultats de la présente étude » (Armağan, 2013, p. 8, notre trad.). Idéalement, l'introduction contient les informations suivantes : i) une hypothèse ou une question de recherche, ii) ce qui est connu à son propos, iii) ce qui est inconnu à son propos, et vi) la méthodologie utilisée (Ghasemi *et al.*, 2019 ; Kearney, 2017 ; Nundy *et al.*, 2022b). Subséquemment, il est probable que la TP soutenue par le texte se trouve dans l'introduction sous forme de question ou d'hypothèse de recherche. Or, une question ou une hypothèse de recherche n'est pas une thèse en soi. D'une part, une thèse prend une forme affirmative tandis qu'une question prend une forme interrogative. Par exemple, la question de recherche « comment écrire une introduction dans un article scientifique? », explicite dans le titre de l'article de Nundy et coll. (2022b, notre trad.), est répondu en conclusion par la thèse selon laquelle une introduction dans un article scientifique prend idéalement la forme d'un bref compte rendu de l'objet étudié, lequel passe d'un sujet de recherche général à un sujet de recherche spécifique, et inclut une recherche documentaire explicitant ce que l'on sait déjà et ce que nous espérons ajouter à ces connaissances (Nundy *et al.*, 2022b, p. 199). D'autre part, une hypothèse peut être infirmée et, le cas échéant, la TP peut prendre la forme d'une négation. De plus, un article peut posséder à la fois une question et une hypothèse de recherche. Dans Nundy et coll. (2022b) par exemple, en plus de la question de recherche présente dans le titre, on présente en introduction une hypothèse selon laquelle une introduction bien constituée produit un effet de captation de l'attention du lecteur. Or, la TP (telle qu'exprimée dans la conclusion) ne reprend aucunement les termes de cette hypothèse. Enfin, bien que le lien entre la question de recherche, l'hypothèse de recherche et la TP ne soit pas toujours explicite ou évident, une relation entre ces composantes argumentatives semble statistiquement probable puisqu'il existe un lien de principe entre la question, l'hypothèse et la conclusion qui constitue ce qui est soutenu en dernière instance. Subséquemment, notre hypothèse subordonnée (à H2.3) est que l'introduction est une macrostructure contenant des informations pertinentes pour la détection des TP (H2.3.3).

7.4.2.3.4 L'identification de la thèse principale à l'aide de la discussion

Dans les textes structurés IMRAD, la discussion est une macrostructure textuelle où différents éléments de la recherche sont discutés, notamment les limites, les résultats, les hypothèses, les inférences et les conclusions. Une revue de la littérature par Skelton et Edwards (2000) montre que la discussion est généralement constituée de 5 composantes soit i) les principales conclusions, ii) les forces et faiblesses de l'étude, iii) les forces et faiblesses par rapport à d'autres études, iv) le sens de l'étude en termes d'explications possibles ou d'implications et v) les questions et recherches futures. Les conclusions principales dans la section discussion sont des éléments suffisamment importants, voire essentiels, pour que cette section soit parfois titrée « conclusion ».³⁵ Ce qui distingue la TP de la conclusion est que la première exprime ce qui est assumé par l'auteur à propos du thème principal en réponse à la question de recherche et la seconde résume l'argumentation qui supporte la TP, ainsi que les implications ou avenues pouvant s'avérer intéressantes pour le développement de la discussion (Burstein et Marcu, 2003 ; Durst, 1987 ; Falakmasir *et al.*, 2014). Puisque la TP est un élément central de la conclusion et de la discussion, alors il est probable que l'ensemble de la discussion soit utile lors de l'identification de la TP. Subséquemment, nous formulons l'hypothèse subordonnée (à H2.3) est que la discussion est une structure contenant des informations permettant la détection des TP (H2.3.4).

7.5 Conclusion

En somme, nous avons présenté une théorie structurelle de l'argumentation dans le discours de haut niveau théorique en nous appuyant sur des notions de la RST, de la théorie IMRAD et de la théorie de l'intertextualité. Cette théorie propose trois niveaux structurels de l'argumentation, à savoir les microstructures argumentaires, les macrostructures argumentaires et les structures argumentaires étendues. Ces niveaux permettent une analyse approfondie des arguments dans les textes de haut niveau théorique, offrant ainsi une compréhension plus claire de leur organisation et de leur présentation. Les hypothèses formulées visent à faciliter l'identification de certains éléments clés de cette théorie dans les textes de haut niveau théorique. Ces éléments sont les contextes de citation et les thèses principales. Nous nous focalisons sur deux éléments clés de l'argumentation, car les premiers constituent les moyens par lesquels thèses, arguments, et contre-arguments sont matériellement connectés à d'autres textes, tandis que les seconds sont au cœur même du texte de haut niveau théorique. La thèse principale jouant un rôle

³⁵ La section discussion de la structure *IMRAD* est identifiée comme telle, peu importe comment cette section est titrée, et celle-ci peut parfois se structurer en sous-sections (Swales, 1990).

crucial, notre premier objectif est de déterminer comment repérer les thèses principales. Après les avoir identifiées, notre intention est d'examiner comment elles se lient à d'autres textes pour créer des structures textuelles plus vastes. Bien qu'une analyse bibliométrique révèle des liens extratextuels, ceux-ci restent généraux dans la bibliographie. En revanche, les contextes de citation spécifient ces relations extratextuelles en fonction de leur rôle dans le texte. Par exemple, un argument ou un contre-argument, relié à un autre texte par un contexte de citation, se positionne par rapport à la thèse principale dans la structure IMRAD, caractérisant ainsi le lien extratextuel par des relations intratextuelles. Par exemple, un argument méthodologique lié à un autre texte par un contexte de citation joue indirectement un rôle méthodologique en soutenant ou contestant un argument ou un contre-argument méthodologique. Il est donc possible de caractériser les liens extratextuels selon le rôle de la composante argumentative dans le texte, qui peut être déduit de la macrostructure IMRAD de laquelle elle provient. En d'autres termes, la macrostructure IMRAD d'un contexte de citation aide à caractériser la relation extratextuelle pour une analyse argumentaire. Réaliser cet objectif implique d'abord d'identifier à quelle macrostructure IMRAD les contextes de citation appartiennent. Notre second objectif est donc de déterminer si une telle démarche est faisable et, si c'est le cas, de trouver une méthode efficace pour identifier à quelle macrostructure IMRAD appartiennent les contextes de citation. Les différentes approches proposées pour la détection de segments textuels exprimant la TP incluent l'exploitation des macrostructures IMRAD, notamment le titre, le résumé, l'introduction ou la discussion, ainsi que l'emplacement et les connecteurs de discours. En combinant ces différentes méthodes d'identification, nous comptons améliorer significativement l'analyse des textes de haut niveau théorique en mettant en évidence la structure argumentaire et les éléments essentiels qui soutiennent la TP. Cette approche théorique peut être un outil précieux pour les chercheurs, les analystes et les lecteurs qui cherchent à comprendre plus en profondeur les arguments et les raisonnements avancés dans ce type de texte. En conclusion, la théorie structurelle de l'argumentation, combinant des concepts de la RST, de la théorie IMRAD et de la théorie de l'intertextualité, offre une approche solide pour analyser les textes de haut niveau théorique et en extraire les informations importantes. Les hypothèses formulées pour identifier les contextes de citation et les thèses principales enrichissent cette approche en proposant différentes stratégies d'analyse. Ainsi, cette théorie a le potentiel de contribuer de manière significative à l'étude et à la compréhension des discours de haut niveau théorique dans divers domaines de recherche.

CHAPITRE 8

La détection de l'appartenance des contextes de citation aux macrostructures argumentatives

Au chapitre précédent, nous avons présenté une théorie de l'argumentation fondée sur la RST, la théorie IMRAD et la théorie de l'intertextualité. Subséquemment, nous avons postulé quelques hypothèses à propos de l'interrelation entre les contextes de citation et les macrostructures IMRAD. Dans ce chapitre, nous nous intéressons à la tâche d'identification de l'appartenance des contextes de citation aux macrostructures de type IMRAD dans les articles scientifiques ou philosophiques publiés dans les revues à comité de lecture.

8.1 Introduction

Le but du *Texte Court de Haut Niveau Théorique* publié dans une *Revue à Comité de lecture* (TECHNETIUM) est de présenter et de justifier une ou plusieurs thèses de manière à ce qu'elles soient acceptées par la communauté philoscience (Suppe, 1993, 1998). La structure argumentative de ces articles prend habituellement une forme IMRAD où chacune des macrostructures qui la composent – introduction, méthodes, résultats et discussion – remplit un sous-objectif distinct. Le format IMRAD standardise et articule le récit énonciatif philoscience en rassemblant des informations homogènes répondant aux sous-objectifs propres à chacune des macrostructures IMRAD³⁶. Ce faisant une tâche importante de l'analyse d'argument est d'identifier ces macrostructures IMRAD dans les textes ainsi structurés. La subdivision du texte en macrostructures IMRAD est une opération pouvant être facilitée par la présence de sous-titres associés aux différentes sections. Or, tous les textes ne titrent pas nécessairement les macrostructures IMRAD et, lorsque ces dernières le sont, elles ne sont pas toujours titrées canoniquement « introduction », « méthodes », « résultats » ou « discussion ». Subséquemment, la détection des macrostructures IMRAD est une tâche importante de l'analyse des arguments. Parallèlement, nous nous intéressons à la composante argumentaire particulière qu'est le contexte de citation. Le contexte de citation renvoie à la pratique consistant à citer des sources dans un contexte spécifique pour entre autres soutenir, illustrer ou contester un point avancé. Cette pratique ne consiste pas seulement à faire référence à une source, mais aussi à intégrer la citation d'une manière significative et pertinente par rapport à

³⁶ Spécifions que le discours philoscience n'est pas exclusivement composé de textes au format IMRAD.

l'argument ou à la discussion spécifique en cours. Premièrement, le contexte de citation s'insère localement dans une structure argumentative de bas niveau à l'intérieur d'un texte. Deuxièmement, le contexte de citation s'insère globalement dans une structure argumentative de haut niveau, une macrostructure IMRAD par exemple, laquelle joue un rôle bien défini par rapport à la thèse principale. Troisièmement, le contexte de citation désigne un autre texte et, ce faisant, il s'insère dans une structure argumentative étendue. Le contexte de citation joue ainsi un rôle important qui est celui d'articuler différents niveaux structurels de l'argumentation. Dans cette étude, nous nous intéressons plus particulièrement aux contextes de citation au niveau argumentaire des macrostructures IMRAD. Notre hypothèse principale est qu'il est possible de détecter l'appartenance des contextes de citation aux macrostructures de type IMRAD (H1). Pour ce faire, différents modèles de détection sont possibles. D'abord, comme vu au chapitre 6, différentes techniques de représentation de la donnée textuelle sont possibles et, si certaines représentations sont plus performantes que d'autres, alors il sera intéressant d'en comparer quelques-unes. Notre première hypothèse subordonnée (H1.1) est que les techniques de représentation impliquant un grand nombre de paramètres, inspirées de GPT ou ChatGPT notamment (qui est composé de ~175 milliards de paramètres), sont plus performantes que les autres. Ensuite, différentes méthodes de classification sont possibles, notamment des méthodes fondées par la classification catégorielle, la classification instanciée, la classification linéaire, la classification arborescente et la classification neuronale. Ainsi, notre deuxième hypothèse subordonnée (H1.2) est que la classification linéaire offre de meilleures performances, car ces méthodes optimisées pour résoudre des problèmes convexes garantissent une solution unique optimale le cas échéant. Subséquemment, notre stratégie consistera à tester diverses représentations textuelles ainsi que différents classifieurs sur un corpus composé de contextes de citation annotés de la macrostructure IMRAD de laquelle ceux-ci proviennent. À cet effet, le corpus *In-text Reference Corpus* (InTeReC) se prête bien à cette tâche, car il est composé de contextes de citation issus de TECHNETIUM et annotés de leur provenance macrostructurale.

8.2 Contexte et état de l'art

La notion de « contexte de citation » réfère à un « passage particulier ou un énoncé dans le texte citant contenant une ou plusieurs références » (Small, 1982, p. 288, notre trad.). Autrement dit, il constitue un segment de texte où on réfère à un ou d'autres textes. Les contextes de citation sont des composantes discursives suffisamment importantes pour donner lieu à un champ de recherche spécifique. L'analyse des contextes de citation permet l'exploration des idées, des thèmes, des thèses et des arguments importants qui sont utilisés et diffusés d'un texte source vers des travaux de citation ultérieurs (Anderson et Lemken,

2023). Une question de recherche de ce domaine est d'identifier dans quelle mesure l'utilisation des œuvres citées est périphérique ou fondamentale par rapport aux principaux arguments des textes citant (Anderson et Lemken, 2019 ; Burton-Jones, 2014 ; Lounsbury et Carberry, 2005). Autrement dit, les contextes de citation sont des composantes argumentatives importantes en ce que certaines d'entre elles sont reliées à des arguments plus centraux, voire aux thèses principales.

8.2.1 Les techniques de représentation

La représentation vectorielle des textes est nécessaire pour tester nos hypothèses. Comme vu au chapitre 6, plusieurs techniques permettent de représenter le texte au moyen de vecteurs numériques. Dans cette étude, nous utiliserons certaines représentations vectorielles classiques et d'autres plus modernes. La première – nommée BOW – est une représentation classique de type « sac de mots » (de l'anglais : *bag of words* ou BOW). Ce type de représentation vectorielle ne tient pas compte de l'ordre ou du contexte des mots. La seconde – nommée DBOW – est une représentation de type doc2vec de variante DBOW. DBOW est générée à l'aide d'un modèle de réseau neuronal. Le modèle apprend à prédire un mot en fonction de son contexte, et les vecteurs résultants représentent chaque mot et document du corpus. DBOW peut capturer certaines relations sémantiques entre les mots, mais ne tient pas compte de l'ordre des mots dans le texte, sinon indirectement au moyen d'une fenêtre de mots qui se déplace de gauche à droite lors de l'apprentissage. La troisième est une représentation de type BERT générée à l'aide de *Transformers* préentraînés sur un très grand corpus de texte. BERT est un modèle bidirectionnel qui peut prendre en compte à la fois le contexte et l'ordre des mots dans le texte. Plusieurs représentations BERT seront utilisées.³⁷ D'abord, les représentations originales du modèle BERT de base et du modèle BERT large sont employées. Ensuite, deux (2) modèles de représentation préentraînés provenant de Hugging Face sont utilisés (Wolf *et al.*, 2020). Le premier est basé sur BERT large et son entraînement a été affiné sur une tâche d'argumentation *NLI*³⁸ et une tâche d'évaluation de similarité sémantique *STS*³⁹ (Reimers et

³⁷ Nous utilisons l'API de Hugging Face pour les modèles de représentations du corpus de type BERT. Hugging Face est connu pour sa bibliothèque de code ouvert nommée *Transformers*, qui permet d'accéder à divers modèles de traitement du langage naturel, dont BERT.

³⁸ Une tâche de *Natural Language Inference (NLI)* consiste à déterminer la relation logique entre deux segments de texte. Étant donné une prémisse et une hypothèse, la tâche consiste à déterminer si l'hypothèse est impliquée, contredite ou neutre par rapport à la prémisse.

³⁹ La similarité sémantique textuelle (*semantic textual similarity* en anglais) est une tâche d'évaluation de la performance d'un modèle de traitement du langage naturel sur sa capacité à mesurer le degré de similarité sémantique entre deux phrases. Un ensemble de données permettant cet entraînement est *STS* (*Semantic Textual*

Gurevych, 2019). Le second est basé sur MPNet (de l'anglais : *Masked and Permuted language modeling Network*) qui s'inspire de BERT en ajoutant une tâche de prédiction des mots permutés à la tâche d'identification des mots masqués (Song *et al.*, 2020). La version MPNet utilisée est celle particulièrement performante suggérée par Hugging Face dont l'entraînement a été affiné sur une panoplie de tâches impliquant des paires dont, entre autres, des couples questions-réponses, titres-résumés, titres-réponses, des citations, etc.⁴⁰ Le dernier modèle BERT utilisé provient de BERT de base, mais avec un entraînement affiné sur une tâche de classification IMRAD. Enfin, le dernier modèle de représentation est développé par l'entreprise OpenAI et connue sous le nom de « text-embedding-ada-002 » qui est généré à l'aide de *Transformers* préentraînés sur un très grand corpus de texte.

8.2.2 Les méthodes de classification

Différentes techniques de classification issues du domaine de l'apprentissage machine permettent d'effectuer cette tâche et notre objectif est d'évaluer leur performance. Parmi ces techniques, certaines permettent la classification multiclassées et, lorsque ce n'est pas le cas, il est possible d'utiliser des stratégies le permettant, notamment celle nommée « un contre un »⁴¹ ou celle nommée « un contre tous »⁴². Aussi, les méthodes de classification qui seront présentées permettront la validation croisée qui est une technique utilisée afin d'évaluer les performances d'un modèle et ses hyperparamètres. L'objectif de la validation croisée est d'estimer dans quelle mesure le modèle se généralisera à de nouvelles données en évaluant ses performances sur un sous-ensemble de données qui n'a pas été utilisé pendant l'entraînement. L'idée de base derrière la validation croisée est de diviser les données en deux ou plusieurs sous-ensembles et d'utiliser un sous-ensemble pour tester le modèle tandis que les autres sous-ensembles

Similarity Benchmark) qui contient des paires de phrases, dont chacune a été annotée avec un score continu indiquant le degré de similitude entre les deux phrases.

⁴⁰ Le site de Hugging Face spécifie que le modèle « all-mpnet-base-v2 » est le plus performant (*provides the best quality*), voir https://www.sbert.net/docs/pretrained_models.html. Pour plus de détail sur les *corpora*, voir <https://huggingface.co/sentence-transformers/all-mpnet-base-v2>

⁴¹ La stratégie « un contre un » implique l'entraînement de plusieurs classificateurs binaires pour chaque paire possible de classes dans l'ensemble de données. Par exemple, s'il y avait 3 classes (A, B et C), alors on entraînerait 3 classificateurs binaires : un pour A vs B, un pour A vs C et un pour B vs C. Ensuite, chaque classificateur binaire prédirait la classe du point de données, et la classe avec le plus de votes serait sélectionnée comme prédiction finale.

⁴² La classification « un contre tous » implique la formation d'un seul classificateur binaire pour chaque classe de l'ensemble de données. Pendant l'entraînement, chaque classificateur binaire est entraîné à distinguer les points de données d'une classe contre ceux de toutes les autres classes. Ensuite, chaque classificateur binaire prédit la probabilité que le point de données appartienne à sa classe correspondante, et la classe avec la probabilité la plus élevée est sélectionnée comme prédiction finale.

sont utilisés pour l'entraînement. Ce processus peut être répété plusieurs fois, chaque sous-ensemble étant utilisé une fois pour les tests et les sous-ensembles restants utilisés pour l'entraînement. La moyenne des résultats permet ensuite d'estimer les performances du modèle. Une validation croisée utilisée est nommée *K-Folds*, où les données sont divisées en K sous-ensembles (*folds*) de taille égale. Le modèle est entraîné sur l'union de $K-1$ sous-ensembles et testé sur le sous-ensemble restant, et ce processus est répété K fois, chaque sous-ensemble étant utilisé une fois pour le test. Les mesures de performance, telles que la précision ou l'*AUC*⁴³, sont calculées pour chaque sous-ensemble, et la moyenne est considérée comme l'estimation finale des performances du modèle. La validation croisée peut également être utilisée pour régler les hyperparamètres d'un modèle. Cela se fait en formant et en évaluant le modèle sur différentes valeurs des hyperparamètres et en choisissant celles donnant les meilleures performances sur les données de validation. Une autre méthode souvent utilisée est la validation croisée stratifiée en K sous-ensembles (*Stratified K-Folds*). L'idée derrière cette approche est de diviser les données en K sous-ensembles (*folds*), tout en veillant à ce que chaque sous-ensemble ait à peu près la même proportion d'exemples positifs et négatifs. Les données sont d'abord partitionnées en K sous-ensembles disjoints, où chacun contient des exemples de l'une des K classes. Pour chaque sous-ensemble, un nombre fixe d'exemples est sélectionné au hasard dans chacun des K sous-ensembles, de sorte que la résultante ait à peu près la même proportion d'exemples de chaque classe que les données d'origine. Cette approche assure que chaque sous-ensemble a une distribution similaire de classes, ce qui peut réduire la variance de l'estimation de validation croisée et fournir une estimation plus précise des performances du modèle. Aussi, cette approche permet de prévenir le surajustement en garantissant que le modèle est évalué sur un échantillon représentatif des données. De plus, spécifions que certaines techniques font des hypothèses à propos des données en jeu et, ce faisant, elles modélisent la donnée. D'autres font des hypothèses concernant la procédure de prédiction et, donc, elles modélisent la fonction algorithmique. Mais, comme le souligne Breiman (2001b), rappelons que plusieurs modèles concurrents peuvent décrire/prédire aussi efficacement les données; que la simplicité est un gage de compréhensibilité et non pas de prédictibilité; et que la réduction de la dimensionnalité n'est pas, non plus, un gage de prédictibilité. Enfin, parmi toutes les approches possibles, nous examinerons plus particulièrement les modèles de

⁴³ L'*AUC* (*Area Under the Curve*) mesure l'aire sous la courbe *ROC* (*Receiver Operating Characteristic*), qui est un tracé du taux de vrais positifs par rapport au taux de faux positifs à différentes valeurs seuils. La courbe *ROC* montre à quel point le modèle peut faire la distinction entre les exemples positifs et négatifs sur une plage de valeurs de seuil possibles.

classification instanciée, de classification linéaire, de classification arborescente, de classification neuronale et de classification basée sur la catégorisation⁴⁴.

8.2.2.1 Les modèles de classification basée sur la catégorisation

La classification basée sur la catégorisation (*clustering*), telle que la K-moyenne et la catégorisation gaussienne, regroupent des points de données similaires en clusters en fonction de leur similarité. Le principe des classificateurs basés sur la catégorisation est de trouver des regroupements émergents automatiquement des données pour être utilisés afin de catégoriser de nouveaux points de données. Cette classification ne nécessite pas de données d'apprentissage étiquetées et peut donc être utilisée pour des tâches d'apprentissage non supervisé. Une fois les catégories identifiées, l'algorithme attribue le nouveau point de données en fonction de sa similarité avec les points de données existants dans chaque catégorie. Spécifions que la classification basée sur la catégorisation peut être utilisée dans les cas particuliers où des données d'apprentissage étiquetées sont disponibles, notamment pour évaluer si les classes ou étiquettes attribuées aux données peuvent correspondre à des catégories automatiquement détectables sans connaissance a priori des classes à découvrir.⁴⁵ Lorsqu'un problème de classification est soluble par ce type d'approche, cela indique qu'il existe un partitionnement implicite dans l'espace des caractéristiques permettant de retrouver les classes recherchées.

8.2.2.1.1 Le modèle des K-moyennes

Le modèle des K-moyennes est un algorithme d'apprentissage automatique non supervisé proposé pour la première fois par Stuart Lloyd en 1957 (Lloyd, 1982). Cet algorithme vise à partitionner un ensemble de données en K regroupements minimisant la somme des distances au carré entre les points de données et leur centre le plus proche. Cet algorithme fonctionne en affinant de manière itérative une estimation initiale pour les centroïdes des K regroupements jusqu'à la convergence du modèle. Les centroïdes représentent le centre de chaque regroupement et l'algorithme attribue chaque point de données au

⁴⁴ Outre ces approches qui seront investiguées, notons qu'il existe aussi des méthodes de classification basées sur le théorème de Bayes, sur des règles ou sur différentes combinatoires.

⁴⁵ Dans ce cas, notons que la classification fondée sur la catégorisation n'identifie pas quelle catégorie correspond à quelle étiquette. Cependant, il est possible de calculer la performance de cette classification par rapport aux données étiquetées. Pour un même nombre de catégories et d'étiquettes, on peut attribuer, une à une, chacune des catégories à chacune des étiquettes, puis pour chaque permutation possible, calculer la performance collective de ces attributions pour chaque permutation. Dans cette étude, la permutation offrant la meilleure performance en termes d'exactitude est sélectionnée de manière à attribuer de manière optimale une catégorie à chaque étiquette.

centroïde le plus proche. Plus précisément, l'algorithme sélectionne au hasard K points de données dans l'ensemble de données pour servir de centroïdes initiaux. Ensuite, chaque point de données est affecté au centroïde le plus proche. Puis, les centroïdes de chaque cluster sont mis à jour en fonction de la moyenne des points de données attribués au regroupement. Enfin, les deux étapes précédentes sont répétées jusqu'à la convergence qui est atteinte lorsque les centroïdes ne changent plus de manière significative. L'algorithme des K -moyennes présente certains avantages tels que sa simplicité, sa rapidité et son efficacité pour les grands ensembles de données. Cependant, il présente également certaines limites, notamment l'obligation de spécifier le nombre de regroupements, et l'algorithme peut converger vers un minimum local plutôt qu'un minimum global, ce qui signifie que le résultat final peut dépendre des centroïdes initiaux. Afin de contrer ce problème, Arthur et Vassilvitskii (2007) proposent une solution nommée « kmean++ » permettant une initialisation des centroïdes qui surpasse substantiellement la méthode standard en termes de vitesse et de précision. Kmean++ sélectionne les centroïdes initiaux en utilisant un échantillonnage basé sur une distribution de probabilité empirique de la contribution des points à l'inertie globale. Plus précisément, l'algorithme échantillonne le premier centroïde au hasard à partir de l'ensemble de données et les autres $k-1$ centroïdes de manière itérative selon l'échantillonnage D^2 où la probabilité qu'un point de données devienne le centroïde suivant est proportionnelle à sa distance au carré du centroïde le plus proche choisi jusqu'à présent.⁴⁶

8.2.2.1.2 Le modèle de mélange gaussien

Le modèle de mélange gaussien est un modèle probabiliste d'apprentissage automatique utilisé pour la catégorisation non supervisée. Le modèle est un « mélange » au sens où il est composé de multiples distributions gaussiennes, également appelées distributions normales. Le modèle de mélange gaussien suppose que chaque point d'un ensemble de données est généré à partir de l'une des nombreuses distributions gaussiennes sous-jacentes. L'algorithme tente d'estimer les paramètres de ces distributions, qui incluent la moyenne et la variance de chaque distribution, ainsi que les proportions relatives des points de données provenant de chaque distribution. Une fois les paramètres du modèle estimés, l'algorithme affecte chaque point de données à la distribution gaussienne qui est la plus susceptible de l'avoir généré. Le modèle de mélange gaussien fonctionne en améliorant de manière itérative l'estimation initiale des

⁴⁶ Nous utilisons l'implémentation de modèle des K -moyennes du *progriciel Scikit-learn* en langage Python. Par défaut, la méthode d'initialisation des centroïdes est « greedy k-means++ » qui diffère de k-means++ par l'identification de plusieurs candidats à chaque étape d'échantillonnage et en choisissant comme centroïde le meilleur parmi eux (Bhattacharya *et al.*, 2019). Notons que le modèle utilisé est composé de 4 centroïdes.

paramètres du modèle. L'algorithme commence par initialiser aléatoirement les paramètres du modèle. Puis, sur la base des estimations actuelles des paramètres, l'algorithme calcule la probabilité que chaque point de données appartienne à chacune des distributions gaussiennes du modèle. Ensuite, l'algorithme met à jour les estimations des paramètres des distributions gaussiennes en fonction des probabilités calculées à l'étape précédente. Enfin, l'algorithme répète les deux étapes précédentes jusqu'à la convergence, ce qui signifie que les paramètres des distributions gaussiennes ne changent plus de manière significative.⁴⁷

8.2.2.2 Les modèles de classification instanciée

Les modèles de classification instanciée, tels que celui des K plus proches voisins, des voisins de rayon, ou des plus proches centroïdes, appartiennent à la famille des algorithmes dont le principe de classification est la similarité des nouveaux cas aux instances présentées lors de l'apprentissage. Ces modèles sont généralement non paramétriques au sens où aucune hypothèse sur la distribution des données sous-jacentes n'est nécessaire. Ceux-ci utilisent toutes les instances de l'ensemble de données d'apprentissage lors de la modélisation et le classement des nouveaux points de données s'effectue en fonction de leur similarité aux données d'apprentissage.

8.2.2.2.1 La méthode des K plus proches voisins

La méthode des K plus proches voisins (*K-nearest neighbors*) est un algorithme d'apprentissage supervisé développé par Fix et Hodges (1951) ainsi que Cover et Hart (1967). Celui-ci fonctionne en trouvant les K voisins les plus proches d'un point de données dans un espace de caractéristiques. La valeur de K est un hyperparamètre qui doit être choisi avant de former le modèle. Une fois que les plus proches K voisins sont identifiés, l'algorithme attribue le nouveau point de données à la classe la plus commune parmi ses K voisins les plus proches. La métrique utilisée pour mesurer la distance entre les points de données est généralement la distance euclidienne, mais d'autres métriques telles que la distance Manhattan ou la distance cosinus peuvent être utilisées. La méthode des K plus proches voisins peut gérer à la fois les problèmes de classification binaire et multiclasse. Dans le cas binaire, l'algorithme attribue le nouveau point de données à l'une des deux classes. Dans le cas multiclasse, l'algorithme attribue le nouveau point de données à la classe la plus courante parmi ses K voisins les plus proches. L'un des avantages de

⁴⁷ Nous utilisons l'implémentation de modèle de mélange gaussien du *progiciel Scikit-learn* en langage Python. Par défaut, le modèle est initialisé à l'aide de la méthode des k -moyennes. Les centroïdes identifiés sont utilisés comme moyenne des points de données de chaque catégorie. Notons que le modèle utilisé est composé de 4 gaussiennes.

l'algorithme est qu'il est simple à mettre en œuvre, mais il peut être computationnellement coûteux d'évaluer les distances entre les points de données. Différentes méthodes permettent d'évaluer ces distances pour trouver les K plus proches voisins, notamment la force brute, *Ball Tree* et *KD Tree*. La méthode de la force brute compare le point de données de test avec tous les points de données d'apprentissage, puis calcule la distance entre eux et renvoie les K points de données avec les distances les plus courtes. La méthode *Ball Tree* fonctionne en partitionnant récursivement les points de données en sous-ensembles topologiques nommées « balles » jusqu'à ce que chacune contienne un nombre fixe de points. Lorsqu'un nouveau point doit être classé, l'algorithme commence à la racine de l'arbre et descend jusqu'au nœud feuille qui contient le point de test. On recherche ensuite les voisins les plus proches dans cette balle et renvoie les K -plus proches voisins. La méthode *KD Tree* fonctionne en partitionnant récursivement les points de données à l'aide d'hyperplans jusqu'à ce que chaque sous-espace contienne un nombre fixe de points. Lorsqu'un nouveau point doit être classé, l'algorithme commence à la racine de l'arbre et descend jusqu'au nœud feuille qui contient le point de test. Il recherche ensuite les voisins les plus proches dans les hyperplans et renvoie les K -plus proches voisins.⁴⁸

8.2.2.2.2 La méthode des voisins de rayon

La méthode des voisins de rayon (*radius neighbors*) est un algorithme d'apprentissage supervisé popularisé par Cover et Hart (1967). Celui-ci est similaire aux K plus proches voisins, mais se distingue par son utilisation de tous les voisins se trouvant dans un rayon donné plutôt qu'un nombre fixe de voisins. Le rayon est un hyperparamètre qui doit être défini avant de former le modèle. Une fois que les voisins dans le rayon sont identifiés, l'algorithme attribue le nouveau point de données à la classe la plus commune parmi le voisinage. En d'autres termes, l'algorithme compte le nombre d'échantillons dans chaque classe du voisinage et sélectionne celle avec le nombre le plus élevé comme classe prédite du nouveau point de données. La métrique utilisée pour mesurer la distance entre les points de données est généralement la distance euclidienne, mais d'autres métriques sont possibles. Comme pour les K plus proches voisins, différentes méthodes permettent d'évaluer les distances pour trouver les voisins de rayon, dont la force

⁴⁸ Nous utilisons l'implémentation des K plus proches voisins du *progriciel Scikit-learn* en langage Python. La configuration par défaut est $K = 5$ et la méthode utilisée est déterminée automatiquement entre la force brute, *Ball Tree* et *KD Tree*.

brute, *Ball Tree* et *KD Tree*. Enfin, cette méthode permet de traiter les problèmes de classification binaire et multiclasse.⁴⁹

8.2.2.2.3 La méthode du plus proche centroïde

La méthode du plus proche centroïde, aussi connu sous le nom d'algorithme de Rocchio, est un algorithme introduit par Rocchio (1971). Cet algorithme fonctionne en calculant d'abord le centroïde de chaque classe dans les données d'apprentissage, où le centroïde est simplement la moyenne de tous les vecteurs de caractéristiques appartenant à cette classe. Pendant la phase de prédiction, l'algorithme calcule la distance entre le nouveau point de données et chacun des centroïdes de classe. Il attribue ensuite le nouveau point de données à la classe avec le centroïde le plus proche. La métrique de distance peut être l'une des mesures de distance standard, telles que la distance euclidienne ou la similarité cosinus. Un avantage de cet algorithme est sa robustesse par rapport aux valeurs aberrantes et au bruit dans les données, car le calcul du centroïde prend la moyenne de tous les vecteurs de caractéristiques de la classe. Cependant, un inconvénient de l'algorithme est qu'il suppose que les données sont normalement distribuées autour des centroïdes de classe, ce qui n'est pas toujours vrai en pratique. De plus, l'algorithme est sensible aux ensembles de données déséquilibrés, car il attribue des poids égaux à tous les vecteurs de caractéristiques d'une classe lors du calcul du centroïde.⁵⁰

8.2.2.3 Les modèles de classification linéaire

La classification linéaire est un type d'algorithmes qui classent les données au moyen de limites de décision linéaires ou de fonctions linéaires séparant les classes dans l'espace des caractéristiques. Les classifieurs linéaires peuvent être entraînés à l'aide de différentes techniques d'optimisation, telles que la régression logistique ou les machines à vecteurs de support. Spécifions qu'un type particulier de classification linéaire permet la classification de données non linéairement séparables dans l'espace original de caractéristiques. Cette approche basée sur une fonction de noyau (*kernel*) utilise une astuce selon laquelle les données en entrée sont transformées dans un espace de plus grande dimension dans lequel une séparation linéaire est possible. Entre autres, cet espace de redescription peut être la résultante de fonctions polynomiale,

⁴⁹ Nous utilisons l'implémentation des voisins de rayon du *progiciel Scikit-learn* en langage Python. La configuration par défaut propose un rayon de 1 et, comme pour les K plus proches voisins, la méthode déterminant le voisinage est déterminée automatiquement entre la force brute, *Ball Tree* et *KD Tree*.

⁵⁰ Nous utilisons l'implémentation des plus proches centroïdes du *progiciel Scikit-learn* en langage Python. La configuration par défaut propose une distance euclidienne.

sigmoïde ou de base radiale (*Radial Basis Function* ou *RBF*). Enfin, les modèles de classification linéaire résolvent des problèmes d'optimisation convexes, et ils le font généralement en minimisant une fonction de perte convexe et/ou en trouvant une limite de décision convexe. Dans ce contexte, la convexité est soit une propriété de la fonction de perte, soit de l'espace de solution (l'ensemble de toutes les limites de décision possibles). Cette propriété implique que tout minimum local est également un minimum global unique, ce qui simplifie considérablement le processus d'optimisation de ces classifieurs (Boyd et Vandenberghe, 2004).

8.2.2.3.1 La descente de gradient stochastique.

Introduit par Robbins et Monro (1951), la descente de gradient stochastique (*Stochastic Gradient Descent* ou *SGD*) est une méthode d'optimisation utilisée afin de minimiser une fonction objectif pouvant résoudre des problèmes convexes (ou non). Cette méthode peut être utilisée afin d'optimiser les paramètres d'un modèle de classification linéaire. Le cas échéant, on définit un modèle qui prédit la classe de sortie en fonction d'une combinaison linéaire des caractéristiques en entrée. Cette combinaison est calculée en multipliant chaque caractéristique d'entrée par un poids, en additionnant les produits et en ajoutant un biais. Les pondérations et le biais sont appris lors de l'entraînement en utilisant la descente de gradient stochastique qui est un algorithme d'optimisation itératif mettant à jour les paramètres du modèle par rapport à une fonction objectif permettant de mesurer la différence entre la classe de sortie prévue et la classe de sortie réelle. La fonction objectif utilisée par un classifieur fondé par *SGD* peut être sélectionnée parmi une variété d'options, notamment la fonction *hinge* qui est définie comme le maximum entre 0 et la différence entre 1 et l'étiquette de classe réelle multipliée par le score prédit. Enfin, cette approche se caractérise par une mise à jour des paramètres suivant le calcul des gradients par rapport à la fonction objectif sur un petit sous-ensemble sélectionné au hasard dans l'ensemble des données d'apprentissage (un seul point de données ou un petit lot de points de données). Spécifions que la taille du sous-ensemble est un hyperparamètre qui peut être réglé pour optimiser les performances. Le taux d'apprentissage, un autre hyperparamètre, permet de contrôler la taille du pas des mises à jour des paramètres et peut également être réglé pour optimiser les performances. Enfin, il est possible d'intégrer une régularisation au modèle afin d'éviter le surapprentissage en ajoutant un terme de pénalité à la fonction objectif de manière à pénaliser les poids excessifs. La force de la régularisation est contrôlée par un hyperparamètre

appelé alpha, qui détermine le compromis entre la minimisation de l'erreur et la réduction de l'ampleur des poids.⁵¹

8.2.2.3.2 Les machines à vecteurs de support (SVM)

Les machines à vecteurs de support (*Support Vector Machine* ou *SVM*) sont des algorithmes d'apprentissage supervisé introduits par Vladimir Vapnik et ses collègues (Cortes et Vapnik, 1995 ; Vapnik, 1963, 1995 ; Vapnik *et al.*, 1996). Leur objectif est de trouver un hyperplan séparant deux ensembles de points de données en fonction de leurs caractéristiques ou attributs. Les SVM essaient de trouver la frontière de décision la plus éloignée des points de données entre deux classes. Lorsque les données ne sont pas linéairement séparables et, donc, qu'il n'y a pas de marge de séparation claire entre les deux classes, on peut utiliser les SVM sur un espace de redescription des données à l'aide d'une fonction de noyau (Aizerman *et al.*, 1964). Les SVM sont des classifieurs binaires, mais ils peuvent être utilisés pour les problèmes de classification multiclassés à l'aide des techniques « un contre un » ou « un contre tous » (Vapnik, 1995). La fonction objectif des SVM consiste à trouver l'hyperplan optimal à l'aide d'un problème d'optimisation convexe. Celui-ci consiste à minimiser le taux d'erreur, tout en maximisant la marge entre les différentes classes. Ce problème d'optimisation est résolu à l'aide de diverses techniques telles que la programmation quadratique, la descente de gradient stochastique (Abeykoon *et al.*, 2022) ou l'optimisation minimale séquentielle (Platt, 1998). Une fois l'hyperplan optimal trouvé, il peut être utilisé pour prédire la classe des nouveaux points de données en calculant la distance entre le point de données et l'hyperplan.⁵²

⁵¹ Nous utilisons l'implémentation de la descente de gradient stochastique du *proiciel Scikit-learn* en langage Python. Par défaut, la fonction objectif utilisée est *hinge*, la fonction de pénalité est L2 ajustée avec le multiplicateur alpha de valeur 0,0001, lequel définit également le taux d'apprentissage tel que $1 / (\alpha * (t + t0))$ où t est l'époque (*epoch*) actuelle de l'itération de l'apprentissage et $t0$ est une époque choisie par une heuristique proposée par Bottou (2012). La stratégie multiclassée est « un contre tous ».

⁵² Nous utilisons les implémentations SVM du *proiciel Scikit-learn* en langage Python. Deux implémentations sont possibles pour le SVM linéaire. « *Linearsvm* » utilise une stratégie multiclassée « un contre tous » ainsi qu'une fonction de perte L2 aussi nommée Hinge au carré. Plus précisément, la fonction de perte Hinge au carré est utilisée pour maximiser la marge, c'est-à-dire la distance entre la frontière de décision (ou hyperplan) et les exemples les plus proches (les vecteurs de support), tout en minimisant la classification erronée des exemples (Lee et Lin, 2013). « *SVC* » utilise une stratégie multiclassée « un contre un », permet la méthode noyau et, par défaut, utilise la fonction de perte Hinge (mais, pas au carré). Nous utilisons le noyau de fonction linéaire (sans redescription) et celui de fonction radiale (*RBF*). De plus, *SVC* inclut un paramètre de régularisation qui induit un compromis entre l'obtention d'une faible erreur d'apprentissage et l'obtention d'un modèle de faible complexité. Nous utilisons la valeur par défaut de 1, c.-à-d. que la marge est étroite et le modèle est moins tolérant aux erreurs

8.2.2.3.3 La régression logistique

La régression logistique est un type de modèle statistique introduit par Berkson (1944) et utilisé pour les tâches de classification binaire où l'objectif est de prédire la probabilité qu'une entrée appartienne à l'une des deux classes possibles. L'idée derrière la régression logistique est de transformer un modèle de régression linéaire, qui prédit une valeur continue, en un modèle qui prédit une probabilité entre 0 et 1. Ceci est réalisé grâce à l'utilisation d'une fonction logistique, qui cartographie toute entrée à valeur réelle à une valeur comprise entre 0 et 1. De plus, la régression logistique peut être étendue aux problèmes de classification multiclassé à l'aide, notamment, des techniques « un contre un », « un contre tous » ou la perte multinomiale (également appelée perte d'entropie croisée ou perte logarithmique). À cet effet, la perte multinomiale quantifie la différence entre les probabilités prédites et les véritables étiquettes de classe sur l'ensemble de la distribution de probabilité. L'objectif de la régression logistique multiclassé est de trouver les paramètres du modèle qui minimisent la perte logarithmique multinomiale moyenne sur toutes les observations. En pratique, des algorithmes d'optimisation comme la descente de gradient ou ses variantes sont utilisés pour trouver les paramètres du modèle qui minimisent la perte pendant le processus d'apprentissage. Ces algorithmes mettent à jour de manière itérative les coefficients du modèle pour trouver le meilleur ajustement pour l'ensemble de données et minimiser l'erreur. Le processus se poursuit jusqu'à la convergence ou un nombre prédéfini d'itérations.⁵³ La régression logistique peut s'utiliser avec validation croisée.⁵⁴

8.2.2.3.4 La régression *Ridge*

La régression *Ridge* est un algorithme de classification linéaire introduit par Hoerl et Kennard (1970b, 1970a) qui fonctionne en attribuant des poids aux données en entrée et en les utilisant pour faire une

de classification. Une valeur proche de zéro impliquerait une marge plus large et un modèle plus tolérant aux erreurs de classification. Enfin, l'usage du noyau RBF implique la définition du coefficient de noyau (gamma) qui contrôle la largeur de la fonction gaussienne utilisée pour calculer la similarité entre les points de données. Une petite valeur gamma se traduit par un grand rayon de similarité et une frontière de décision plus lisse, tandis qu'une grande valeur gamma se traduira par un rayon de similarité plus petit et une frontière de décision plus complexe qui correspond plus étroitement aux données d'apprentissage. Nous sélectionnons la valeur gamma par défaut qui est calculée comme l'inverse du nombre de caractéristiques multiplié par la variance moyenne pour les données d'apprentissage.

⁵³ Nous utilisons l'implémentation de la régression logistique du *progiciel Scikit-learn* en langage Python avec une la stratégie de décision multiclassé par défaut qui est fondée par l'entropie croisée.

⁵⁴ Nous utilisons l'implémentation de la régression logistique avec validation croisée du *progiciel Scikit-learn* en langage Python avec la stratégie de décision multiclassé par défaut qui est fondée par l'entropie croisée. Par défaut, le nombre de sous-ensembles (*folds*) est de 5.

prédiction sur la classe en sortie. L'algorithme est une version régularisée de la régression linéaire et est particulièrement utile pour les ensembles de données à haute dimensionnalité avec un potentiel de corrélation entre les caractéristiques. Cet algorithme fonctionne en définissant d'abord un modèle linéaire qui prédit la classe de sortie en fonction d'une combinaison linéaire des données en entrée. Cette combinaison linéaire est calculée en multipliant chaque caractéristique d'entrée par un poids, en additionnant les produits et en ajoutant un terme de biais. Les poids et le terme de biais sont appris au cours de l'entraînement à l'aide d'une fonction objectif qui pénalise les poids importants. Cette fonction intègre un terme de régularisation conçu pour éviter le surajustement et consiste en une pénalité sur la grandeur des poids et est proportionnelle au carré de la norme L2 du vecteur poids. La norme L2 est calculée en additionnant les carrés des poids individuels. Pendant l'entraînement, cet algorithme trouve les poids qui minimisent la fonction objectif, sous réserve de la pénalité de régularisation. La force de la régularisation est contrôlée par un hyperparamètre appelé alpha, qui détermine le compromis entre la minimisation de l'erreur et la réduction de l'ampleur des poids. Une valeur élevée d'alpha conduit à plus de régularisation, tandis qu'une valeur inférieure conduit à moins de régularisation. Pour faire une prédiction sur un nouveau point de données, le classifieur *Ridge* prend la combinaison linéaire des caractéristiques d'entrée en utilisant les poids appris et le terme de biais. Si la valeur résultante est supérieure à un seuil, le classifieur prédit la classe positive, sinon il prédit la classe négative.⁵⁵ La régression *Ridge* peut être utilisée avec une validation croisée afin de sélectionner les meilleurs hyperparamètres du modèle.⁵⁶

8.2.2.3.5 La classification passive-agressive

Introduite par Crammer et coll. (2006), la classification passive-agressive est un type d'algorithme de classification d'apprentissage « en ligne », c.-à-d. que l'apprentissage à partir de nouvelles données est possible au fur et à mesure de leur disponibilité sans avoir à reprendre tout le processus d'apprentissage du modèle. *A contrario*, les algorithmes de classification qui ne sont pas « en ligne » nécessitent que toutes les données de formation soient disponibles au départ, et le modèle doit être formé à partir de zéro chaque fois que de nouvelles données sont ajoutées. La classification passive-agressive est également

⁵⁵ Nous utilisons l'implémentation de la régression *Ridge* du *progiciel Scikit-learn* en langage Python. Par défaut, la valeur du paramètre alpha est de 1. Pour la classification multiclasse, le problème est traité comme une régression multiple et la classe prédite correspond à la sortie avec la valeur la plus élevée.

⁵⁶ Nous utilisons l'implémentation de la régression *Ridge* avec validation croisée du *progiciel Scikit-learn* en langage Python. Par défaut, la validation croisée est de type « leave-one-out » (*LOOCV*) où toutes les observations sauf une sont utilisées dans l'ensemble d'apprentissage. L'hyperparamètre testé est la force de régularisation (alpha).

considérée comme un type d'algorithme de classification linéaire, car elle modélise la relation entre les entités en entrée et les classes en sortie à l'aide d'une fonction linéaire. Cependant, son fonctionnement est efficace même lorsque la relation entre les caractéristiques et les classes n'est pas linéaire, grâce à l'utilisation de l'astuce du noyau ou d'autres techniques de transformation non linéaire. Le classifieur passif-agressif fonctionne en effectuant une série de mises à jour d'un vecteur de poids, qui est utilisé pour calculer un score pour chaque instance d'entrée. Le score est ensuite comparé à une valeur de seuil, et l'instance est affectée à l'une des deux classes selon que le score est supérieur ou inférieur au seuil. Pendant l'entraînement, des instances étiquetées sont présentées au classifieur passif-agressif qui fait des prédictions basées sur le vecteur de poids actuel. Si la prédiction est correcte, l'algorithme n'apporte aucune modification au vecteur de poids. Cependant, si la prédiction est incorrecte, l'algorithme met à jour le vecteur de poids de manière à minimiser l'erreur. Le nom « passif agressif » vient du fait que l'algorithme est conçu pour être très agressif dans la correction des erreurs, mais passif dans les situations où la bonne réponse est obtenue. Cela rend l'algorithme particulièrement efficace dans les situations où les données sont bruitées ou lorsqu'il y a des changements fréquents dans la distribution des données.⁵⁷

8.2.2.4 Les modèles de classification arborescente

La classification arborescente est basée sur des structures décisionnelles ramifiées, telles que les arbres de décision et les forêts aléatoires, et utilise ces structures arborescentes pour classer les données. Le principe des classifieurs arborescents est de diviser récursivement l'espace des caractéristiques en sous-ensembles plus homogènes par rapport à la variable cible. Chaque nœud de l'arbre correspond à un sous-ensemble de l'espace des caractéristiques, et les arêtes correspondent aux règles qui définissent comment diviser les données. À terme, l'algorithme assigne ainsi un nouveau point de données à la classe associée au nœud feuille de l'arbre qui le contient.

8.2.2.4.1 Les arbres de décision

Les arbres de décision sont un type d'algorithme d'apprentissage supervisé introduit par Morgan et Sonquet (1963) et utilisé pour les tâches de classification. L'algorithme fonctionne en construisant une structure arborescente dans laquelle chaque nœud interne représente un test sur une caractéristique particulière des données d'entrée, et chaque nœud feuille représente une étiquette de classe. L'arbre de

⁵⁷ Nous utilisons l'implémentation passive-agressive du *progiciel Scikit-learn* en langage Python. Par défaut, la fonction objectif de type *hinge* s'inspire de la méthode nommée *PA-I* dans Crammer et coll. (2006).

décision est construit de manière récursive en sélectionnant la fonctionnalité qui fournit le plus de gain d'informations, qui mesure la réduction de l'entropie ou de l'impureté de l'ensemble de données qui résulte de la division des données sur cette fonctionnalité. Pendant le processus de formation, l'arbre de décision est construit en partitionnant de manière récursive les données d'entrée en sous-ensembles plus petits en fonction des caractéristiques sélectionnées. L'arbre est agrandi jusqu'à ce qu'un critère d'arrêt soit satisfait, comme atteindre une profondeur maximale ou avoir un nombre minimal d'échantillons à chaque nœud feuille. Une fois l'arbre construit, il peut être utilisé pour classer de nouvelles instances en parcourant l'arbre du nœud racine à un nœud feuille en fonction des valeurs des caractéristiques de l'instance d'entrée. Les arbres de décision peuvent gérer à la fois des données catégorielles et numériques, et ils peuvent gérer des valeurs manquantes dans les données d'entrée.⁵⁸

8.2.2.4.2 Les forêts aléatoires

Les forêts aléatoires sont un type d'algorithme d'apprentissage supervisé introduit par Breiman (2001a). Dans cet algorithme, plusieurs arbres de décision sont formés sur différents sous-ensembles aléatoires des données d'entraînement, et leurs résultats sont combinés pour produire la sortie finale. L'algorithme de forêts aléatoires fonctionne comme suit :

- i. Des sous-ensembles aléatoires de données d'apprentissage sont sélectionnés.
- ii. Pour chaque sous-ensemble des données d'apprentissage, un arbre de décision est formé à l'aide d'un sous-ensemble aléatoire de caractéristiques.
- iii. Lorsqu'une nouvelle instance doit être classée, chaque arbre de décision de la forêt classe l'instance, et la sortie de la forêt est déterminée par un vote majoritaire ou une moyenne.

⁵⁸ Nous utilisons l'implémentation d'arbres de décision du *progiciel Scikit-learn* en langage Python. Par défaut, la mesure utilisée pour générer l'arborescence de classification est l'impureté de Gini. Cette mesure indique la probabilité de mal classer une observation en calculant la somme des carrés des proportions des nombres de points de données dans toutes les catégories cibles par nœud, soustraite de 1 (Gini, 1921).

Comme les arbres de décision, les forêts aléatoires peuvent gérer à la fois des données catégorielles et numériques. Cependant, les forêts aléatoires peuvent être difficiles à interpréter, car la relation entre les éléments en entrée et ceux en sortie est répartie sur plusieurs arbres.⁵⁹

8.2.2.4.3 Les classifieurs d'arbre supplémentaire

Introduits par Geurts, Ernst et Wehenkel (2006), les classifieurs d'arbres supplémentaires, également connus sous le nom d'arbres extrêmement aléatoires, sont un type d'algorithme d'apprentissage supervisé basé sur les forêts aléatoires. Ce type d'algorithme implémente un estimateur qui ajuste un ensemble d'arbres de décision (ou arbres supplémentaires) sur divers sous-échantillons de l'ensemble de données. La différence entre les forêts aléatoires et les arbres supplémentaires est que les premiers sélectionnent de manière aléatoire un sous-ensemble de données à chaque nœud de l'arbre, tandis que les seconds sélectionnent des fractionnements arborescents aléatoires pour chacune des données. Autrement dit, le classifieur d'arbre supplémentaire applique des divisions au hasard, puis sélectionne automatiquement celle qui entraîne la plus faible impureté des nœuds enfant résultants. Par conséquent, le classifieur d'arbre supplémentaire tend à réduire la variance du modèle par rapport aux forêts aléatoires.⁶⁰

8.2.2.4.4 L'amplification de gradient

La classification par amplification de gradient (*Gradient Boosting*) est un algorithme d'apprentissage automatique proposé par Friedman (2001). Il s'agit d'une méthode d'ensemble qui combine plusieurs modèles, généralement des arbres de décision, pour former un modèle prédictif plus précis. L'idée générale derrière l'amplification de gradient est de former séquentiellement de nouveaux modèles qui prédisent les erreurs des modèles précédents, puis de les ajouter à l'ensemble pour réduire l'erreur globale. Le modèle final est obtenu en combinant les prédictions de tous les modèles de l'ensemble. La descente de gradient est utilisée pour minimiser l'erreur de chaque nouveau modèle. Pendant l'apprentissage, l'algorithme calcule le gradient de la fonction objectif par rapport aux valeurs prédites, puis met à jour les paramètres du modèle dans le sens du gradient. Ce processus est répété pour chaque nouveau modèle, résultant en une séquence de modèles qui améliorent progressivement la précision de l'ensemble final.

⁵⁹ Nous utilisons l'implémentation des forêts aléatoires du *progiciel Scikit-learn* en langage Python. Par défaut, le nombre d'arbres est de 100 et la mesure utilisée pour générer l'arborescence de classification est l'impureté de Gini.

⁶⁰ Nous utilisons l'implémentation des arbres supplémentaires du *progiciel Scikit-learn* en langage Python. Par défaut, le nombre d'arbres est de 100 et la mesure utilisée pour générer l'arborescence de classification est l'impureté de Gini.

L'algorithme sous-jacent est régulé par plusieurs hyperparamètres, tels que le taux d'apprentissage, le nombre d'arbres et la profondeur maximale de chaque arbre. Le taux d'apprentissage contrôle la taille du pas effectué dans la direction du gradient, et un taux d'apprentissage plus faible conduit généralement à de meilleures performances, mais à une convergence plus lente. Le nombre d'arbres et la profondeur maximale de chaque arbre contrôlent la complexité du modèle et peuvent aider à éviter le surajustement.⁶¹

8.2.2.5 Les modèles de classification neuronale

La classification neuronale est un type d'apprentissage automatique inspiré de la structure et de la fonction du cerveau humain. Les réseaux de neurones artificiels sont constitués de couches de neurones artificiels interconnectés qui effectuent des calculs sur les données en entrée. Le neurone artificiel, ou nœud, est l'unité de base des réseaux de neurones artificiels. Un neurone reçoit des entrées, effectue un calcul et génère une sortie. Chaque entrée est pondérée par un poids, reflétant son importance relative dans le calcul et la sortie est le résultat d'une fonction d'activation, une sigmoïde, une tangente hyperbolique ou une ReLU (Unité de Rectification Linéaire) par exemple. Ainsi, le neurone artificiel se compose de trois composants, soit la couche d'entrée, les poids et la fonction d'activation. La couche d'entrée reçoit les données d'entrée, qui sont multipliées par les poids, et le résultat est transmis à la fonction d'activation pour générer la sortie. Le principe de la classification neuronale est d'approximer une fonction mettant en correspondance les données d'entrée et les étiquettes de sortie à l'aide d'un processus d'apprentissage ajustant le poids des connexions entre les neurones. Cela se fait en propageant de manière itérative les données d'entrée à travers le réseau et en ajustant les pondérations en fonction de l'erreur entre les étiquettes de sortie prédites et celles réelles. Un cas particulier de la classification neuronale est l'apprentissage profond qui utilise plusieurs couches de neurones interconnectés pour apprendre des représentations hiérarchiques des données d'entrée.

8.2.2.5.1 Le perceptron

Le perceptron, aussi nommé neurone de McCulloch-Pitts ou porte linéaire, est un type d'algorithme de réseau neuronal artificiel utilisé pour l'apprentissage supervisé de classifieurs binaires. Introduit par Rosenblatt (1958), le perceptron est un algorithme simple constitué d'un seul neurone avec une fonction

⁶¹ Nous utilisons l'implémentation de l'amplification de gradient du *progiciel Scikit-learn* en langage Python. Par défaut, le taux d'apprentissage est de 0,1, le nombre d'arbres est de 100 et la profondeur maximale est de 3.

d'activation de type seuil, lequel peut être utilisé pour classer les données d'entrée dans l'une de deux catégories, sur la base d'une combinaison linéaire des caractéristiques d'entrée. Le perceptron fonctionne en prenant un ensemble de caractéristiques d'entrée et en multipliant chaque caractéristique par une valeur de poids correspondante. Ces caractéristiques pondérées sont ensuite additionnées pour produire une valeur scalaire unique, qui est comparée à une valeur seuil. Si la valeur scalaire est supérieure au seuil, le perceptron délivre une classification positive ; sinon, il produit une classification négative. Pendant l'apprentissage, l'algorithme perceptron est présenté avec un ensemble d'instances d'entrée étiquetées, et il ajuste les valeurs de poids en fonction des erreurs commises lors de la classification. Plus précisément, pour chaque instance, le perceptron calcule la classification de sortie en fonction des valeurs de pondération actuelles, puis ajuste les pondérations en fonction de la différence entre la sortie prévue et la sortie réelle au moyen d'une SGD. La différence entre la prédiction et la véritable valeur cible est calculée en utilisant une fonction de coût. L'erreur calculée est ensuite rétropropagée à travers le réseau pour déterminer la contribution de chaque poids à l'erreur totale. Pour chaque poids, la dérivée partielle de la fonction de coût par rapport à ce poids est calculée, donnant le gradient. L'algorithme perceptron répète ce processus pour plusieurs itérations, jusqu'à ce que toutes les instances d'entrée soient correctement classées ou qu'un nombre maximum d'itérations soit atteint. Bien que l'algorithme perceptron soit relativement simple et efficace, il se limite aux données linéairement séparables et peut ne pas converger pour des données qui ne sont pas linéairement séparables. Cependant, des variantes de l'algorithme perceptron, telles que les perceptrons multicouches, peuvent être utilisées pour classer les données non linéaires.⁶²

8.2.2.5.2 Le perceptron multicouche

Introduit par Rosenblatt (1958) et popularisé par Rumelhart et coll. (1985), le perceptron multicouche (*multilayer perceptron* ou *MLP*) est une généralisation du perceptron simple. Il comprend plusieurs couches de neurones, créant ainsi un réseau dense. Les MLP sont formés d'une couche d'entrée, de plusieurs couches cachées et d'une couche de sortie. Il s'agit d'un réseau de neurones à anticipation, ce qui signifie que les informations ne circulent que dans une seule direction, de la couche d'entrée à la couche de sortie, à travers une série de couches cachées. Il est mathématiquement prouvé qu'un *MLP* à trois couches utilisant la fonction d'activation sigmoïdale peut approximer n'importe quelle fonction

⁶² Nous utilisons l'implémentation de perceptron du *progiciel Scikit-learn* en langage Python. Par défaut, le nombre maximal d'itérations avant la convergence du modèle est de 1000 et la stratégie de décision multiclasse est « un contre tous ».

multivariée continue à n'importe quel degré de précision (Cybenko, 1989 ; Funahashi, 1989 ; Hornik *et al.*, 1989). Dans un perceptron multicouche, plusieurs perceptrons sont disposés en couches. La première couche est la couche d'entrée, qui reçoit les données d'entrée. La sortie de chaque perceptron dans la couche d'entrée est connectée à chaque perceptron dans la première couche cachée, et ainsi de suite pour les couches cachées suivantes. La couche finale est la couche de sortie, qui génère la sortie du réseau. Chaque perceptron dans la couche de sortie correspond à une classe ou une valeur spécifique à prédire. La fonction d'activation utilisée dans un *MLP* est généralement une fonction non linéaire, telle que la fonction sigmoïde ou *ReLU*. Bref, pendant l'apprentissage, les poids du *MLP* sont mis à jour à l'aide d'un algorithme de rétropropagation, qui calcule le gradient de la fonction objectif par rapport aux poids. La fonction objectif mesure la différence entre la sortie prédite du *MLP* et la sortie réelle. Les poids sont ajustés dans le sens opposé du gradient pour minimiser la fonction de perte. Un algorithme d'optimisation souvent utilisé dans les *MLP* pour mettre à jour les poids pendant l'entraînement est l'*Adaptive Moment Estimation (Adam)*. *Adam* est un type d'algorithme de descente de gradient stochastique qui utilise à la fois les moments de premier et de second ordre des gradients pour adapter le taux d'apprentissage pour chaque poids. Cela permet à cet algorithme de converger plus rapidement et de manière plus fiable que les algorithmes traditionnels de descente de gradient stochastique, en particulier dans les cas où les gradients sont clairsemés ou bruyants.⁶³

8.2.2.5.3 La classification de séquence BERT

La classification de séquence BERT (de l'anglais : *BERT for sequence classification*) est une variété du modèle BERT fondé par le modèle de base, le modèle large ou toute autre variante entraînée préalablement sur un large jeu de données textuelles. Lors de la classification de séquence BERT, le modèle préentraîné sélectionné est affiné au moyen d'un corpus spécifique constitué de séquences de mots annotées des classes qu'on vise à modéliser. Pour ce faire, chaque séquence est soumise en entrée en adjoignant un sigle spécial qui sera utilisé comme représentation globale de la séquence lors de la classification. Lorsqu'une séquence passe à travers le modèle BERT, un vecteur de sortie associé au sigle spécial est extrait, lequel est par la suite soumis en entrée à une dernière couche dense dont le nombre

⁶³ Le paramétrage par défaut du *MLP* dans *Scikit-learn* est une activation de type *ReLU* et une optimisation des poids de type *Adam*. Par défaut, le nombre de couches cachées est de 1 et le nombre de neurones artificiels des couches cachées est de 100, mais nous avons ajusté ces paramètres à 2 couches cachées de 200 neurones artificiels chacune afin de permettre l'approximation de n'importe quelle fonction et de contenir au moins le 2/3 du nombre minimal de neurones de la couche d'entrée (la couche cachée de *Doc2vec* étant habituellement composée de 300 neurones artificiels).

de dimensions correspond au nombre de classes à prédire. Au moyen d'une fonction objectif appropriée,⁶⁴ le modèle est affiné au moyen du corpus spécifique à la tâche de classification où seuls les poids de la nouvelle couche dense sont initialisés aléatoirement, car les autres poids proviennent du modèle BERT préentraîné. Lors de cet affinage, tous les poids du modèle BERT ainsi que ceux de la nouvelle couche dense sont mis à jour et, une fois le modèle affiné, il est possible de faire des prédictions sur de nouvelles séquences. Ces dernières sont passées à travers le modèle affiné et la classe avec la probabilité la plus élevée est choisie comme prédiction suivant l'application d'une fonction *softmax* à la sortie.

8.3 Les méthodes

La tâche de reconnaissance de l'appartenance de ces composantes argumentatives à une macrostructure IMRAD s'inscrit à l'intérieur de l'analyse d'argument telle que présentée au chapitre 5. En bref, elle implique i) la constitution du corpus, ii) le décodage du corpus, iii) l'analyse du corpus, iv) la représentation des résultats et v) l'interprétation des résultats et la discussion. Pour les étapes iv) et v), nous présenterons des résultats et de courtes discussions en alternance afin de faciliter l'interprétation.

8.3.1 La constitution du corpus

InTeReC est un corpus développé par Marc Bertin et Iana Atanassova (2018) constitué de contextes de citation annotés de la macrostructure IMRAD de laquelle ils proviennent. Pour des raisons pragmatiques identifiées par les auteurs, notons que ce corpus n'est pas composé de tous les types de contextes de citation, mais seulement de contextes simples de citation (ne contenant qu'une seule citation) et qui cite directement (plutôt qu'indirectement ou de manière *perfunctory*). Les contextes de citation du corpus InTeReC proviennent de 90 071 articles issus de 7 journaux académiques publiés en accès libre à la *Public Library Of Science (PLOS)* de 2001 à 2013. Ces articles sont partitionnés en sections identifiables à l'aide de balises XML pour ne garder que celles correspondant à « introduction », « méthode », « résultats » et « discussion ». Ensuite, les sections sont partitionnées en phrases pour ne conserver que celles qui sont des contextes simples et directs de citation. Les contextes directs de citation sont identifiés en vérifiant si le verbe dans les phrases est membre d'un ensemble de verbes généralement utilisés dans des contextes

⁶⁴ Nous utilisons le modèle *BertForSequenceClassification* du progiciel *transformers* développé par Hugging Face, lequel est affiné pour la classification à l'aide d'une fonction d'entropie croisée (*cross-entropy*). La formule de l'entropie croisée est : $L = - \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(p(y_{ij}))$ où C est le nombre de classes, y_{ij} est 1 si l'échantillon i appartient à la classe j , sinon y_{ij} est 0, et $p(y_{ij})$ est la probabilité prédite par le modèle que l'échantillon i appartient à la classe j .

de citation (Bertin *et al.*, 2015)⁶⁵. À terme, 314 023 phrases annotées de leur section IMRAD sont identifiées comme étant des contextes simples et directs de citation, lesquelles constituent le corpus InTeReC. Plus précisément, les phrases sont annotées des étiquettes « I », « M », « R », « D », « MR » et « RD » correspondant respectivement à l'introduction, la méthode, les résultats, la discussion ainsi que certains mélanges tels que méthode et résultats ou résultats et discussion. Afin d'éviter toute ambiguïté, les phrases annotées « MR » ou « RD » sont éliminées du corpus qui, à terme, est constitué de 303 947 phrases annotées IMRAD. Les 10 076 phrases restantes pourront être utilisées ultérieurement afin de tester notre approche sur des cas indéterminés, c.-à-d. qu'il sera possible de déterminer l'appartenance des contextes de citation classés « MR » ou « RD » aux macrostructures IMRAD.

8.3.2 Le décodage du corpus

Le corpus InTeReC est composé de phrases écrites en langage naturel, lesquelles doivent être transformées en une représentation vectorielle afin de permettre les analyses ultérieures impliquant des manipulations computationnelles. Préalablement, une telle représentation nécessite la décomposition des segments textuels (ou DOMIFs) correspondant à des contextes de citation en une suite d'unités lexicales atomiques (ou UNIFs) correspondant à des mots. Une fois ces dernières identifiées, nous déterminons la catégorie grammaticale de chacune ainsi que son lemme⁶⁶. Dans cette étude, plusieurs représentations vectorielles sont utilisées de manière à identifier celles les plus performantes pour la tâche investiguée. Plus précisément, les méthodes de représentation vectorielle BOW, DBOW, BERT et « text-embedding-ada-002 » sont comparées. La représentation classique BOW est pondérée TFIDF et exempt de mots rares (présents dans moins de 50 extraits textuels) et de mots fonctionnels (*stopwords*). Les autres techniques utilisées sont appliquées directement sur le texte. La représentation de type DBOW contient une couche cachée de 300 neurones. Les représentations BOW et DBOW sont produites dans Python à l'aide du progiciel *Gensim* (Řehůřek et Sojka, 2010). Les représentations de type BERT sont produites dans Python à l'aide du progiciel *transformers* provenant de Hugging Face et donnent des vecteurs de dimension 1024 pour la version large et des vecteurs de dimension 768 pour la version de base (Wolf *et al.*, 2020). La représentation « text-embedding-ada-002 » est produite dans Python à l'aide de l'API de

⁶⁵ Par exemple, les verbes *show, use, include, suggest, identify, find, require, associate, involve, lead, perform, follow, obtain, generate, base, determine, contain, calculate, carry, report, observe, express* et *see* sont des exemples de verbe caractérisant la relation de citation.

⁶⁶ Pour cette étape de tokenisation, de lemmatisation et d'annotation de la catégorie grammaticale, nous utilisons Stanza qui est un outil de traitement du langage naturel développé par le Stanford NLP Group (Qi *et al.*, 2020).

OpenAI et donne des vecteurs de dimension 1536. Enfin, nous proposons d'entraîner et d'évaluer les outils de détection des macrostructures des contextes de citation sur différents types de représentation de manière à identifier les plus performantes.

8.3.3 Les méthodes de détection des macrostructures argumentatives

Dans le contexte de cette étude, l'analyse du corpus n'est pas un examen exhaustif des arguments, mais une enquête sur les différentes méthodes permettant l'identification de macrostructures argumentatives IMRAD. Plus précisément, il s'agit de déterminer l'appartenance des contextes de citation du corpus InTeReC aux classes « introduction » (I), « méthode » (M), « résultats » (R) et « discussion » (D).

8.3.4 L'évaluation des résultats

Deux méthodes d'évaluation seront utilisées pour évaluer la détection de l'appartenance des contextes de citation aux macrostructures IMRAD. La première permet l'évaluation des modèles de classification préalablement énoncés, lorsqu'appliqués aux 314 023 contextes de citation annotés de leur appartenance aux macrostructures IMRAD. Ce jeu de données est dit « déterminé » dans la mesure où l'appartenance IMRAD est déterminée par les annotations. Nous évaluerons les différents modèles de représentation et de classification en termes de précision et de rappel, notamment à l'aide de la F-mesure, l'exactitude, la F-mesure macro et la F-mesure pondérée. L'évaluation est effectuée sur un sous-ensemble aléatoire constitué de 20% des données déterminées. La deuxième méthode permet l'évaluation de n'importe lequel des modèles précédemment énoncés, lorsqu'appliqués aux 10 076 contextes de citation qui ne sont pas clairement associés à une macrostructure IMRAD. Ce jeu de données est dit « indéterminé » dans la mesure où l'appartenance IMRAD n'est pas clairement déterminée par les annotations. Cette méthode sera utilisée afin d'évaluer plus en détail un modèle qui apparaîtra particulièrement performant de représentation et de classification, c.-à-d. la classification de séquence BERT. Rappelons que les cas indéterminés sont classés « MR » ou « RD », c.-à-d. qu'ils proviennent de sections textuelles titrées « méthodes et résultats » (MR) ou encore « résultats et discussion » (RD). En principe, les cas indéterminés annotés « MR » appartiennent soit à la macrostructure « méthodes », soit à la macrostructure « résultats ». Ceux annotés « RD » appartiennent, en principe, soit à la macrostructure « résultats », soit à la macrostructure « discussion ». Lorsqu'un modèle prédictif fait une prédiction à propos de ces cas indéterminés, soit la prédiction est possiblement vraie, de manière ambiguë, soit la prédiction est nécessairement fautive. La raison est qu'on ne connaît pas la véracité des prédictions « méthodes » ou « résultats » pour les cas « MR », ou encore les prédictions « résultats » ou « discussion »

pour les cas « RD ». Contrairement aux cas précédents qui sont ambigus, on connaît la fausseté avérée des prédictions « introduction » et « discussion » pour les cas « MR » et « introduction » et « méthodes » pour les cas « RD ». Ainsi, il est possible de calculer la valeur de rappel maximale d'un tel modèle en postulant la véracité des prédictions ambiguës. La valeur de rappel maximale est une évaluation réelle du modèle prédictif dans la mesure où cette valeur permet la réfutation du modèle évalué. En effet, l'éventualité d'une prédiction massive des catégories dont la fausseté est avérée se traduirait par une faible valeur de rappel maximale et, subséquentement, établirait la faillite du modèle prédictif. Par conséquent, nous évaluerons un modèle de classification au moyen de la valeur de rappel maximale sur les cas indéterminés en interprétant les résultats de manière binaire entre les cas possiblement vrais et les cas nécessairement faux.⁶⁷

8.4 Les résultats

L'entraînement des modèles de classification IMRAD est appliqué sur un sous-ensemble aléatoire constitué de 80% des données déterminées du corpus InTeReC., c.-à-d. 243 157 contextes de citation dont 91 056 proviennent de la discussion (D), 71 172 proviennent de l'introduction (I), 53 844 proviennent des méthodes (M) et 27 085 proviennent des résultats (R). Puis, les modèles entraînés sont évalués sur le sous-ensemble aléatoire constitué du 20% restants des données déterminées, c.-à-d. 60 790 contextes de citation dont 22 833 proviennent de la discussion (D), 13 454 proviennent de l'introduction (I), 17 672 proviennent des méthodes (M) et 6 831 proviennent de résultats. Les mesures de validation utilisées sont la F-mesure, l'exactitude, la macro-moyenne et la moyenne pondérée. La F-mesure, également connue sous le nom de mesure F1, est calculée comme la moyenne harmonique de la précision et du rappel. L'exactitude est une mesure simple qui représente la proportion d'instances correctement classées par rapport au nombre total d'instances. La macro-moyenne (F-mesure macro) est une méthode permettant d'agréger les mesures de performance de plusieurs classes. La macro-moyenne donne un poids égal à toutes les classes, quel que soit le nombre d'instances dans chaque classe. La moyenne pondérée (F-mesure pondérée) est une méthode d'agrégation des mesures de performance sur plusieurs classes en tenant compte du nombre d'instances dans chaque classe. On mesure la performance de chaque classe,

⁶⁷ En l'occurrence, une prédiction est nécessairement fautive lorsqu'un cas indéterminé annoté MR est prédit appartenir à la macrostructure « introduction » ou « discussion » ou lorsqu'un cas annoté RD est prédit appartenir à une macrostructure « introduction » ou « méthodes ». Inversement, une prédiction est possiblement vraie lorsqu'un cas indéterminé annoté MR est prédit appartenir à la macrostructure « méthodes » ou « résultats » ou lorsqu'un cas annoté RD est prédit appartenir à une macrostructure « résultats » ou « discussion ».

puis pondère ces mesures par le nombre d'instances dans chaque classe avant de calculer la moyenne sur toutes les classes.

8.4.1 Les caractéristiques importantes de chaque classe IMRAD

Avant de procéder à l'exploration des différentes méthodes de représentation et de classification, nous effectuons l'analyse exploratoire des caractéristiques textuelles les plus susceptibles de permettre l'identification des macrostructures auxquelles les contextes de citation appartiennent. Une manière d'y parvenir est la régression logistique, car la fonction de décision de cette approche est composée de coefficients de caractéristiques (également appelé poids), qui représentent la force et la direction de la relation entre chaque caractéristique textuelle et la classe cible, en l'occurrence, entre chaque mot et une macrostructure IMRAD. Ces coefficients déterminent dans quelle mesure les caractéristiques en entrée influencent la probabilité prédite d'appartenir à une classe particulière. Les coefficients positifs impliquent une relation positive avec la classe positive, tandis que les coefficients négatifs impliquent une relation négative. Des coefficients d'amplitude plus élevée suggèrent des influences plus fortes. Dans le tableau 8.1 suivant, on voit les 10 mots dont l'amplitude du coefficient est plus élevée, par ordre décroissant, pour les macrostructures « discussion », « introduction », « méthodes », et « résultats ».

Tableau 8.1 Mots de coefficient est plus élevée pour les macrostructures IMRAD

Rang	Discussion		Introduction		Méthodes		Résultats	
	Top-mot	Coef	Top-mot	Coef	Top-mot	Coef	Top-mot	Coef
1	interestingly	3,34	example	2,74	described	7,70	figure	7,18
2	indeed	3,22	worldwide	2,61	calculated	6,89	fig	7,10
3	example	3,19	recently	2,57	performed	6,23	table	5,16
4	future	3,07	increasingly	2,37	used	4,83	previously	3,93
5	may	2,95	promising	2,31	determined	4,33	next	3,67
6	however	2,90	currently	2,30	details	4,00	observed	2,70
7	finding	2,83	however	2,22	obtained	3,47	whether	2,69
8	interesting	2,77	instance	2,18	according	3,45	consistent	2,67
9	study	2,76	approaches	2,01	carried	3,42	wanted	2,63
10	seems	2,57	widely	1,99	version	3,34	previous	2,60

Sans surprise, la majorité des mots les plus représentatifs semblent s'accorder au rôle argumentaire des macrostructures IMRAD auxquelles ils appartiennent. Les mots caractéristiques de la discussion indiquent des relations argumentaires cohérentes avec les sous-objectifs d'une discussion, notamment la

monstration de découvertes (*finding*) accompagnée d'élaborations (*exemple*), d'interprétations (*interestingly, indeed*), d'affirmations prudentes (*may, seems*), d'antithèses ou de limitations (*however*) ainsi que des implications futures (*future*). Les mots caractéristiques de l'introduction sont cohérents avec les sous-objectifs de cette macrostructure, entre autres la mise en place d'un arrière-plan théorique (*approaches*) possiblement récent (*recently, currently*), intéressant (*increasingly, promising*) ou grandiose (*worldwide, widely*), laquelle peut s'accompagner d'élaborations (*example, instance*) ou d'antithèses (*however*). Les mots caractéristiques des méthodes indiquent des relations argumentaires cohérentes avec les sous-objectifs de cette macrostructure, notamment l'élaboration (*version, details*) de procédures (*carried, performed*) ou de méthodes (*used*) pour décrire (*described*), calculer (*calculated*), déterminer (*determined*) quelque chose, possiblement certains résultats (*obtained*), et ces méthodes peuvent être attribuées (*according*). Enfin, les mots caractéristiques des résultats montrent des relations argumentaires cohérentes avec les sous-objectifs de cette macrostructure, entre autres l'exposition de quelque chose qui peut être observé (*observed*) ou recherché (*wanted*), qui suit une certaine articulation (*previous, previously, next*), qui peut être visualisé au moyen de figures (*figure, fig*) ou de tableaux (*table*), et qui peut être interprété (*whether*) ou évalué (*consistent*). Outre la macrostructure de méthodes dont 7 des 10 mots les plus représentatifs sont des verbes, la catégorie grammaticale du verbe ne semble pas la plus déterminante pour l'identification de l'appartenance des contextes de citation aux autres macrostructures IMRAD.

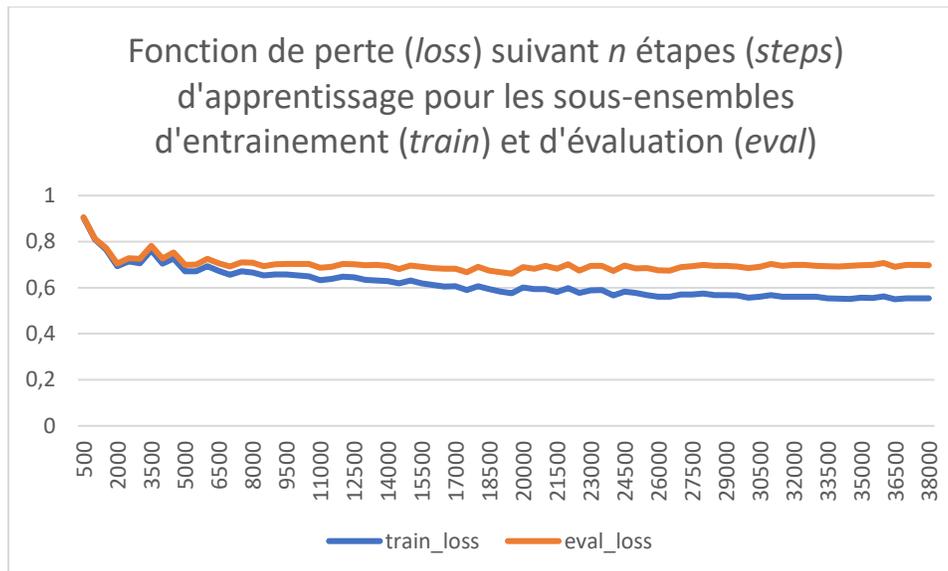
8.4.2 Les résultats du modèle de classification de séquences BERT

Le modèle de classification de séquence BERT remplit un double objectif. Dans notre cas, son entraînement débute avec le modèle BERT de base, lequel est affiné de manière à optimiser la classification de nos données. D'une part, ce modèle permet de générer une représentation textuelle sophistiquée de nos données textuelles. D'autre part, ce modèle permet la classification efficiente de nos données. À la figure 8.1 suivante, on voit les valeurs de la fonction de perte (*loss*) à différentes étapes (*steps*) de l'affinage pour le sous-ensemble d'entraînement et pour le sous-ensemble d'évaluation. Un passage complet du modèle sur toutes les données du sous-ensemble d'apprentissage – nommé une époque (*epoch*)⁶⁸ – correspond à 3800 sous-étapes (*steps*). Le processus d'affinage s'est opéré en 38 000 étapes, soit 10 époques. Une analyse de la fonction de perte permet d'évaluer s'il y a stabilisation ou s'il y a surapprentissage (*overfitting*)

⁶⁸ Lors de l'entraînement d'un modèle d'apprentissage machine, en particulier les réseaux de neurones, les données d'entraînement sont souvent passées plusieurs fois au modèle. Chaque fois que le modèle voit l'intégralité de l'ensemble d'entraînement, on dit qu'une époque est terminée.

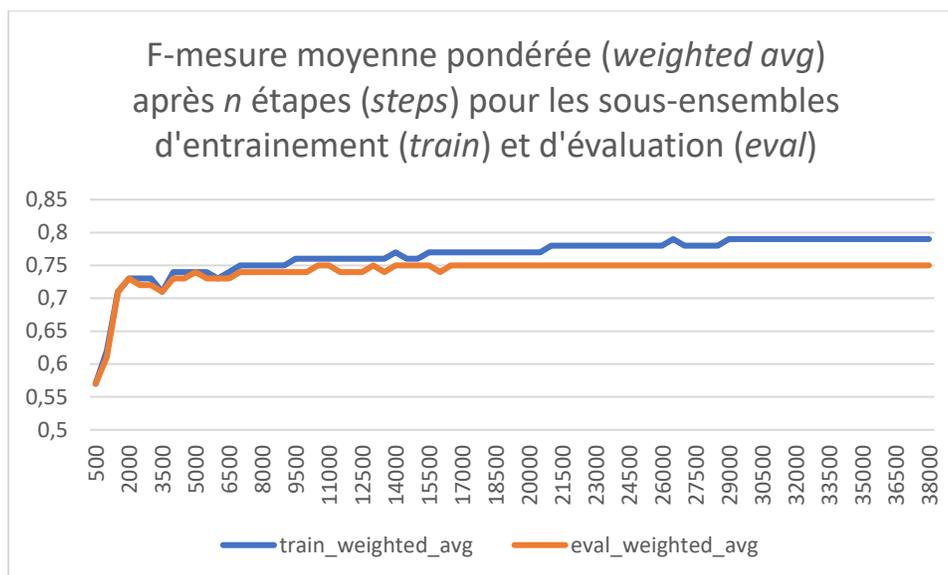
du modèle. La stabilisation du modèle se produit lorsque la fonction de perte atteint un plateau, indiquant que le modèle a appris à partir des données d'entraînement et ne s'améliore plus de manière significative avec des itérations supplémentaires. Le surapprentissage se produit lorsque le modèle apprend trop bien les détails et les bruits spécifiques des données d'entraînement, au détriment de sa capacité à généraliser sur des données non vues. Pour évaluer cela, on observe la fonction de perte non seulement sur l'ensemble d'entraînement, mais aussi sur l'ensemble de validation. Si la perte sur l'ensemble de validation commence à augmenter tandis que la perte sur l'ensemble d'entraînement continue de diminuer, cela peut être un signe de surapprentissage. Il est important de suivre ces deux métriques au fil du temps durant l'entraînement pour identifier le moment où le modèle commence à surapprendre. Dans notre cas, on note que le découplage des valeurs de la fonction de perte entre les données d'apprentissage et les données d'évaluation se réalise tôt lors de l'affinage du modèle, mais ces valeurs ne s'éloignent pas de manière catastrophique entre ces deux sous-ensembles, et les courbes ne s'inversent pas. Le modèle s'améliore considérablement pendant les 2-3 premières époques (7600-11000 étapes) pour ensuite s'améliorer très lentement sur les données d'évaluation. On trouve la valeur de fonction de perte optimale de 0,661 pour les données d'évaluation (0,576 pour les données d'apprentissage) après 19500 étapes (environ 5 époques). Par la suite, les valeurs de la fonction de perte s'améliorent relativement aux données d'apprentissage, mais celles-ci se dégradent très lentement relativement aux données d'évaluation. Cette inversion de la fonction de perte pour les données d'évaluation par rapport à la continuelle amélioration de la fonction de perte pour les données d'apprentissage indique un problème de surapprentissage après environ 5 époques. Par conséquent, le modèle semble se stabiliser rapidement sans amélioration significative après l'étape 19 500 du processus d'affinage.

Figure 8.1 Valeurs de la fonction de perte (*loss*) à différentes étapes (*steps*) de l'affinage pour le sous-ensemble d'entraînement (*train*) et pour le sous-ensemble d'évaluation (*loss*)



Similairement, on voit à la figure 8.2 suivante les valeurs de F-mesure moyenne pondérée (arrondie à 2 décimales) à différentes sous-étapes (*steps*) de l'affinage pour le sous-ensemble d'entraînement et le sous-ensemble d'évaluation. On remarque que le modèle améliore sa performance sur les données d'évaluation lors des 10 500 premières étapes (environ 3 époques) pour éventuellement se stabiliser autour d'une F-mesure moyenne pondérée d'environ 0,75. Parallèlement, la F-mesure moyenne pondérée du sous-ensemble d'apprentissage semble aussi se stabiliser après 29 000 étapes pour une valeur approximative de 0,79. Les valeurs de la F-mesure moyenne pondérée confirment notre analyse précédente selon laquelle le modèle semble se stabiliser rapidement vers la sous-étape 7 500 (F-mesure moyenne pondérée de $\sim 0,74$ sur les données d'évaluation) sans amélioration significative après la sous-étape 16 500 du processus d'affinage (F-mesure moyenne pondérée stable de $\sim 0,75$ sur les données d'évaluation). Puisque le modèle s'améliore sur les données d'apprentissage sans gain significatif sur les données d'évaluation après la sous-étape 21 000, on peut conclure que l'apprentissage subséquent est localisé et n'est pas généralisable. Le temps de calcul pour les 10 époques étant considérable (environ une semaine avec une carte graphique RTX 3090), il est utile de savoir que le modèle converge rapidement après 7500 sous-étapes et qu'il tend au surapprentissage après 21 000 sous-étapes. Le modèle qui sera utilisé subséquentement est celui optimal suggéré par l'analyse de la fonction de perte en accord avec l'analyse de la F-mesure moyenne pondérée, c.-à-d. le modèle à la sous-étape 19 500.

Figure 8.2 Valeurs de F-mesure moyenne pondérée (*weighted_avg*) à différentes sous-étapes de l’affinage pour le sous-ensemble d’entraînement (*train*) et le sous-ensemble d’évaluation (*loss*)



Le tableau 8.2 suivant montre les valeurs de précision, de rappel, de F-mesure (F1) et le support de chaque classe IMRAD du modèle optimal sur les données d’évaluation. De plus, on voit la valeur d’exactitude ainsi que les valeurs de macro moyenne et de moyenne pondérée pour la précision, le rappel et la F-mesure.

Tableau 8.2 Valeurs de précision, de rappel, de F-mesure (F1) et de support de chaque classe IMRAD du modèle BERT optimal sur les données d’évaluation

	Précision	Rappel	F1	Support
Discussion	0,749485	0,748566	0,749025	22833
Introduction	0,647823	0,752118	0,696086	13454
Méthodes	0,88266	0,899842	0,891168	17672
Résultats	0,606576	0,386181	0,471914	6831
Exactitude			0,752607	60790
Macro moyenne	0,721636	0,696677	0,702048	60790
Moyenne pondérée	0,749641	0,752607	0,747491	60790

On observe que la classe la moins performante – la classe « résultats » – est celle dont le support est le plus faible. *A contrario*, la classe la plus performante – la classe « méthodes » – n’est pas celle dont le support est le plus élevé. Cela peut s’expliquer du fait que les énoncés méthodologiques, contrairement à ceux provenant des autres macrostructures, ne portent pas sur l’objet de recherche, mais sur la manière de l’investiguer. Néanmoins, le fait que la classe la plus performante n’est pas celle contenant le plus grand

nombre d'exemplaires montre que l'étendue du support ne suffit pas à expliquer la performance. Au tableau 8.3 suivant, on voit les mesures de performance du modèle optimal sur les données d'apprentissage.

Tableau 8.3 Valeurs de précision, de rappel, de F-mesure (F1) et de support de chaque classe IMRAD du modèle BERT optimal sur les données d'apprentissage

	Précision	Rappel	F1	Support
Discussion	0,773872	0,772283	0,773077	91056
Introduction	0,678647	0,78239	0,726835	53844
Méthodes	0,900696	0,916161	0,908363	71172
Résultats	0,66126	0,435038	0,524808	27085
Exactitude			0,779069	243157
Macro moyenne	0,753619	0,726468	0,733271	243157
Moyenne pondérée	0,777363	0,779069	0,774781	243157

La différence entre la F-mesure des classes IMRAD pour les données d'apprentissage et la F-mesure des classes IMRAD pour les données d'évaluation est de 0,024052 pour « discussion »; 0,030749 pour « introduction »; 0,017195 pour « méthodes » et 0,052894 pour « résultats ». L'écart plus prononcé entre la F-mesure de la classe « résultats » pour les données d'apprentissage et celle pour les données d'évaluation peut s'expliquer par un support plus faible qui semble insuffisant pour une modélisation optimale de cette classe.

8.4.3 Les résultats des modèles de classification basée sur la catégorisation

Les deux (2) modèles de classification basée sur la catégorisation qui ont été testés sur le corpus InTeRec sont les K-moyennes (noté « KMeans ») et le mélange gaussien (noté « Gaussian Mixture »). Ces deux modèles sont appliqués à diverses représentations vectorielles des contextes de citation, c.-à-d. BOW (noté « BOW »), DBOW (noté « DBOW »), OpenAI (noté « OAI ») et BERT dont nous avons testé cinq (5) variantes, soit BERT de base (noté « BERT_base »), BERT large (noté « BERT_large »), BERT affiné sur d'autres tâches ou *corpora*, notamment BERT large affiné sur une tâche de NLI et STS (noté « BERT_nli_stsb ») ainsi que BERT de base affiné sur plusieurs paires (noté « BERT_mpnet ») et, enfin, BERT de base affiné sur la tâche de détection de l'appartenance des contextes de citation aux macrostructures

IMRAD au moyen de corpus InTeReC (noté « BERT_imrad »).⁶⁹ Au tableau 8.4 suivant, on voit la performance de ces techniques de classification appliquées aux diverses représentations vectorielles exprimée en termes de F-mesure pour chacune des classes (notées « F1 (D) » pour discussion, « F1 (I) » pour introduction, « F1 (M) » pour méthodes et « F1 (R) » pour résultats), d'exactitude (notée « Exactitude »), de F-mesure moyenne (notée « F1 (moy.) ») et de F-mesure pondérée (notée « F1 (pond.) »).⁷⁰

Tableau 8.4 Performance des modèles de classification basée sur la catégorisation appliqués aux diverses représentations vectorielles exprimée en termes de F-mesure pour chacune des classes (notées « F1 (D) » pour discussion, « F1 (I) » pour introduction, « F1 (M) » pour méthodes et « F1 (R) » pour résultats), d'exactitude (notée « Exactitude »), de F-mesure moyenne (notée « F1 (moy.) ») et de F-mesure pondérée (notée « F1 (pond.) »)

Modèle		F1 (D)	F1 (I)	F1 (M)	F1 (R)	Exactitude	F1 (moy.)	F1 (pond.)
Kmeans	BOW	0,435	0,167	0,478	0,015	0,37914	0,27358	0,34077
	DBOW	0,341	0,239	0,300	0,118	0,27868	0,24944	0,28140
	BERT_base	0,443	0,353	0,574	0,208	0,41448	0,39463	0,43498
	BERT_large	0,444	0,333	0,535	0,109	0,38054	0,35547	0,40846
	BERT_nli_stsb	0,433	0,360	0,543	0,193	0,39870	0,38235	0,42204
	BERT_mpnet	0,458	0,240	0,402	0,063	0,34149	0,29066	0,34903
	OAI	0,481	0,302	0,561	0,094	0,39349	0,35932	0,42091
	BERT_imrad	0,689	0,684	0,798	0,382	0,66583	0,63808	0,68483
Gaussian Mixture	BOW	0,410	0,245	0,492	0,251	0,35690	0,34958	0,37949
	DBOW	0,382	0,264	0,247	0,162	0,28745	0,26376	0,29185
	BERT_base	0,534	0,343	0,506	0,188	0,42765	0,39274	0,44470
	BERT_large	0,482	0,320	0,623	0,092	0,41757	0,37919	0,44335
	BERT_nli_stsb	0,476	0,308	0,513	0,186	0,39506	0,37061	0,41673
	BERT_mpnet	0,467	0,220	0,384	0,216	0,35085	0,32169	0,35988
	OAI	0,499	0,293	0,589	0,103	0,40610	0,37090	0,43499
	BERT_imrad	0,695	0,672	0,549	0,319	0,57725	0,55887	0,60532

Lors de l'évaluation des modèles, nous nous concentrerons sur la moyenne pondérée de la moyenne harmonique des mesures de précision et de rappel de chaque classe – la F-mesure pondérée – car cette mesure permet de synthétiser la performance d'un modèle sur chacune des classes en tenant compte des

⁶⁹ Spécifions que ces modèles de représentation seront notés de la même manière lorsqu'ils seront utilisés avec les autres modèles de classification.

⁷⁰ La précision et le rappel ne sont pas présentés dans ce tableau par souci d'économie de l'espace, puisque la F-mesure exprime la moyenne harmonique de ces mesures.

différentes valeurs de support. Par rapport au modèle de classification de séquence BERT, les modèles de classification basée sur la catégorisation offrent des performances suboptimales. Parmi ces derniers, le modèle le moins performant est celui des K-moyennes associé à une représentation DBOW pour une F-mesure pondérée de $\sim 0,2814$. Le modèle le plus performant est celui des K-moyennes associé à une représentation BERT_imrad pour une F-mesure pondérée de $\sim 0,68483$. Cette performance suboptimale, mais relativement élevée, indique qu'il est possible d'identifier l'appartenance des contextes de citation aux macrostructures IMRAD au moyen d'un apprentissage non supervisé. Le deuxième modèle le plus performant est celui de mélange gaussien associé à une représentation BERT_imrad pour une F-mesure pondérée de $\sim 0,60532$. Le troisième modèle le plus performant est celui de mélange gaussien associé à une représentation BERT_base pour une F-mesure pondérée de $\sim 0,4447$. Le modèle de classification le plus performant parmi ceux testés est celui des K-moyennes. Or, quand on compare les modèles de classification en fonction des techniques de représentation, on observe que le mélange gaussien est légèrement plus performant pour 6 types de représentation sur 8. Rappelons que les K-moyennes postulent que les données d'une classe s'organisent autour d'un centroïde tandis que le mélange gaussien postule que les données d'une classe s'organisent conformément à une distribution gaussienne. Subséquemment, on peut conclure que, parmi les techniques de représentation testées, celle la plus performante est BERT_imrad dans laquelle les données de chaque classe s'organisent, dans une certaine mesure, autour de centroïdes plutôt qu'une distribution gaussienne, ce qui est l'inverse pour les autres types de représentation à l'exception de BERT_nli_stsb. D'ailleurs, ce qu'ont en commun les représentations BERT_imrad, BERT_nli_stsb et BERT_mpnet est le processus d'affinage qui implique un apprentissage supervisé sur des données annotées, ce qui modifie significativement la représentation initiale comme le montrent nos résultats. Pour sa part, la représentation DBOW apparaît être la technique de représentation la moins performante parmi celles testées. Il est probable que cette performance sous-optimale ne soit pas propre à la technique elle-même, mais au paramétrage particulier que nous avons utilisé. Le cas échéant, la difficulté d'identifier un paramétrage optimal pourrait constituer une limite importante de cette approche. Enfin, étant donné que la représentation la plus performante – BERT_imrad – est la seule qui offre des performances intéressantes, mais qu'elle implique un apprentissage supervisé préalable sur nos données, alors on ne peut pas recommander la classification basée sur la catégorisation pour identifier de manière non supervisée l'appartenance des contextes de citation aux structures IMRAD.

8.4.4 Les résultats des modèles de classification instanciée

Les trois (3) modèles de classification instanciée qui ont été testés sur le corpus InTeRec sont les k plus proches voisins (noté « K Neighbors Classifier »), les voisins de rayon (noté « Radius Neighbors Classifier ») et le plus proche centroïde (noté « Nearest Centroid »). Au tableau 8.5 suivant, on voit la performance de ces techniques de classification instanciée appliquées aux mêmes représentations vectorielles précédemment décrites.

Tableau 8.5 Performance des modèles de classification instanciée appliquées aux diverses représentations vectorielles exprimée en termes de F-mesure pour chacune des classes (notées « F1 (D) » pour discussion, « F1 (I) » pour introduction, « F1 (M) » pour méthodes et « F1 (R) » pour résultats), d'exactitude (notée « Exactitude »), de F-mesure moyenne (notée « F1 (moy.) ») et de F-mesure pondérée (notée « F1 (pond.) »)

Modèle		F1 (D)	F1 (I)	F1 (M)	F1 (R)	Exactitude	F1 (moy.)	F1 (pond.)
K Neighbors Classifier	BOW	0,563	0,314	0,712	0,149	0,53427	0,43440	0,50462
	DBOW	0,471	0,258	0,448	0,056	0,39194	0,30830	0,37070
	BERT_base	0,658	0,392	0,846	0,304	0,63463	0,55024	0,61430
	BERT_large	0,655	0,389	0,845	0,300	0,63130	0,54735	0,61148
	BERT_nli_stsb	0,630	0,415	0,795	0,152	0,60189	0,49780	0,57644
	BERT_mpnet	0,585	0,355	0,750	0,153	0,55544	0,46066	0,53349
	OAI	0,629	0,409	0,827	0,202	0,61163	0,51705	0,59015
	BERT_imrad	0,736	0,659	0,883	0,432	0,73420	0,67732	0,72739
Radius Neighbors Classifier	BOW	0,136	0,100	0,550	0,073	0,16513	0,17159	0,24102
	DBOW	0,557	0,000	0,188	0,002	0,40094	0,18666	0,26401
	BERT_base	0,548	0,002	0,027	0,003	0,37970	0,14482	0,21426
	BERT_large	0,548	0,001	0,029	0,003	0,37995	0,14518	0,21475
	BERT_nli_stsb	0,546	0,000	0,004	0,001	0,37615	0,13782	0,20646
	BERT_mpnet	0,640	0,139	0,733	0,024	0,57014	0,38392	0,48682
	OAI	0,546	0,000	0,000	0,000	0,37560	0,13652	0,20512
	BERT_imrad	0,614	0,375	0,612	0,117	0,54774	0,42975	0,50500
Nearest Centroid	BOW	0,517	0,457	0,755	0,372	0,54177	0,52513	0,55649
	DBOW	0,432	0,315	0,422	0,211	0,36995	0,34480	0,37826
	BERT_base	0,445	0,443	0,708	0,324	0,50615	0,48004	0,50744
	BERT_large	0,457	0,451	0,749	0,330	0,52217	0,49675	0,52635
	BERT_nli_stsb	0,522	0,430	0,753	0,291	0,53524	0,49926	0,54305
	BERT_mpnet	0,438	0,321	0,649	0,294	0,44650	0,42518	0,45693
	OAI	0,514	0,465	0,777	0,346	0,54851	0,52567	0,56087
	BERT_imrad	0,717	0,693	0,877	0,483	0,72497	0,69258	0,73186

Les modèles de classification instanciée testés offrent des performances inférieures au modèle de classification de séquence BERT, mais les modèles les plus performants ne sont pas très loin derrière. Le modèle le moins performant est celui des voisins de rayon associé à une représentation OAI pour une F-mesure pondérée de $\sim 0,20512$. Le modèle le plus performant est celui du plus proche centroïde associé à une représentation BERT_imrad pour une F-mesure pondérée de $\sim 0,73186$. Notons que la représentation DBOW est la moins performante pour deux (2) des trois (3) modèles de classification testés. Pour tous les modèles testés, la représentation BERT_imrad est largement plus performante que les autres représentations. Toutes représentations confondues, le modèle de classification des voisins de rayon est le moins performant, ce qui indique que les exemplaires dans un rayon fixe autour d'un point de donnée ne permettent pas d'identifier adéquatement l'appartenance du point à une classe. Le troisième modèle le plus performant, qui n'est pas associé à une représentation BERT_imrad, est celui des k plus proches voisins associé à une représentation BERT_base pour une F-mesure pondérée de $\sim 0,6143$. De plus, on observe que cette méthode est plus performante pour cinq (5) types de représentation sur les huit (8) testées. La performance des plus proches voisins, comparativement à celles des voisins de rayon, semble indiquer qu'il y a des exemplaires voisins permettant d'identifier l'appartenance d'un point à une classe, mais que ceux-ci ne sont pas regroupés dans un rayon fixe autour de ce point. Pour les représentations BERT_base, BERT_large, BERT_nli_stsb, BERT_mpnet et OAI, les k plus proches voisins contiennent plus d'informations à propos de l'appartenance d'un point à une classe que la moyenne des points d'une classe donnée. Néanmoins, cette observation ne s'applique pas à la représentation BERT_imrad puisque l'approche du plus proche centroïde est plus performante pour ce type de représentation (ce qui est consistant avec l'observation qu'une approche basée sur les centroïdes couplée à une représentation BERT_imrad est plus performante pour les modèles de classification fondée sur la catégorisation). Cependant, notons que les approches des plus proches voisins et du plus proche centroïde, couplées à la représentation BERT_imrad, sont les seules à offrir des performances intéressantes et celles-ci sont relativement similaires. Subséquemment, on peut inférer que les informations locales et les informations globales, dans la représentation BERT_imrad, permettent toutes deux l'identification de l'appartenance d'un point à une classe. Or, puisque la représentation BERT_imrad nécessite l'entraînement du modèle de classification de séquence BERT sur nos données, et que la performance de ce modèle est supérieure, nous ne pouvons pas recommander l'usage des modèles d'instanciation testés pour l'identification de l'appartenance des contextes de citation aux macrostructures IMRAD.

8.4.5 Les résultats des modèles de classification linéaire

Les huit (8) modèles de classification linéaire qui ont été testés sur le corpus InTeRec sont deux (2) variantes SVM, soit le noyau linéaire (notés « SVC_linear ») et celui répondant à une fonction de base radiale (notés « SVC_RBF »), la régression logistique (notée « Logistic Regression »), la régression logistique avec validation croisée (notée « Logistic Regression CV »), la régression Ridge (notée « Ridge Classifier »), la régression Ridge avec validation croisée (notée « Ridge Classifier CV »), la descente de gradient stochastique (notée « SGD Classifier ») et la classification passive-agressive (notée « Passive Agressive Classifier »). Au tableau 8.6 suivant, on voit la performance de ces techniques de classification linéaire appliquées aux représentations vectorielles précédemment décrites.

Tableau 8.6 Performance des modèles de classification linéaire appliquées aux diverses représentations vectorielles exprimée en termes de F-mesure pour chacune des classes (notées « F1 (D) » pour discussion, « F1 (I) » pour introduction, « F1 (M) » pour méthodes et « F1 (R) » pour résultats), d'exactitude (notée « Exactitude »), de F-mesure moyenne (notée « F1 (moy.) ») et de F-mesure pondérée (notée « F1 (pond.) »)

Modèle		F1 (D)	F1 (I)	F1 (M)	F1 (R)	Exactitude	F1 (moy.)	F1 (pond.)
SVC Linear	BOW	0,675	0,417	0,845	0,328	0,65178	0,56621	0,62830
	DBOW	0,577	0,000	0,493	0,000	0,45651	0,26748	0,35998
	BERT_base	0,730	0,606	0,876	0,356	0,71837	0,64207	0,70295
	BERT_large	0,732	0,606	0,880	0,375	0,72153	0,64823	0,70708
	BERT_nli_stsb	0,713	0,589	0,855	0,247	0,69632	0,60101	0,67456
	BERT_mpnet	0,683	0,369	0,835	0,054	0,64129	0,48517	0,58680
	OAI	0,743	0,615	0,884	0,372	0,73050	0,65355	0,71412
	BERT_imrad	0,759	0,698	0,894	0,455	0,75846	0,70129	0,75038
SVC RBF	BOW	0,684	0,414	0,851	0,331	0,65989	0,57009	0,63325
	DBOW	0,587	0,077	0,551	0,013	0,48189	0,30705	0,39931
	BERT_base	0,739	0,620	0,882	0,353	0,72777	0,64865	0,71101
	BERT_large	0,738	0,600	0,885	0,357	0,72563	0,64524	0,70764
	BERT_nli_stsb	0,725	0,612	0,865	0,260	0,71017	0,61568	0,68861
	BERT_mpnet	0,693	0,427	0,856	0,212	0,66350	0,54694	0,62746
	OAI	0,747	0,627	0,891	0,390	0,73719	0,66362	0,72209
	BERT_imrad	0,759	0,696	0,894	0,456	0,75874	0,70137	0,75040
Logistic Regression	BOW	0,671	0,438	0,844	0,360	0,65243	0,57801	0,63454
	DBOW	0,574	0,161	0,508	0,045	0,46529	0,32187	0,40378
	BERT_base	0,729	0,606	0,873	0,389	0,71717	0,64913	0,70519
	BERT_large	0,732	0,605	0,877	0,403	0,72081	0,65424	0,70912
	BERT_nli_stsb	0,710	0,585	0,852	0,304	0,69428	0,61297	0,67820
	BERT_mpnet	0,678	0,424	0,832	0,214	0,64548	0,53664	0,61408
	OAI	0,741	0,615	0,883	0,402	0,72922	0,66015	0,71627

	BERT_imrad	0,760	0,697	0,894	0,477	0,75909	0,70683	0,75304
Logistic Regression CV	BOW	0,677	0,429	0,841	0,342	0,65422	0,57240	0,63233
	DBOW	0,573	0,185	0,510	0,049	0,46664	0,32946	0,41011
	BERT_base	0,729	0,605	0,873	0,387	0,71703	0,64852	0,70487
	BERT_large	0,732	0,606	0,877	0,404	0,72089	0,65480	0,70939
	BERT_nli_stsb	0,710	0,585	0,852	0,300	0,69414	0,61179	0,67761
	BERT_mpnet	0,677	0,433	0,834	0,236	0,64696	0,54489	0,61906
	OAI	0,741	0,615	0,883	0,402	0,72922	0,66015	0,71627
	BERT_imrad	0,760	0,695	0,894	0,475	0,75886	0,70596	0,75249
Ridge Classifier	BOW	0,676	0,406	0,833	0,294	0,64721	0,55220	0,61891
	DBOW	0,577	0,082	0,512	0,004	0,46369	0,29374	0,38419
	BERT_base	0,726	0,572	0,863	0,225	0,70362	0,59661	0,67556
	BERT_large	0,730	0,569	0,869	0,254	0,70791	0,60561	0,68133
	BERT_nli_stsb	0,711	0,556	0,837	0,111	0,68197	0,55371	0,64590
	BERT_mpnet	0,682	0,381	0,819	0,051	0,63928	0,48321	0,58435
	OAI	0,740	0,588	0,876	0,275	0,71958	0,61974	0,69376
	BERT_imrad	0,760	0,697	0,894	0,472	0,75943	0,70569	0,75261
Ridge Classifier CV	BOW	0,684	0,390	0,833	0,272	0,65088	0,54495	0,61615
	DBOW	0,577	0,082	0,511	0,004	0,46356	0,29362	0,38405
	BERT_base	0,726	0,571	0,863	0,223	0,70347	0,59593	0,67518
	BERT_large	0,730	0,568	0,868	0,250	0,70739	0,60409	0,68039
	BERT_nli_stsb	0,711	0,557	0,837	0,111	0,68212	0,55381	0,64603
	BERT_mpnet	0,682	0,386	0,819	0,053	0,63955	0,48493	0,58560
	OAI	0,741	0,597	0,878	0,295	0,72217	0,62783	0,69888
	BERT_imrad	0,760	0,697	0,893	0,473	0,75937	0,70573	0,75255
SGD Classifier	BOW	0,688	0,162	0,809	0,210	0,62882	0,46739	0,55316
	DBOW	0,526	0,138	0,528	0,012	0,44257	0,30091	0,38284
	BERT_base	0,702	0,624	0,865	0,326	0,70211	0,62926	0,68986
	BERT_large	0,727	0,563	0,873	0,090	0,70067	0,56325	0,66167
	BERT_nli_stsb	0,706	0,560	0,839	0,056	0,67740	0,54018	0,63919
	BERT_mpnet	0,679	0,128	0,797	0,025	0,61193	0,40734	0,51790
	OAI	0,721	0,483	0,858	0,021	0,68337	0,52096	0,62976
	BERT_imrad	0,761	0,696	0,894	0,472	0,75925	0,70551	0,75260
Passive Aggressive Classifier	BOW	0,645	0,399	0,824	0,281	0,61745	0,53703	0,60152
	DBOW	0,458	0,237	0,506	0,008	0,41257	0,30224	0,37251
	BERT_base	0,710	0,504	0,853	0,093	0,67873	0,54008	0,63683
	BERT_large	0,686	0,596	0,870	0,390	0,69038	0,63556	0,68641
	BERT_nli_stsb	0,622	0,475	0,818	0,299	0,61517	0,55343	0,61008
	BERT_mpnet	0,578	0,438	0,809	0,294	0,58279	0,52990	0,58233

	OAI	0,702	0,635	0,870	0,456	0,69995	0,66588	0,70839
	BERT_imrad	0,747	0,688	0,896	0,452	0,74935	0,69561	0,74404

Les modèles de classification linéaire testés offrent des performances parfois inférieures, parfois supérieures au modèle de classification de séquence BERT. Le modèle le moins performant est le modèle SVM à noyau linéaire associé à une représentation DBOW pour une F-mesure pondérée de $\sim 0,35998$. Le modèle le plus performant est celui de la régression logistique associé à une représentation BERT_imrad pour une F-mesure pondérée de $\sim 0,75304$, ce qui est légèrement supérieur au modèle de classification de séquence BERT. Pour toutes les méthodes de classification linéaire testées, la représentation BERT_imrad est la plus performante et la représentation DBOW est la moins performante. À cet effet, tous les modèles de classification linéaire testés et associés à la représentation BERT_imrad font légèrement mieux que le modèle de classification de séquence BERT à l'exception de la classification passive-agressive. Le modèle le plus performant qui n'est pas associé à une représentation BERT_imrad est le SVM avec noyau RBF associé à une représentation OAI pour une F-mesure pondérée de $\sim 0,72209$. D'ailleurs, la représentation OAI est la seconde plus performante après BERT_imrad pour tous les classifieurs linéaires testés à l'exception de la descente de gradient stochastique (chacun avec des performances relativement élevées pour une moyenne de F-mesure pondérée de $\sim 0,71$). La troisième représentation qui est la plus performante est BERT_large pour tous les classifieurs linéaires testés à l'exception du SVM avec noyau RBF (chacun avec des performances moins intéressantes pour une moyenne de F-mesure pondérée de $\sim 0,69$). Ce dernier est le modèle le plus performant pour la moitié des types de représentation, et la régression logistique (simple ou avec validation croisée) est plus performante pour l'autre moitié des types de représentation. Cette observation indique une relation linéaire entre le logarithme des chances d'appartenir aux classes IMRAD et les points de donnée dans l'espace de représentation original pour la moitié des types de représentation testés (BOW, DBOW, BERT_large et BERT_imrad), et qu'une redescription des données dans un espace de plus grande dimension offre un léger gain de performance pour l'autre moitié des types de représentation (BERT_base, BERT_nli_stsb, BERT_mpnet et OAI). Étonnamment, tous les modèles de classification linéaires testés, à l'exception de la classification passive-agressive, font légèrement mieux que le modèle de classification de séquence BERT lorsqu'associé à la représentation BERT_imrad. Subséquemment, on peut inférer que le modèle de classification de séquence BERT, qui apparaît être un excellent moyen d'affiner la représentation BERT, ne produit pas nécessairement le classifieur le plus performant pour la tâche qui nous intéresse. Cependant, la différence est marginale entre la performance du modèle de classification de séquence BERT et celle des classifieurs

linéaires associés à la représentation BERT_imrad. De plus, la performance du SVM avec noyau RBF associé à la représentation OAI est particulièrement intéressante, car la représentation OAI, qui ne provient pas d'un long processus d'affinage, s'obtient très rapidement et le SVM avec noyau RBF peut être entraîné en quelques heures, ce qui est beaucoup plus rapide que la semaine complète nécessaire à l'entraînement du modèle de classification de séquence BERT. Spécifions qu'il est possible d'affiner la représentation OAI similairement au modèle de classification de séquence BERT, notamment à l'aide de l'API d'OpenAI (l'algorithme étant propriétaire et payant). Le cas échéant, on peut s'attendre à des performances encore meilleures que celles obtenues au moyen du modèle de classification de séquence BERT (mais qui lui a l'avantage d'être libre d'accès). Enfin, notons que le SVM avec noyau RBF associé à la représentation BERT de base, sans aucun affinage, offre une performance légèrement inférieure au SVM-RBF-OAI, mais comparable, aussi rapidement et en libre accès.

8.4.6 Les résultats des modèles de classification arborescente

Les quatre (4) modèles de classification arborescente qui ont été testés sur le corpus InTeRec sont les arbres de décision (notés « Decision Tree Classifier »), les forêts aléatoires (notées « Random Forest Classifier »), les classifieurs d'arbre supplémentaire (notés « Extra Tree Classifier ») et l'amplification de gradient (notée « Gradient Boosting Classifier »). Au tableau 8.7 suivant, on voit la performance de ces techniques de classification arborescente appliquées aux représentations vectorielles précédemment décrites.

Tableau 8.7 Performance des modèles de classification arborescente appliquées aux diverses représentations vectorielles exprimée en termes de F-mesure pour chacune des classes (notées « F1 (D) » pour discussion, « F1 (I) » pour introduction, « F1 (M) » pour méthodes et « F1 (R) » pour résultats), d'exactitude (notée « Exactitude »), de F-mesure moyenne (notée « F1 (moy.) ») et de F-mesure pondérée (notée « F1 (pond.) »)

Modèle		F1 (D)	F1 (I)	F1 (M)	F1 (R)	Exactitude	F1 (moy.)	F1 (pond.)
Decision Tree Classifier	BOW	0,560	0,353	0,746	0,258	0,53948	0,47932	0,53441
	DBOW	0,411	0,252	0,379	0,139	0,33530	0,29536	0,33598
	BERT_base	0,538	0,366	0,716	0,209	0,51393	0,45722	0,51478
	BERT_large	0,528	0,369	0,725	0,213	0,51267	0,45877	0,51475
	BERT_nli_stsb	0,510	0,328	0,639	0,157	0,46644	0,40859	0,46768
	BERT_mpnet	0,470	0,284	0,582	0,158	0,42568	0,37342	0,42637
	OAI	0,514	0,327	0,680	0,181	0,48153	0,42538	0,48329
	BERT_imrad	0,665	0,583	0,844	0,372	0,66483	0,61599	0,66595
Extra Tree Classifier	BOW	0,558	0,319	0,716	0,202	0,52165	0,44875	0,51107
	DBOW	0,399	0,247	0,362	0,130	0,32377	0,28464	0,32444

	BERT_base	0,506	0,335	0,666	0,184	0,47710	0,42261	0,47842
	BERT_large	0,510	0,339	0,669	0,180	0,47970	0,42445	0,48117
	BERT_nli_stsb	0,475	0,292	0,582	0,145	0,42744	0,37363	0,42870
	BERT_mpnet	0,446	0,269	0,522	0,146	0,39434	0,34588	0,39536
	OAI	0,475	0,289	0,599	0,148	0,43201	0,37769	0,43311
	BERT_imrad	0,672	0,585	0,844	0,370	0,66830	0,61798	0,66904
Random Forest Classifier	BOW	0,680	0,297	0,821	0,240	0,63548	0,50939	0,58664
	DBOW	0,566	0,109	0,515	0,012	0,45938	0,30071	0,38801
	BERT_base	0,684	0,415	0,832	0,133	0,64989	0,51624	0,60586
	BERT_large	0,681	0,393	0,838	0,113	0,64632	0,50630	0,59920
	BERT_nli_stsb	0,669	0,293	0,788	0,002	0,61316	0,43788	0,54518
	BERT_mpnet	0,650	0,137	0,765	0,010	0,58339	0,39029	0,49780
	OAI	0,683	0,250	0,829	0,019	0,63119	0,44551	0,55536
BERT_imrad	0,756	0,688	0,892	0,466	0,75437	0,70054	0,74790	
Gradient Boosting Classifier	BOW	0,655	0,191	0,761	0,252	0,59679	0,46484	0,53789
	DBOW	0,571	0,092	0,502	0,019	0,45929	0,29633	0,38323
	BERT_base	0,693	0,479	0,841	0,276	0,66768	0,57236	0,64195
	BERT_large	0,691	0,466	0,848	0,278	0,66598	0,57058	0,64024
	BERT_nli_stsb	0,676	0,436	0,797	0,113	0,63497	0,50551	0,59487
	BERT_mpnet	0,651	0,259	0,770	0,062	0,59332	0,43552	0,53261
	OAI	0,696	0,449	0,843	0,237	0,66499	0,55657	0,63280
BERT_imrad	0,759	0,696	0,892	0,473	0,75779	0,70491	0,75150	

Les modèles de classification arborescente testés offrent des performances inférieures au modèle de classification de séquence BERT à l'exception de deux (2) modèles. Le modèle le moins performant est celui des arbres supplémentaires associé à une représentation DBOW pour une F-mesure pondérée de $\sim 0,32444$. En fait, DBOW est la représentation la moins performante pour tous les modèles testés de classification arborescente. Le modèle le plus performant est celui de l'amplification de gradient associé à une représentation BERT_imrad pour une F-mesure pondérée de $\sim 0,7515$. Le second modèle le plus performant est celui des forêts aléatoires associé à la représentation BERT_imrad pour une F-mesure pondérée de $\sim 0,7479$. Pour toutes les méthodes de classification arborescente testées, la représentation BERT_imrad est plus performante que les autres. Le modèle le plus performant qui n'est pas associé à une représentation BERT_imrad est l'amplification de gradient associé à une représentation BERT_base pour une F-mesure pondérée de $\sim 0,64195$. Pour tous les modèles testés, l'amplification de gradient est le modèle le plus performant pour toutes les techniques de représentation fondées par l'apprentissage

profond (BERT, ses variantes et OpenAI). Les deux (2) représentations plus simples testées, BOW et DBOW, sont plus performantes avec les forêts aléatoires. Or, il est reconnu que les forêts aléatoires sont beaucoup moins sensibles au bruit et aux données aberrantes comparativement à l'amplification de gradient (Hastie *et al.*, 2009). Subséquemment, nos résultats montrent que nos représentations issues des techniques de représentation fondées par l'apprentissage profond sont relativement moins bruitées ou comportent moins de données aberrantes. Enfin, étant donné que les performances dignes d'intérêt sont obtenues au moyen de la représentation issue du modèle de classification de séquence BERT, que la différence entre ce classifieur et les meilleurs modèles de classification arborescente est marginale, et que cette performance est légèrement inférieure aux meilleurs modèles de classification linéaire, nous ne recommandons pas cette approche.

8.4.7 Les résultats des modèles de classification neuronale

Les trois (3) modèles de classification neuronale qui ont été testés sur le corpus InTeReC sont le perceptron (noté « Perceptron »), le perceptron multicouche (noté « MLP Classifieur ») et le modèle de classification de séquences BERT (noté « BERT for Sequence »). Notons que le modèle de classification de séquence BERT est aussi utilisé pour générer la représentation BERT entraînée sur le corpus InTeReC et, donc, il n'est pas appliqué sur les autres représentations vectorielles comme pour les autres modèles. Au tableau 8.8 suivant, on voit la performance de ces techniques de classification neuronale appliquées aux différentes représentations vectorielles précédemment décrites.

Tableau 8.8 Performance des modèles de classification neuronale appliquées aux diverses représentations vectorielles exprimée en termes de F-mesure pour chacune des classes (notées « F1 (D) » pour discussion, « F1 (I) » pour introduction, « F1 (M) » pour méthodes et « F1 (R) » pour résultats), d'exactitude (notée « Exactitude »), de F-mesure moyenne (notée « F1 (moy.) ») et de F-mesure pondérée (notée « F1 (pond.) »)

Modèle		F1 (D)	F1 (I)	F1 (M)	F1 (R)	Exactitude	F1 (moy.)	F1 (pond.)
Perceptron	BOW	0,573	0,448	0,809	0,345	0,58442	0,54370	0,58824
	DBOW	0,519	0,011	0,369	0,180	0,37506	0,26963	0,32485
	BERT_base	0,321	0,552	0,799	0,387	0,53122	0,51500	0,51867
	BERT_large	0,677	0,574	0,715	0,384	0,61716	0,58754	0,63230
	BERT_nli_stsb	0,637	0,409	0,813	0,273	0,60235	0,53305	0,59679
	BERT_mpnet	0,513	0,454	0,798	0,260	0,55695	0,50630	0,55451
	OAI	0,706	0,627	0,864	0,023	0,69581	0,55507	0,65777
	BERT_imrad	0,738	0,681	0,888	0,445	0,73859	0,68804	0,73613
MLP Classifieur	BOW	0,665	0,000	0,811	0,264	0,58626	0,43483	0,51501
	DBOW	0,548	0,296	0,534	0,086	0,46782	0,36602	0,43621

	BERT_base	0,671	0,523	0,836	0,369	0,65090	0,59969	0,65215
	BERT_large	0,666	0,539	0,841	0,366	0,65695	0,60281	0,65487
	BERT_nli_stsb	0,661	0,528	0,825	0,314	0,64011	0,58184	0,64007
	BERT_mpnet	0,577	0,430	0,793	0,272	0,57185	0,51828	0,57326
	OAI	0,630	0,558	0,862	0,300	0,64062	0,58758	0,64457
	BERT_imrad	0,752	0,694	0,889	0,448	0,75108	0,69560	0,74456
BERT for Sequence	BERT_imrad	0,749	0,696	0,891	0,472	0,75261	0,70205	0,74749

Le modèle de classification de séquence BERT (associé nécessairement à la représentation BERT_imrad) est le plus performant des modèles de classification neuronale testés pour une F-mesure pondérée de $\sim 0,74749$. Le modèle le moins performant est le modèle perceptron associé à une représentation DBOW pour une F-mesure pondérée de $\sim 0,32485$. En fait, DBOW est la représentation la moins performante et BERT_imrad est la représentation la plus performante de tous les modèles de classification neuronale testés. Le modèle le plus performant qui n'est pas associé à une représentation BERT_imrad est le perceptron associé à une représentation OAI pour une F-mesure pondérée de $\sim 0,65777$. Or, le perceptron nécessite une séparabilité linéaire des classes. Subséquemment, on pourrait inférer que les classes IMRAD sont linéairement séparables pour la représentation OAI (ainsi que BOW), ce que semblent indiquer les classifieurs linéaires non neuronaux, bien que la méthode RBF du SVM offre un gain performatif marginal pour la représentation OAI. Pour les autres représentations, le classifieur testé le plus performant (après le modèle de classification de séquence BERT) est le perceptron multicouche qui offre des performances légèrement supérieures au perceptron. Or, contrairement au perceptron, le perceptron multicouche permet des partitions non linéaires. Subséquemment, on pourrait inférer qu'une part significative des points de donnée dans les classes IMRAD ne sont pas linéairement séparables pour les représentations DBOW, BERT_base, BERT_large, BERT_nli_stsb, BERT_mpnet et BERT_imrad. Cette observation est corroborée par certains classifieurs linéaires non neuronaux, notamment lorsqu'on compare le SVM linéaire et le SVM à noyaux RBF. En effet, tous les types de représentation performant mieux avec le SVM à noyaux RBF. Or, ce qui distingue ces deux approches, c'est que la première est opérée sur les points de donnée dans l'espace original et la seconde est appliquée sur une redescription des mêmes points de donnée dans un espace de plus grande dimension, lequel permet le partitionnement linéaire de ce qui ne l'était pas dans l'espace original. Néanmoins, la différence de performance est très faible entre ces deux approches, notamment pour les représentations BOW, BERT_base, BERT_large, OAI et BERT_imrad (moins de 0,01 d'écart pour la F-mesure moyenne pondérée), ce qui montre que la grande majorité des points de

donnée sont linéairement séparables en classes IMRAD dans ces types de représentation. Une raison expliquant pourquoi les SVM linéaire et RBF performant mieux que les perceptrons simple et multicouche pour notre tâche est que les premiers résolvent un problème d'optimisation convexe impliquant un minimum global unique. Le cas échéant, la convexité garantit qu'il n'y a pas de minima local qui pourrait piéger l'algorithme d'optimisation. Le perceptron simple peut ne pas converger vers un minimum global (ce qui semble être notre cas étant donné la faiblesse des résultats) et le perceptron multicouche peut s'enliser dans un minimum local (ce qui semble aussi être notre cas étant donné la faiblesse des résultats). Le fait que tous les modèles de classification linéaire testés offrent des performances élevées indique que l'appartenance des contextes de citation aux macrostructures IMRAD est un problème convexe et qu'une stratégie efficace pourrait consister à représenter les données de manière à ce qu'une résolution soit possible au moyen d'une fonction de perte convexe et/ou d'une limite décisionnelle convexe. Inversement, en raison des problèmes de convergence et de minimum local, nous ne recommandons pas l'usage des perceptrons simple ou multicouche. Pour sa part, le modèle de classification de séquence BERT est particulièrement efficace pour le problème qui nous intéresse. La force de cette approche n'est pas tant sa capacité à classer les données, mais son aptitude à redécrire les points de données dans le même espace d'origine de manière à faciliter la classification subséquente. Puisque le problème de classification est solvable par les modèles de classification linéaire faisant l'assomption d'un problème convexe, et que la performance est optimale suivant la transformation des données pas le modèle de classification de séquence de BERT, force à croire que cette transformation semble converger vers une solution convexe du problème. Or, le modèle de classification de séquence BERT introduit en principe des non-linéarités par ses fonctions d'activation, ses interactions complexes et la profondeur globale de son réseau de sorte que les transformations représentationnelles peuvent converger vers une solution non convexe. Subséquemment, il est possible de vérifier si la transformation représentationnelle des données lors de l'affinage de BERT avec la classification de séquence BERT converge vers une solution convexe du problème, facilitant ainsi sa résolution. D'une part, les modèles de classification linéaire doivent performer mieux avec la représentation issue du modèle de classification de séquence BERT comparativement à la représentation BERT originale. D'autre part, les modèles de classification linéaire doivent performer également ou mieux avec la représentation issue du modèle de classification de séquence BERT que ne le fait le modèle de classification de séquence BERT lui-même.

8.4.8 La détection de l'appartenance des cas indéterminés aux macrostructures IMRAD

Les résultats précédemment présentés ont permis d'évaluer l'application de plusieurs modèles de représentation et de classification sur nos données déterminées. Parmi ceux-ci, certains se sont avérés particulièrement performants, notamment la représentation issue du modèle de classification de séquence BERT ainsi que son classifieur. La question est maintenant de savoir si ce modèle peut être appliqué sur des données indéterminées en espérant une performance similaire. Subséquemment, nous proposons une approche hybride composée de deux (2) étapes. D'abord, un modèle de représentation et de classification est entraîné sur des données déterminées. Ensuite, le modèle entraîné est appliqué sur des données indéterminées. Maintenant, l'objectif est de valider cette approche hybride. Pour ce faire, nous utiliserons le modèle de classification de séquence BERT, puis nous l'appliquerons sur nos données indéterminées afin de déterminer l'appartenance des cas indéterminés aux macrostructures IMRAD. L'application de ce modèle sur nos données indéterminées est possible, mais la performance pourrait être médiocre si le modèle est incapable de généralisation, ou si les données indéterminées étaient caractérisées autrement. En principe, les cas indéterminés annotés « MR » appartiennent soit à la macrostructure « méthodes », soit à la macrostructure « résultats ». Ceux annotés « RD » appartiennent, en principe, soit à la macrostructure « résultats », soit à la macrostructure « discussion ». La valeur de rappel maximale est calculée en divisant le nombre de cas indéterminés attribués à une catégorie possiblement vraie par le nombre de cas total (le nombre de cas indéterminés attribués à une catégorie possiblement vraie additionné du nombre de cas indéterminés attribués à une catégorie nécessairement fausse). Au tableau 8.9 suivant, on voit le résultat de prédiction du modèle sur les 10 076 cas indéterminés.

Tableau 8.9 Résultat de prédiction du modèle BERT_imrad sur les cas indéterminés

	<i>MR</i>	<i>RD</i>
<i>Discussion</i>	8	5927
Introduction	17	1430
Méthodes	38	934
Résultats	8	1714
Total	71	10005

Pour les cas indéterminés annotés « MR », 46 prédictions sont possiblement vraies et 25 sont fausses. Pour les cas indéterminés annotés « RD », 7641 prédictions sont possiblement vraies et 2364 sont fausses. En somme, 7687 prédictions sont possiblement vraies et 2389 sont fausses, ce qui nous donne une valeur de rappel maximale de 0,76. Notons que cette performance est similaire à la performance des meilleurs

modèles appliqués sur les données d'évaluation. Si le modèle fait aussi bien sur les données indéterminées que sur les données d'évaluation, alors on peut s'attendre à des performances comparables, ce qui semble grossièrement être le cas. Généralement, on peut s'attendre à ce que certains articles suivent clairement une structure IMRAD dans un corpus de TECHNETIUM, au moyen de titres ou de sous-titres notamment. Dans un tel corpus, on peut également s'attendre à ce que d'autres articles ne suivent pas aussi clairement la structure IMRAD, en n'utilisant pas les titres ou sous-titres canoniques ou en mélangeant certaines sections par exemple. Le cas échéant, un tel corpus est composé d'un sous-ensemble de données déterminées et un autre composé de données indéterminées. En somme, nous avons montré qu'il est possible de résoudre cette problématique à l'aide d'une variante de l'identification des données indéterminées au moyen de données déterminées. Dans un premier temps, nous avons montré que plusieurs approches issues du domaine de l'apprentissage machine permettent la modélisation de la tâche de détection de l'appartenance des contextes de citations aux macrostructures IMRAD. Cette modélisation nécessite un ensemble de données où la provenance IMRAD des contextes de citation est déterminée. Suivant la modélisation de cette tâche au moyen des données déterminées, nous avons utilisé et évalué cette modélisation sur des données indéterminées afin d'identifier la fonction IMRAD de la structure d'où provient les contextes de citation indéterminés.

8.5 Conclusion

Dans ce chapitre, nous avons montré que l'identification des contextes de citation aux macrostructures IMRAD est possible et nous avons déployé une analyse approfondie de différentes méthodologies pour cette tâche. L'expérimentation a montré une correspondance entre les mots les plus discriminants et les fonctions argumentatives des macrostructures IMRAD. Elle s'est poursuivie avec l'application de diverses techniques de représentation vectorielle du texte et de méthodes de classification, engendrant un total de 153 modèles testés. Plus précisément, l'application de 8 techniques de représentation vectorielle et de 19 méthodes de classification sur nos données textuelles déterminées génère 152 modèles de détection de l'appartenance des contextes de citation aux macrostructures IMRAD. À ceux-ci s'ajoute le modèle de classification de séquence BERT pour un total de 153 modèles. Les 20 méthodes de classification ont été regroupées en familles soit i) les modèles de classification basée sur la catégorisation (2 modèles), ii) les modèles de classification instanciée (3 modèles), iii) les modèles de classification linéaire (8 modèles), iv) les modèles de classification arborescente (4 modèles), et v) les modèles de classification neuronale (3 modèles). Les données utilisées ont semblé adéquates pour cette étude. Le support – le nombre de cas dans chaque classe – semble avoir été suffisant pour une modélisation de la détection de l'appartenance

des contextes de citation à la macrostructure IMRAD de laquelle ils proviennent, à l'exception de la macrostructure « résultats » dont le support semble suboptimal. L'évaluation quantitative de l'application d'un modèle performant sur nos données indéterminées a montré la capacité du modèle à généraliser son apprentissage machine sur ce type de données. Enfin, il est possible de tirer plusieurs conclusions de notre analyse de la correspondance Mots-Macrostructures IMRAD, des techniques de représentation, des méthodes de classification, des modèles combinant une technique de représentation et une méthode de classification ainsi que de l'application du modèle de classification de séquence BERT sur nos données indéterminées.

La correspondance Mots-Macrostructures IMRAD : Les mots les plus représentatifs des différentes sections IMRAD concordent avec leurs fonctions argumentatives respectives. Dans la section de discussion, les mots-clés reflètent divers aspects de l'argumentation, tels que la présentation de découvertes, enrichie d'exemples, ainsi que l'utilisation d'inférences prudentes. On note également la présence de termes marquant des contrastes ou des limitations et des références aux développements futurs. Ces éléments linguistiques correspondent étroitement aux objectifs sous-jacents de la section de discussion. Dans la section de l'introduction, les mots prédominants s'alignent avec la mise en place d'un contexte théorique, souvent sous-tendu par des notions d'actualité, d'importance ou d'ampleur. De plus, ces introductions intègrent souvent des exemples et des contrastes, en accord avec les buts de cette section. Les termes dominants dans la section de méthodes soulignent la précision et la spécificité des actions entreprises, comme l'usage ou l'exécution de procédures. Ces mots facilitent la description, le calcul et la détermination de divers éléments, souvent liés aux résultats obtenus. L'emploi fréquent de verbes dans cette section reflète la nature active et procédurale des méthodes. Les mots-clés de la section résultats illustrent la présentation de données ou de constatations, souvent marquées par des références temporelles ou séquentielles, et la capacité de visualiser ces informations via des figures ou des tableaux. Ils incluent également des termes liés à l'interprétation et à l'évaluation, soulignant l'importance de l'analyse et de l'appréciation des résultats. Bref, l'alignement des mots les plus représentatifs avec les fonctions des sections IMRAD confirme la pertinence et l'efficacité des choix linguistiques dans ces textes. Chaque section utilise un lexique spécifique qui soutient ses objectifs argumentatifs. Bien que les verbes soient prédominants dans la section des méthodes, ils ne sont pas aussi déterminants dans les autres sections IMRAD. Cela souligne une distinction dans la manière dont les différentes sections communiquent leurs contenus. La compréhension de ces régularités linguistiques peut être cruciale pour améliorer les

techniques de classification automatique des textes, en permettant de mieux identifier et classer les différents segments d'un document en fonction de leur rôle et contenu spécifique.

Les techniques de représentation : La technique de représentation vectorielle qui s'est avérée la plus performante pour notre tâche de classification est BERT_imrad qui est obtenue à la suite de l'affinage du modèle BERT de base au moyen du modèle de classification de séquence BERT sur nos données textuelles déterminées. L'entraînement sur nos données s'est opéré sur 10 époques (10 passages complets sur les données), mais nos résultats montrent que 5 époques auraient été suffisantes (ce qui aurait permis d'effectuer cette tâche en la moitié du temps, donc quelques jours dans notre cas). En fait, la représentation BERT_imrad s'est avérée plus performante pour les 20 méthodes de classification testées, ce qui montre l'efficacité de cette approche. À l'inverse, DBOW s'est avérée la représentation la moins performante pour 19 des 20 méthodes de classification testées. Il est peu probable que la technique de représentation DBOW soit moins performante en soi, notamment en comparaison à une représentation plus simple comme BOW. Lareau (2022), par exemple, a montré par une étude de cas que la représentation DBOW peut être plus performante que la représentation BOW pour une tâche complexe de lecture et d'analyse de texte de haut niveau théorique. La représentation DBOW ayant été entraînée à partir de rien sur nos données textuelles, une paramétrique suboptimale peut avoir faussé le processus de représentation. Autrement, il est possible que l'ampleur du corpus soit insuffisante pour générer une représentation DBOW performante. Au total, 28 des 153 modèles combinant une technique de représentation et une méthode de classification offrent une F-mesure moyenne pondérée supérieure à 0,7. Subséquemment, nous nommons l'ensemble de ces modèles « top-28 » que nous considérons relativement performant pour la tâche qui nous intéresse. En ne considérant que le top-28, la seconde représentation la plus performante est OAI obtenue d'OpenAI « text-embedding-ada-002 ». En se concentrant uniquement sur le « top-28 », on constate que BERT large et BERT de base occupent respectivement la troisième et la quatrième position, bien que l'écart de performance entre ces deux types de représentations soit faible. Il est important de noter que les modèles BERT_nli_stsb et BERT_mpnet, qui sont des versions de BERT affinées pour des tâches de classification autres que IMRAD, ne figurent pas parmi le top-28. Qui plus est, ces dernières performant moins bien que les représentations originales dans 91% des cas (100% pour le top-28). Cela suggère que, dans le cas de notre étude, le transfert d'apprentissage d'une tâche à l'autre n'est pas efficace, surtout si l'on considère que l'affinement d'un modèle BERT sur des problèmes de type NLI, STSB ou MPNET semble réduire la performance du modèle BERT original pour des tâches de classification IMRAD. Cependant, étant donné que BERT_mpnet est

reconnu pour sa haute performance dans diverses tâches, il est clair que le modèle acquiert des connaissances supplémentaires par rapport au modèle original, et que ces connaissances pourraient être pertinentes pour notre problème spécifique. Il est donc crucial de comparer les performances entre un modèle de base et sa version dérivée pour confirmer la validité du transfert de connaissances. Puisque ce transfert n'a pas eu lieu dans notre cas, une perspective plus adéquate pourrait être de considérer ces modèles non pas tant comme des vecteurs de transfert de connaissances, mais plutôt comme des outils pour restructurer les données dans leur espace de représentation. Cette restructuration met en exergue certains aspects des données tout en minimisant d'autres. Si les éléments mis en avant par cette restructuration correspondent aux caractéristiques définissant les classes dans notre problème de classification, cela se manifeste par une amélioration des performances par rapport à l'espace de représentation initial. En revanche, si la restructuration atténue des aspects cruciaux pour la définition des classes de notre problème, cela entraîne une baisse des performances en comparaison à l'espace original. Cela est corroboré par le fait que le nombre de paramètres de la technique de représentation vectorielle n'est pas garante d'une meilleure performance puisqu'une représentation moins complexe peut être plus performante. Par exemple, la représentation BERT_imrad (110 millions de paramètres produisant des vecteurs de 768 dimensions) est systématiquement plus performante que BERT_large (340 millions de paramètres produisant des vecteurs de 1024 dimensions) et AOI (plus de 175 milliards de paramètres produisant des vecteurs de 1536 dimensions). Néanmoins, le nombre de paramètres semble un bon indicateur de performance lorsqu'on fait abstraction de la restructuration des données, du moins pour les six (6) modèles du top-28 dans lesquels on retrouve à la fois BERT_base, BERT_large et OAI. En résumé, bien que le nombre de paramètres dans une technique de représentation soit un élément important influençant la performance, ce n'est pas le facteur le plus déterminant.

Les méthodes de classification : Il n'aurait pas été étonnant que le modèle générant la représentation la plus performante – la classification de séquence BERT – soit aussi la méthode de classification la plus performante. Or, ce n'est pas le cas. Bien que la différence ne soit pas énorme, sept (7) modèles de la famille des classifieurs linéaires et deux (2) de la famille des classifieurs arborescents sont plus performants que le modèle de classification de séquence BERT, soit le SVM à noyau linéaire, le SVM à noyau RBF, la régression logistique, la régression logistique avec validation croisée, la régression Ridge, la régression Ridge avec validation croisée, la descente de gradient, les forêts aléatoires et l'amplification de gradient. Globalement, les modèles testés les plus performants appartiennent à la famille des méthodes de classification linéaire (75% du top-28). Plus précisément, tous les modèles du top-28 n'utilisant pas la

représentation BERT_imrad sont composés d'une méthode de classification de la famille des classifieurs linéaires. Dans le top-28, seulement quatre (4) méthodes de classification offrent des performances intéressantes pour chacune des quatre (4) représentations les plus performantes (BERT_base, BERT_large, OAI et BERT_imrad), soit le SVM linéaire, le SVM à noyaux RBF, la régression logistique et la régression logistique avec validation croisée. Tous les autres modèles du top-28 à l'exception du classifieur passif-agressif associé à OAI sont des modèles composés de la représentation BERT_imrad. Subséquemment, nous recommandons l'usage d'un de ces quatre (4) classifieurs pour le problème qui nous intéresse, car ils performant bien avec différents types de représentation. De plus, ces quatre (4) méthodes de classification permettent de vérifier si le problème de classification en est un convexe, car les SVM sont conçus pour maximiser la marge entre les classes et la régression logistique utilise une fonction de perte convexe. Dans notre cas, puisque la classification de séquence BERT permet la résolution de problèmes convexes sans si restreindre et que les SVM ainsi que la régression logistique ne résolvent que des problèmes convexes (parfois avec la méthode noyaux), étant donné la faible différence entre ces méthodes pour notre problème de classification, force à croire que l'identification des contextes de citation aux macrostructures IMRAD est un problème convexe.

Les modèles combinant une technique de représentation et une méthode de classification : Quatre (4) modèles particuliers possèdent des performances et des caractéristiques intéressantes pour résoudre notre problème de classification. D'abord, le modèle de classification de séquence BERT, car il permet la génération de la représentation BERT_imrad et son classifieur offre une performance relativement élevée (F1 pondérée de $\sim 0,748$). Bref, ce modèle est performant, libre d'accès, mais relativement long à entraîner. Ensuite, le modèle associant la régression logistique avec validation croisée à la représentation vectorielle issue de la classification de séquence BERT, car il offre la performance la plus élevée (F1 pondérée de $\sim 0,753$). Ce modèle est performant, libre d'accès, implique le précédent modèle long à entraîner, et permet de vérifier si le problème est convexe. En termes de performance, le modèle composé d'un SVM à noyaux RBF et de la représentation vectorielle d'OpenAI offre une performance impressionnante (F1 pondérée de $\sim 0,722$). Ce modèle est performant, relativement rapide (il n'implique pas le modèle BERT_imrad long à entraîner), permet de vérifier si le problème est convexe, mais il est propriétaire (son fonctionnement est une boîte noire pour les usagers). Enfin, le modèle composé d'un SVM à noyaux RBF et de la représentation vectorielle BERT de base offre une performance intéressante (F1 pondérée de $\sim 0,711$). Ce modèle est performant, relativement rapide, permet de vérifier si le problème est convexe et il est libre d'accès.

L'application du modèle de classification de séquence BERT sur nos données indéterminées : L'application du modèle de classification de séquence BERT sur nos données indéterminées indique un potentiel de généralisation du modèle vers les cas indéterminés. La mesure de rappel maximal de 0,76 observée pour l'application du modèle aux données indéterminées est conséquente à la mesure de rappel moyenne réelle de $\sim 0,7$ observée sur l'application du modèle de classification de séquence BERT sur le sous-ensemble d'évaluation des données déterminées. Comme le montre la correspondance Mots-Macrostructures IMRAD, les contextes de citation semblent contenir des informations suffisamment discriminantes pour permettre l'identification de leur appartenance aux macrostructures IMRAD. Par conséquent, nous recommandons l'approche en deux (2) étapes qui consiste à entraîner un classifieur sur des données déterminées pour ensuite l'appliquer sur des données indéterminées lors d'un problème de classification tel que le nôtre. Entre autres, cette approche peut être utilisée lorsqu'un corpus est composé de TECHNETIUM où les structures IMRAD sont facilement identifiables (à l'aide de sous-titres notamment) ainsi que d'autres où ces structures sont plus difficilement identifiables, en utilisant notamment les contextes de citation.

CHAPITRE 9

La détection des énoncés de thèse principale

Au chapitre précédent, nous explorions la tâche d'identification de l'appartenance des contextes de citation aux macrostructures de type IMRAD. Les contextes de citation sont des composantes argumentatives jouant le rôle d'articulation entre les structures argumentatives intratextuelles et les structures argumentatives extratextuelles. Les macrostructures IMRAD répondent à des sous-objectifs qui participent conjointement à l'objectif principal des textes scientifiques ou philosophiques, c.-à-d. le soutien d'une ou de plusieurs thèses principales. Dans ce chapitre, nous nous intéressons à la tâche d'identification des thèses principales dans les textes de haut niveau théorique publiés dans les revues à comité de lecture.

9.1 Introduction

Les pratiques philoscience sont des activités à la fois sociales et cognitives. Sociales, d'une part, car elles sont effectuées au sein d'une communauté de pratique (Wenger *et al.*, 2002) et cognitives, d'autre part, car elles impliquent des procédés intellectuels comme l'observation, la description, la théorisation, l'argumentation, l'expérimentation, le jugement, l'évaluation, la généralisation, etc. Ultiment, ces pratiques sont effectuées ou publicisées à l'aide du langage, car le dialecte spécialisé de la science et de la philosophie est le moyen par lequel on fait sens du monde, de soi et des autres (Lemke, 1990, p. ix). Ainsi, le discours philoscience est peuplé de performances linguistiques complexes dont le procédé central est l'argumentation, c.-à-d. l'articulation d'arguments, de contre-arguments, de conclusions ou de thèses (Cope *et al.*, 2013 ; Kuhn, 1993). Ce faisant, une tâche importante de l'analyse du discours philoscience est l'identification des thèses principales, laquelle peut être assistée par ordinateur (Levy *et al.*, 2014). Plusieurs méthodes permettent l'identification de thèses principales, notamment celles basées sur les mots-clés, le positionnement, l'apprentissage automatique, le résumé automatique, ou l'analyse des structures syntaxiques et sémantiques. Les méthodes basées sur les mots-clés impliquent l'identification de mots ou d'expressions spécifiques qui indiquent la présence d'un énoncé de thèse principale. En recherchant ces mots-clés, il est possible de localiser les énoncés de la thèse principale dans le texte (Knott et Dale, 1994 ; Levy *et al.*, 2014, 2017). Une autre approche est fondée sur le positionnement des énoncés de la thèse principale dans le texte, ce qui est possible lorsque ceux-ci se trouvent à des positions spécifiques dans le texte. Le cas échéant, l'analyse de la structure argumentative ainsi que de la position des composantes argumentatives permet d'identifier les énoncés de la thèse centrale (Duan *et al.*, 2019).

L'apprentissage machine est une approche consistant à entraîner un modèle d'apprentissage automatique sur un ensemble de données étiqueté par des annotateurs (humains ou automates) ayant préalablement identifié les énoncés de thèse principale. Le modèle apprend à classer les phrases ou les passages exprimant ou n'exprimant pas la thèse principale en fonction de diverses caractéristiques textuelles. Une fois entraîné, le modèle peut être utilisé pour détecter automatiquement les énoncés exprimant une thèse principale dans de nouveaux textes (Lippi et Torroni, 2015). L'approche du résumé automatique utilise les techniques de résumé textuel pour extraire l'idée principale ou l'affirmation centrale d'un texte donné. Ces méthodes visent à condenser le texte en un texte plus court qui capture les informations essentielles dont, notamment, l'énoncé de la thèse principale (Bhatnagar *et al.*, 2022 ; Li et Zhou, 2020). L'approche fondée sur l'analyse des structures syntaxiques et sémantiques du discours utilise des techniques de traitement du langage naturel telles que l'annotation des parties du discours, l'analyse des dépendances ou la reconnaissance des entités nommées. En identifiant les relations entre les mots et les phrases, ces techniques peuvent aider à identifier la thèse principale d'un texte donné (Levy *et al.*, 2017 ; Lippi et Torroni, 2015). Dans cette étude, nous nous intéressons plus précisément aux approches fondées par le positionnement, les mots-clés et l'analyse des structures syntaxiques ou sémantiques. Telle qu'énoncée au chapitre 7, notre hypothèse principale est que la détection efficiente des thèses principales dans le TExte Court de Haut NivEau Théorique publié dans une RevUe à CoMité de lecture (TECHNETIUM) est possible au moyen de ces approches (H2). Une première hypothèse subordonnée est que cette détection peut s'effectuer via l'emplacement (H2.1). Une seconde hypothèse subordonnée est que certains connecteurs de discours peuvent être utilisés pour l'identification des thèses principales (H2.2). Une troisième hypothèse subordonnée est que certaines structures syntaxiques ou sémantiques permettent la détection des thèses principales (H2.3). Plus précisément, nous postulons que : a) certaines structures périphériques aux macrostructures IMRAD entretiennent des relations privilégiées à la thèse principale, notamment le titre (H2.3.1) en raison de sa fonction de préparation thématique et le résumé (H2.3.2) en raison de sa fonction de résumé, et b) la thèse principale joue un rôle argumentatif privilégié dans certaines macrostructures IMRAD, notamment l'introduction (H2.3.3) et la discussion (H.2.3.4) en raison de la relation de solution (*solutionhood*) qui les relie.

9.2 Contexte et état de l'art

La thèse principale réfère à l'idée centrale ou la proposition fondamentale que présente un texte, un article, un essai ou une recherche. Elle est le pilier autour duquel s'articule l'ensemble du travail, orientant l'analyse, la discussion et les conclusions. Dans le cadre de la théorie communicationnelle de

l'argumentation, la thèse principale est un élément – une composante argumentative ou un segment de texte par exemple – qui a la particularité d'être supporté ou attaqué par d'autres éléments, mais qui ne supporte ou n'attaque aucun autre élément. Dans cette perspective, la thèse principale est l'objet de discussion et d'analyse, soutenue ou contestée par d'autres éléments argumentatifs dans le texte, mais elle-même n'agit pas en tant que support ou contre-argument pour d'autres propositions. Techniquement, nous définissons la thèse principale comme un objet textuel abstrait ayant la propriété d'être, soit vrai, soit faux – une proposition – lequel est soutenu ou attaqué par une ou plusieurs propositions, mais qui ne soutient ou n'attaque aucune autre proposition. Ainsi, la notion de proposition est sémantique en raison de son rapport à la vérité⁷¹, et celle de thèse est logique ou épistémique en raison du lien (logique ou épistémique) envers des arguments (ou contre-arguments). Une thèse est donc une proposition dont la vérité est soutenue par au moins un argument, et la thèse principale est une thèse qui ne soutient ni n'attaque d'autres thèses (Lacharité, 1989).⁷²

Dans le cadre des textes structurés IMRAD, dans les TECHNETIUM notamment, la notion de thèse principale peut également être définie comme une solution à une question de recherche. Le cas échéant, la macrostructure de discussion semble être un endroit privilégié où trouver les thèses principales puisque celle-ci est un lieu de discussion de solutions à une question de recherche. Ces deux définitions ne sont pas contradictoires. La première souligne un aspect structural vertical ou hiérarchique de la thèse principale. Elle est conséquente avec la théorie communicationnelle de l'argumentation définissant celle-ci en termes d'articulations de raisons qui soutiennent et de raisons qui s'opposent à une thèse principale. La seconde souligne un aspect structural horizontal ou linéaire de l'argumentation. En effet, l'argumentation est un phénomène de communication se réalisant dans une forme linguistique particulière. Ce faisant, l'argumentation peut prendre la forme d'un récit philoscience de forme linéaire (Suppe, 1998). D'ailleurs, le texte philoscience est souvent accompagné de figures, de graphes, de tableaux, de puces (numériques ou iconiques) ou de sous-titres afin de souligner certains aspects structuraux difficilement (re)présentables au moyen d'un récit linéaire. Cependant, pour des raisons

⁷¹ Notons que la notion de vérité est complexe et fait l'objet de nombreux débats. Différentes approches sont proposées, notamment celles basées sur la correspondance, la cohérence, la pragmatique, la linguistique, la sémantique, etc. Entre autres, une approche pragmatique permet de relier cette notion à l'action (Misak, 2018).

⁷² En tant qu'objet abstrait, une même thèse principale peut être exprimée plusieurs fois dans un texte au moyen de divers énoncés. Ceux-ci peuvent être strictement identiques ou approximativement identiques lorsque les mêmes termes sont utilisés, ou encore être similaires en faisant l'objet de paraphrases ou de reformulations. Subséquemment, nous utiliserons indifféremment les termes « thèse principale » et « énoncé de thèse principale ».

pragmatiques, nous adopterons la définition verticale de la notion de thèse principale puisque le corpus que nous utiliserons – AbstrCT – se base sur celle-ci lors du processus d’annotation des thèses principales.

9.2.1 Approche fondée par le positionnement

Comme vu précédemment, le positionnement de la thèse principale dans un TECHNETIUM suit généralement une structure conventionnelle, qui a été étudiée et décrite dans diverses recherches sur la rédaction académique et la communication scientifique. Généralement, la thèse principale se trouve au début du texte dans la section « introduction » et/ou à la fin du texte dans la section « discussion ». Des chercheurs tels que Swales et Feak (2004) ont expliqué comment les introductions dans les textes académiques suivent souvent un modèle de création d’un espace de recherche (*Create a Research Space* ou *CARS*), dans lequel les auteurs présentent leur thèse ou leur revendication principale après avoir établi un contexte et énoncé le besoin de leur recherche. Sous forme de réponse à l’objectif de recherche, la thèse principale est souvent élaborée dans la section « discussion » ou « conclusion ». Cette partie interprète les résultats, discute de leurs implications et montre comment ils contribuent à répondre à la question de recherche ou à atteindre l’objectif de recherche. De nombreux chercheurs ont observé cette structure dans les articles scientifiques (Day et Gastel, 2011 ; Gustavii, 2000). Bref, la recherche semble indiquer que les thèses principales tendent vers un positionnement spécifique dans les TECHNETIUM.

9.2.2 Approche fondée par les marqueurs

Les relations discursives sont matériellement fondées par des marqueurs jouant le rôle de connecteurs entre les parties du discours (Weber et coll., 2019). Le connecteur de discours est un cas particulier de marqueur discursif dont la spécificité est de signaler une relation rhétorique en permettant une structuration du discours en vue de l’atteinte d’un objectif global (Lewis, 2000). La recherche propose aujourd’hui des outils informatiques de reconnaissance des relations discursives à l’aide de connecteurs de discours. Par exemple, le *Penn Discourse TreeBank* (PDTB) est une ressource linguistique qui représente la structure du discours au sein d’un corpus de textes provenant du *Penn Treebank* et principalement constitué d’articles du *Wall Street Journal*. Plutôt que de se concentrer sur la structure de surface de la phrase, comme c’est généralement le cas avec la plupart des *treebanks*, le PDTB se concentre sur la structure du discours. Ce faisant, le PDTB utilise une variété de relations de discours, telles que la cause, le contraste et l’élaboration, pour annoter le texte. Ces relations peuvent être explicites (signalées par des connecteurs indépendamment du contexte), quasi explicites (signalées par des connecteurs dépendamment du contexte) ou implicites (lorsque les relations sont inférées du contexte sans

connecteurs explicites) (Scheffler et Stede, 2016). Le PDTB a été développé à l'Université de Pennsylvanie et le processus d'annotation implique des annotateurs formés qui balisent les relations de discours en indiquant le connecteur de discours explicite (s'il est présent) et en identifiant le type de relation entre les étendues de texte connectées. Le schéma d'annotation PDTB comprend un ensemble de types de relations prédéfinis, tels que « cause », « contraste », « élaboration », « évaluation », « temporel » et bien d'autres. Chaque relation est associée à des caractéristiques linguistiques et contextuelles spécifiques, fournissant un ensemble d'informations pour l'analyse du discours. Le corpus annoté qui en résulte favorise l'étude de phénomènes de discours, le développement de modèles informatiques de compréhension du discours et la recherche sur les structures et la cohérence du discours. Le PDTB a été largement utilisé pour entraîner des modèles automatiques de l'analyse du discours, ce qui peut avoir des applications dans la détection automatique des arguments, la génération de texte, la réponse aux questions, et d'autres domaines du traitement automatique de la langue naturelle. Au tableau 9.1 suivant, on voit quelques exemples de relations discursives organisées en trois niveaux typologiques et exemplifiées par des connecteurs.

Tableau 9.1 Exemples de relations discursives organisées en trois niveaux typologiques et exemplifiées par des connecteurs

Niveau 1	Niveau 2	Niveau 3	Exemples de connecteur
Relations temporelles	Synchrone	-	when, while, still
	Asynchrone	Précédent	later, then, until
		Successeur	once, since, after
Relations contingentes	Cause	Cause-raison	because, about
		Effet (positif)	hence, accordingly
		Effet négatif	no evidence, too late
	Cause + croyance	Cause + croyance	with
		Effet + croyance	indicating
	Cause + acte langage (a. l.)	Cause + acte lang.	
		Effet + acte lang.	
	Condition (cond.)	Arg1 = condition	when, if
		Arg2 = condition	as long as
	Cond. + a. l.	-	
Condition négative (cond. nég.)	Arg1 = cond. nég.	Otherwise, else	
	Arg2 = cond. nég.	unless	
Cond. nég. + a. l.	-		
Pragmatique	Arg1 = but	by, for that purpose	

		Arg2 = but	so that, in order
Relations de comparaison	Concession (conc.)	Arg1 = dénie	even if, although
		Arg2 = dénie	but, however
	Conc. + a. l.	Arg2 = nié + a. l.	
	Contraste	-	on the other hand
	Similarité	-	similarly, so too
Relations d'expansion	Conjonction	-	and, additionally
	Disjonction	-	or, as an alternative
	Équivalence	-	in other words
	Exception	Arg1 = exception	otherwise
		Arg2 = exception	except
	Instanciation	Arg1 = instance	in general
		Arg2 = instance	for exemple
	Niveau de détail	Arg1 = détail	in short, in sum
		Arg2 = détail	specifically, in that
	Méthode	Arg1 = comment	thereby
Arg2 = comment		by	
Substitution	Arg1 = substitut	rather than	
	Arg2 = substitut	instead	

9.2.3 Approche fondée par l'analyse des structures syntaxiques ou sémantiques

L'analyse du contenu sémantique et syntaxique des textes implique de comprendre à la fois la signification des mots et des phrases (sémantique) ainsi que les règles et la structure du langage (syntaxe). L'analyse de ces structures vise à reconnaître les aspects sémantiques ou syntaxiques afin d'identifier des composants spécifiques dans le texte, tels que les thèses principales. La sémantique fait référence au sens véhiculé par des mots et des expressions dans un contexte spécifique. L'analyse sémantique implique de comprendre les relations entre les mots, la signification des phrases et le contexte global d'une section de texte. L'analyse sémantique est utilisée pour l'analyse des sentiments, la modélisation de sujets et la compréhension des structures thématiques dans le texte. En analyse des arguments, cela peut aider à identifier la position ou les principaux éléments thématiques d'un argument. La syntaxe concerne l'agencement de mots et d'expressions pour créer des phrases bien formées. Cela implique de comprendre les structures grammaticales, la construction des phrases et les relations entre les différentes parties d'une phrase. L'analyse syntaxique identifie la structure grammaticale des phrases pour comprendre les relations entre les différents mots et permet l'identification des catégories grammaticales (comme les noms, les verbes, les adjectifs). Pour l'analyse des arguments, l'analyse syntaxique peut aider à identifier

la structure des arguments, comme les prémisses et les conclusions. La combinaison de l'analyse sémantique et syntaxique permet une compréhension sophistiquée du texte, permettant l'identification de composants complexes comme ceux trouvés dans les arguments. Alors que l'analyse sémantique donne un aperçu de ce qui est dit, l'analyse syntaxique révèle la façon dont cela est dit, toutes deux cruciales pour disséquer les arguments structurés dans les textes. Les propriétés sémantiques et syntaxiques peuvent être capturées par des modèles comme Word2Vec, BERT ou GPT qui représentent des mots dans un espace de grande dimension où les mots sémantiquement similaires sont plus proches les uns des autres, et où le contexte contenant entre autres des informations syntaxiques peut influencer la représentation textuelle.

9.3 Les méthodes

Les énoncés exprimant une thèse principale, aussi nommés énoncés de thèse ou propositions de thèse, c.-à-d. les segments de texte où on exprime une thèse principale, sont des composantes discursives importantes étudiées en philosophie, en linguistique, en analyse du discours et de l'argumentation ainsi qu'en pédagogie. Dans un contexte plus large, tout domaine qui s'occupe de l'analyse ou de la rédaction argumentative ou analytique – que ce soit en philosophie, en littérature, en histoire, en sciences, etc. – abordera la notion de thèse principale, car elle est fondamentale pour structurer et guider une argumentation. L'analyse des thèses principales permet l'exploration des idées, des thèmes, et des arguments qui sont utilisés et diffusés dans un texte. Une question de recherche est d'identifier quels sont les énoncés d'une thèse principale dans les textes de haut niveau théorique et, plus spécifiquement, dans les TECHNETIUM. Pour ce faire, un corpus composé de tels textes dans lesquels les énoncés de la thèse principale sont identifiés sera nécessaire afin de mesurer la performance de différentes méthodes de détection des thèses principales. Malheureusement, il existe très peu de corpus accessibles dans lesquels les thèses principales sont identifiées, et encore moins des corpus composés de TECHNETIUM annotés de leurs thèses principales. Néanmoins, il existe au moins un corpus accessible qui répond à ce critère de sélection, sous réserve cependant. AbstrCT est un corpus développé par Mayer, Cabrio et Villata (2020) constitué de résumés d'articles scientifiques dont les phrases sont annotées de leur rôle argumentaire, notamment celui de thèse principale. Un avantage d'AbstrCT est qu'il est composé de TECHNETIUM. Un désavantage est que les TECHNETIUM qui composent ce corpus sont exclusivement des articles scientifiques issus du domaine de la médecine. Bref, AbstrCT est un échantillon de TECHNETIUM issus d'un domaine spécifique de la science. En utilisant ce corpus pour notre étude, il semble raisonnable de pouvoir généraliser certaines de nos conclusions au sous-domaine scientifique qu'est la médecine, un peu moins

pour la science en général, et encore moins pour le domaine philosophique. Cependant, nos résultats contribueront certainement à mettre en lumière des preuves de concept ou des pistes de recherche intéressantes pour tous ces domaines. Dans ce corpus, spécifions que la définition de thèse principale est compatible avec la nôtre. Selon les auteurs, une thèse principale exprime la conclusion d'un processus inférentiel et, en ce sens, elle est instanciée par un énoncé plus général et conclusif qu'un énoncé de thèse subordonnée ou de prémisse :

[M]ajor claims are more a general/concluding claim, which is supported by other, more specific, claims. [...] The major claims are more general statements, for which the other claims serve as evidence. The concluding statements do not have to occur at the end of the abstract, and may also occur at the beginning of the text as an introductory claim. (Mayer *et al.*, 2018, p. 143)

Spécifions que les auteurs distinguent les segments textuels qui sont des composantes argumentatives de ceux qui ne le sont pas. Parmi les composantes argumentatives, on trouve les prémisses (*evidence*), les thèses subordonnées (*claim*) et les thèses principales (*major claim*). En général, les prémisses peuvent être reliées à des thèses subordonnées par des relations de support ou d'attaque, et les thèses subordonnées peuvent être reliées à des thèses principales au moyen des mêmes relations.

9.3.1 La constitution du corpus

La constitution du corpus AbstrCT s'inspire de la méthodologie de Trenta, Hunter et Riedel (2015) qui a mené à la constitution du corpus RCT (de l'anglais : *Randomized Controlled Trial* ou RCT) composé de 169 résumés d'articles scientifiques portant sur quatre (4) maladies identifiées à l'aide de mots-clés, c.-à-d. glaucome, hypertension, hépatite B, et diabète (Mayer *et al.*, 2018). Pour sa part, le corpus AbstrCT ajoute 500 résumés d'articles scientifiques issus de la base de données MEDLINE via PubMed en ajoutant le terme « neoplasm » aux mots-clés et en précisant qu'il doit s'agir d'un essai clinique contrôlé randomisé (ECR). En suivant le protocole d'annotation des composantes argumentatives proposé par Mayer (2020), deux annotateurs ont effectué l'annotation des résumés en indiquant deux (2) types de relation pouvant relier trois (3) types de composantes argumentatives nommées « évidence », « thèse » ou « thèse principale ».⁷³ À terme, le corpus AbstrCT est composé de 669 résumés qui sont annotés de composantes telles que la prémisse ou évidence (*#Evi*), la thèse subordonnée (*#Claim*) ou la thèse principale (*#MajCl*),

⁷³ Comme le soulignent Mayer *et al.* (2020), l'accord interannotateur pour cette tâche s'est révélé élevé en ce que le kappa de Fleiss (calculé sur 30 résumés) donne 0,72 pour les composants argumentatifs et 0,68 pour la distinction entre les affirmations et les preuves.

ainsi que des relations telles que le support (*#Sup*) ou l'attaque (*#Att*). Le tableau 9.2 suivant montre la répartition des types de composantes et d'annotations du corpus AbstrCT en fonction du thème principal déterminé à l'aide des mots-clés correspondants (tiré de Mayer *et al.*, 2020).⁷⁴

Tableau 9.2 Répartition des types de composantes et d'annotations du corpus AbstrCT en fonction du thème principal

<i>Dataset</i>	<i>#Evi</i>	<i>#Claim</i>	<i>#MajCl</i>	<i>#Sup</i>	<i>#Att</i>
<i>Neoplasm</i>	2193	993	93	1763	298
<i>Glaucoma</i>	404	183	7	334	33
<i>Hepatitis</i>	80	27	5	65	1
<i>Diabetes</i>	72	36	11	44	8
<i>Hypertension</i>	59	26	9	53	2
<i>Total</i>	2808	1265	125	2259	342

Suivant l'inspection du corpus AbstrCT, on observe que seulement 113 résumés contiennent au moins un énoncé de thèse principale. Puisque la tâche qui nous intéresse est la détection des thèses principales (et non pas la détection des autres composantes argumentatives identifiées dans le corpus AbstrCT), nous constituons notre corpus à partir d'AbstrCT en portant une attention particulière aux résumés contenant au moins un énoncé de thèse principale. Chaque texte composant le corpus est constitué de deux (2) documents, c.-à-d. un premier constitué du résumé complet de format TXT et un second résultant de la segmentation du même résumé en une suite de segments textuels où seuls ceux qui sont des composantes argumentatives sont inscrits, chacun séparé par un saut de ligne et annoté en début de ligne par un identifiant, le type de composante argumentative (*premise, claim, major claim*) ainsi que l'emplacement du premier et du dernier caractère du segment dans le résumé. Ensuite, dans chacun de ces documents où sont annotés les constituants, sont ajoutées les relations argumentaires chacune séparée par un saut de ligne et annotée d'un identifiant de relation, du type de relation (*support, partial-attack, etc.*) ainsi que l'identifiant des composantes argumentatives qui sont ainsi reliées. Le tableau 9.3 suivant montre un exemple tiré du corpus AbstrCT, mais adapté au format tabloïde afin de favoriser la lisibilité.

⁷⁴ Le corpus AbstrCT est disponible au <https://gitlab.com/tomaye/abstrct/>. Notons que le nombre de thèses principales (*#MajCl*) reporté dans ce tableau par les auteurs est erroné. Le corpus AbstrCT contient plutôt 126 thèses principales.

Tableau 9.3 Exemple tiré du corpus AbstrCT

T1	Major Claim	1	83	The impact of prolonged subclinical hyperthyroidism on quality of life is unclear.
T5	Premise	1322	1433	After 6 months, none of the quality of life parameters in the low-TSH group was different from baseline values.
T6	Premise	1434	1543	In the euthyroid group, motivation was significantly improved (Multidimensional Fatigue Index-20, P = 0.003),
T7	Premise	1544	1620	although this parameter did not differ from the reference group at baseline.
T8	Premise	1621	1735	A probable worsening in role limitations as a result of physical problems (Short Form-36; P = 0.050) was observed.
T9	Premise	1736	1781	No improvement in the SDQ score was observed.
T10	Claim	1782	1897	In summary, quality of life in patients with DTC and long-term subclinical hyperthyroidism in general is preserved.
T11	Claim	1898	1969	Restoration of euthyroidism in general does not affect quality of life.
R1	Partial-Attack	Arg1:T7	Arg2:T6	
R2	Support	Arg1:T5	Arg2:T11	
R3	Support	Arg1:T6	Arg2:T10	
R4	Support	Arg1:T9	Arg2:T11	
R5	Partial-Attack	Arg1:T8	Arg2:T10	

Spécifions que le nom de chaque document correspond à un identifiant unique propre à PubMed (PMID) permettant de retrouver les textes complets desquels proviennent les résumés.

9.3.1.1 La complétion des données textuelles

Afin de tester nos hypothèses à propos du rôle argumentatif privilégié de certaines macrostructures IMRAD (l'introduction et la discussion) ainsi que certaines structures périphériques aux macrostructures IMRAD (le titre et le résumé), notre corpus doit contenir ces structures et macrostructures, car celles-ci peuvent fournir des informations importantes afin d'identifier la thèse principale dans un texte. Or, le corpus AbstrCT ne les contient pas toutes de sorte qu'il est incomplet pour tester ces hypothèses. Subséquemment, nous ajoutons une tâche de complétion des données textuelles de manière à ajouter les structures et macrostructures « titre », « introduction » et « discussion » aux résumés annotés comportant au moins une thèse principale.⁷⁵ Dès lors, nous nommons le nouveau corpus produit « AbstrCT+ ».

⁷⁵ Étant donné le petit nombre de textes comportant au moins une thèse principale dans le résumé, l'ajout du titre, de l'introduction et de la discussion est opéré manuellement, c.-à-d. que le document original est inspecté et les

9.3.2 Le décodage du corpus

Le corpus AbstrRCT+ est constitué de 669 textes dont 559 ne contiennent que la structure du résumé et 110 contiennent les structures « titre » et « résumé » ainsi que les macrostructures « introduction » et « discussion ». Toutes ces structures et macrostructures sont segmentées en phrases, les phrases sont décomposées en une série de mots et les mots sont annotés de leur lemme ainsi que de leur fonction dans la phrase (leur catégorie grammaticale).⁷⁶ Les thèses principales dans le résumé, déjà identifiées dans le document ne contenant que les composantes argumentatives du résumé, sont identifiées dans le résumé complet au moyen de la distance Levenshtein qui est calculée entre chaque chaîne de caractère identifiée dans les composantes argumentatives correspondant aux thèses principales et toutes les chaînes de caractère identifiées comme phrase dans les résumés du corpus AbstrRCT+.⁷⁷ La phrase AbstrRCT+ ayant une distance minimale avec la thèse principale identifiée dans AbstrRCT est annotée comme étant une thèse principale. Enfin, une représentation vectorielle des phrases du corpus AbstrRCT+ est fournie à l'aide d'une technique *doc2vec*, c.-à-d. DBOW. Un des avantages d'utiliser cette approche par rapport à d'autres représentations par plongement tels que BERT ou « text-embedding-ada-002 » est la possibilité d'utiliser autre chose que des unités lexicales littérales lors de la représentation des textes, notamment l'usage des *lemma*. La lemmatisation est une opération de réduction des formes fléchies d'un mot à une forme commune qui correspond habituellement à l'entrée du mot dans un dictionnaire. Cette opération permet de réduire la dimensionnalité du vocabulaire d'un corpus donné en regroupant sous un même lemme des mots partageant une similitude sémantique. De plus, il est possible de réduire davantage le vocabulaire en ne gardant que certaines catégories grammaticales. Ces techniques de réduction sont parfois utilisées lors de l'analyse thématique, car elles permettent de contraster les thématiques en regroupant et en se concentrant sur les unités lexicales indiquant des informations thématiques importantes (Malaterre *et al.*, 2020b, 2021 ; Malaterre et Lareau, 2022, 2023d, 2023a). Or, puisque certaines de nos hypothèses

structures et macrostructures sont extraites à la main au moyen des sous-titres et autres indices permettant de les identifier. Lors de cette opération, seulement 110 des 113 textes se sont avérés à la fois accessibles et de langue anglaise (un texte étant en mandarin et un autre en italien).

⁷⁶ Nous utilisons le *procyon* Stanza dans Python pour la segmentation, la tokenisation, la lemmatisation et l'annotation de la catégorie grammaticale (Qi *et al.*, 2020).

⁷⁷ La distance de Levenshtein, également connue sous le nom de distance d'édition, est une mesure de la différence entre deux chaînes de caractères. Elle est définie comme le nombre minimum d'opérations d'édition (insertions, suppressions ou substitutions de caractères) nécessaires pour transformer une chaîne en une autre. Cette mesure est appropriée pour l'identification des phrases qui correspondent aux thèses principales dans le résumé complet, car les composantes argumentaires qui sont déjà identifiées comme thèses principales sont des segments de texte provenant intégralement du résumé complet. Autrement, cette mesure ne serait pas appropriée pour identifier des paraphrases ou des reformulations, mais ce n'est pas le cas ici.

indiquent que certaines structures et macrostructures entretiennent des liens avec la thèse principale par le biais d'un thème, il semble raisonnable d'utiliser une représentation vectorielle permettant de contraster des distinctions thématiques. Subséquemment, nous utilisons une représentation DBOW obtenue à partir des *lemma* dont la catégorie grammaticale est exclusivement l'adverbe, l'adjectif, l'auxiliaire, le verbe, le nom, le pronom ou le nom propre. Parallèlement, nous comparerons l'approche de représentation textuelle DBOW à celle plus sophistiquée développée par l'entreprise OpenAI et connue sous le nom de « text-embedding-ada-002 » (laquelle avait donné de bons résultats pour la tâche investiguée au chapitre précédent).

9.3.3 Les méthodes de détection de la thèse principale

9.3.3.1 L'identification de la thèse principale dans le corps du texte à l'aide de l'énonciation de la thèse principale dans le résumé.

Les TECHNETIUM sont généralement constitués de structures argumentatives de type IMRAD. Ce format est un agencement linéaire de structures textuelles jouant chacune un rôle argumentatif spécifique et dont certaines semblent propices à l'énonciation de thèses principales. La discussion est une macrostructure dans laquelle l'énonciation de la thèse principale est plus susceptible de se produire. Parallèlement, le corpus AbstrCT+ comporte des résumés dans lesquels au moins une thèse principale est identifiée. Subséquemment, nous postulons qu'il est possible d'identifier l'énoncé (ou les énoncés) de la thèse principale dans la macrostructure de discussion à l'aide de l'énoncé (ou des énoncés) de thèse principale dans le résumé. Plus précisément, l'énoncé de thèse principale est identifiable dans la mesure où ce dernier entretient une certaine relation de similitude envers son énonciation dans le résumé. Ainsi, nous proposons d'identifier les énoncés de thèse principale dans les discussions à l'aide de la mesure de similarité cosinus⁷⁸ maximale entre les énoncés de thèse principale déjà identifiés dans les résumés et toutes les phrases de la discussion. Afin d'évaluer cette méthode, les énoncés de thèse principale provenant de la structure « résumé » de 12 textes (environ 10% du corpus AbstrCT+) choisis au hasard seront présentés et comparés aux énoncés qui sont respectivement les plus similaires dans la macrostructure « discussion ».

⁷⁸ Puisque la représentation textuelle utilisée est fondée sur le plongement lexical, la mesure de distance appropriée est ici le cosinus et non pas la distance de Levenshtein. De plus, la thèse principale dans la discussion pourrait ne pas être strictement identique à son énonciation dans le résumé, ou faire l'objet d'une paraphrase ou d'une reformulation. Une représentation par plongement lexical ainsi qu'une mesure cosinus sont plus aptes à rendre compte de ce phénomène.

9.3.3.2 L'identification de la thèse principale à l'aide de l'emplacement

Un autre moyen d'utiliser l'emplacement lors de l'identification des énoncés de thèse principale est d'évaluer si le positionnement d'un énoncé dans une macrostructure permet de prédire le rôle de thèse principale. Il est possible de tester cette hypothèse dans la mesure où les énoncés de thèse principale sont annotés comme tels dans la structure du résumé, et les résultats de l'étape précédente permettront de déterminer les énoncés de thèse principale dans la macrostructure de discussion. Subséquemment, nous déterminerons la probabilité d'un énoncé de jouer le rôle de thèse principale en fonction de son positionnement dans la structure « résumé » ou la macrostructure « discussion ». Puisque tous les résumés et toutes les discussions ne contiennent pas le même nombre de phrases, alors une normalisation du positionnement sera effectuée, c.-à-d que le rang d'une phrase dans la structure ou macrostructure sera divisé par le nombre total de phrases dans cette structure ou macrostructure.

9.3.3.3 L'identification de la thèse principale à l'aide de connecteurs de discours.

Pour l'approche des connecteurs de discours, nous utilisons la version 3 du PDTB (PDTB-3) qui inclut 53 631 marqueurs de relations internes aux phrases (Intra-S) ou externes aux phrases (Extra-S) (Prasad et coll., 2019).⁷⁹ En outre, les textes du corpus PDTB-3 sont décomposés en paragraphes, lesquels sont décomposés en phrases, lesquelles sont décomposées en structures arborescentes de manière à identifier les parties discursives atomiques, puis les relations entre les parties (dans les phrases) et entre les phrases (dans les paragraphes) sont identifiées. Ainsi, nous postulons que certains marqueurs rhétoriques permettent d'identifier les phrases exprimant la thèse principale. Nous testerons cette hypothèse sur le corpus AbstrCT+.

9.3.3.4 L'identification de la thèse principale à l'aide de propriétés syntaxiques ou sémantiques

Pour l'approche des propriétés syntaxiques ou sémantiques, nous postulons que le titre, le résumé, l'introduction ou la discussion ont une similarité maximale avec l'énoncé de la thèse principale dans le résumé ainsi que la discussion. Pour tester cette hypothèse, nous calculons la valeur de similarité cosinus entre la structure ou la macrostructure investiguée et chacune des phrases de la structure ou de la macrostructure dans laquelle nous cherchons la thèse principale (le résumé ou la discussion). À partir de

⁷⁹ Les 53 631 tokens peuvent être regroupés en 174 types distincts. Étant donné qu'un même marqueur peut avoir plusieurs rôles, les 53 631 tokens peuvent également être regroupés en 418 catégories en tenant compte du type et de la fonction, comme indiqué dans l'annexe B.

la représentation issue de la technique DBOW, les représentations vectorielles des phrases de la structure ou de la macrostructure investiguée sont additionnées, puis cette somme est divisée par le nombre total de phrases de manière à produire un vecteur synthétique représentant la structure ou la macrostructure investiguée. La similarité cosinus est calculée entre ce vecteur synthétique moyen et chacune des phrases de la structure ou de la macrostructure dans laquelle nous cherchons la thèse principale (le résumé ou la discussion). La phrase entretenant une similarité cosinus maximale envers la structure ou la macrostructure investiguée est identifiée comme jouant le rôle de thèse principale. Pour ce qui est de la représentation « text-embedding-ada-002 », aucun vecteur synthétique n'est produit, car la structure ou la macrostructure est directement représentée par un vecteur propre.⁸⁰ La similarité cosinus est calculée entre ce vecteur propre et chacune des phrases de la structure ou de la macrostructure dans laquelle nous cherchons la thèse principale (le résumé ou la discussion). De la même manière, que précédemment, la phrase entretenant une similarité cosinus maximale envers la structure ou la macrostructure investiguée est identifiée comme jouant le rôle de thèse principale.

9.3.4 L'évaluation des résultats

La détection de(s) énoncé(s) de thèse principale dans les textes est un cas particulier de classification de données mal équilibrées. On considère un ensemble de données mal balancé ou déséquilibré lorsque les éléments de l'ensemble ne sont pas répartis également entre les classes. Or, parmi toutes les phrases d'un TECHNETIUM, un nombre très restreint correspond à l'énonciation d'une thèse principale, ce qui est disproportionné par rapport aux phrases qui ne sont pas des thèses principales. Par exemple, seulement 110 des 669 résumés du corpus AbstrCT contiennent au moins un énoncé de thèse principale et, parmi ces 110 résumés, la grande majorité ne contient qu'une seule phrase énonçant la thèse principale dans l'ensemble de toutes les phrases du résumé. Ce faisant, on peut considérer cette tâche de détection des énoncés de thèse principale comme très difficile et s'attendre à de relatives faibles performances en termes de précision, de rappel ou de F-mesure. Dans ce contexte, il peut être intéressant de non seulement

⁸⁰ Les structures et macrostructures sont composées d'un certain nombre de phrases qui sont représentées par des vecteurs de valeurs continues. Une manière de représenter les structures et les macrostructures dans le même espace de représentation est d'extraire la moyenne arithmétique des phrases les composant. Une autre approche consiste à représenter directement les phrases, les structures et les macrostructures dans le même espace. Dans le cas de DBOW, procéder à l'entraînement du modèle avec les phrases, les structures et les macrostructures pourrait biaiser le modèle. La raison est que l'entraînement DBOW est appliqué sur tout le corpus qui autrement serait trop petit sans tous les résumés (même ceux ne contenant aucune thèse principale). Or, les textes correspondant aux résumés sans thèse principale ne sont pas présents dans le corpus (contrairement aux résumés avec thèse principale). À l'inverse, le modèle de représentation « text-embedding-ada-002 » est déjà entraîné de sorte qu'il est possible de représenter des phrases, des structures et des macrostructures sans risque de biaiser la résultante.

évaluer la performance brute de la détection, mais aussi d'évaluer comment les diverses méthodes investiguées permettent de restreindre l'espace de recherche. Par exemple, on peut se demander quelle est la performance d'une méthode à identifier un certain nombre de candidats parmi lesquels se trouvent le(s) énoncé(s) de thèse principale. La mesure d'évaluation top-k, qui est notamment utilisée lors de tâches comme la reconnaissance d'entités nommées, la désambiguïsation d'entités, le classement, etc., permet d'évaluer si la véritable réponse (ou le véritable cas) se trouve parmi les k meilleures prédictions (ou les k cas les plus probables) produites par un modèle donné (voir Runeson *et al.*, 2007 ; Sun *et al.*, 2011 ; Thung *et al.*, 2013). Dans le contexte de la mesure top-k, on peut calculer une F-mesure d'un modèle sur les k meilleurs prédictions afin d'évaluer à quel point le modèle parvient à détecter les bonnes réponses parmi les prédictions les plus probables. Cette approche est utile lorsque l'espace de réponses possibles est large, que les données sont déséquilibrées, ou qu'il est plus pratique d'évaluer un modèle sur sa capacité à produire la bonne réponse parmi un certain nombre de prédictions plutôt que de simplement déterminer si la première prédiction est correcte ou non. Ainsi, les mesures classiques de précision, de rappel et de F-mesure peuvent être adaptées pour les prédictions top-k. La précision top-k divise le nombre d'entités pertinentes correctement prédites par le nombre total de meilleures prédictions (le nombre k multiplié par le nombre de textes dans notre cas):

$$\text{Précision top } k = \frac{\text{Vrais positifs parmi les prédictions}}{\text{Total de prédictions}}$$

Le rappel top-k divise le nombre d'entités pertinentes correctement prédites par le nombre total d'entités pertinentes dans le jeu de données :

$$\text{Rappel top } k = \frac{\text{Vrais positifs parmi les prédictions}}{\text{Total d'entités pertinentes}}$$

La F-mesure top-k est une adaptation de la F-mesure classique aux définitions précédentes :

$$\text{F-mesure top } k = 2 \times \left(\frac{\text{Précision top } k \times \text{Rappel top } k}{\text{Précision top } k + \text{Rappel top } k} \right)$$

Étant donné la constitution d'AbstRCT+⁸¹ et le fonctionnement de nos modèles de détection⁸², un modèle parfait prédira 110 thèses principales lorsque $k=1$, 121 thèses principales lorsque $k=2$ et 123 thèses principales lorsque $k=3$. En termes de précision, ce modèle parfait débutera avec une précision de 1 (110/110) pour $k=1$, puis 0,55 (121/220) pour $k=2$, pour ensuite se dégrader de manière linéaire en fonction du nombre de k . En termes de rappel, le modèle parfait débutera avec un rappel de $\sim 0,89$ (110/123) pour $k=1$, puis $\sim 0,98$ (121/123) pour $k=2$, pour ensuite plafonner à 1 (123/123) pour toutes les autres valeurs de k . Par conséquent, le modèle parfait débutera avec une F1 de $\sim 0,94$ pour $k=1$, puis $\sim 0,71$ pour $k=2$, pour ensuite se dégrader de manière linéaire en fonction du nombre de k . Le tableau 9.4 suivant montre les valeurs de précision top- k (P- tk), de rappel top- k (R- tk) et de F-mesure top- k (F1- tk) d'un modèle parfait. Dans ce tableau, « vp » correspond au nombre de cas vrais positifs. Notons que ce format tabloïde sera repris lors de la présentation des résultats.

Tableau 9.4 Précision top- k (P- tk), rappel top- k (R- tk) et F-mesure top- k (F1- tk) d'un modèle parfait

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	110	1	0,894	0,944
2	121	0,550	0,984	0,706
3	123	0,373	1	0,543
4	123	0,280	1	0,437
5	123	0,224	1	0,366

Lors de la présentation des modèles de détection des thèses principales, spécifions que plus les mesures de performance sont similaires à celles d'un modèle parfait, plus le modèle évalué est performant. À l'inverse, moins les mesures de performances d'un modèle sont similaires à celles d'un modèle parfait, moins celui-ci est performant. Deux (2) interprétations de ces mesures sont particulièrement informatives pour mettre en lumière cette performance. Premièrement, on peut comparer la valeur de précision top- k lorsque $k=1$ par rapport au modèle parfait, et prendre note du comportement subséquent en fonction du nombre k . Si les valeurs de précision top- k sont décroissantes en fonction de k , cela indique que les prédictions sont ordonnées selon la vraisemblance d'être une thèse principale en fonction de k . Deuxièmement, on peut noter le nombre k pour lequel la valeur de rappel top- k est maximale. Cette

⁸¹ 123 thèses principales réparties dans 110 textes, soit 123 énoncés de thèse principale dans 110 résumés et 123 énoncés de thèse principale dans 110 discussions.

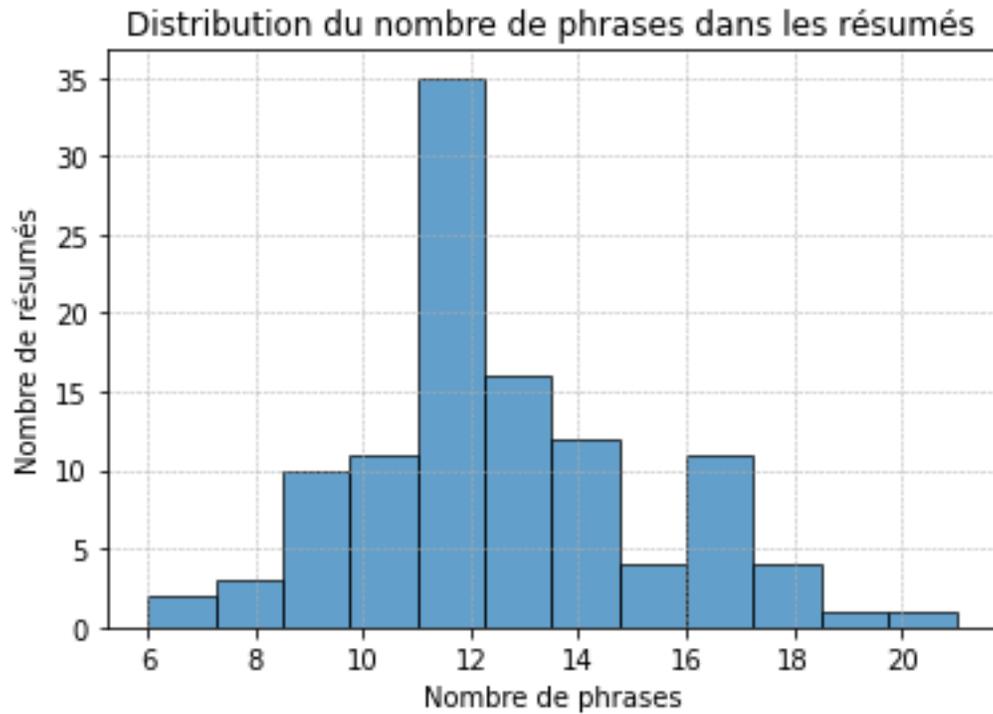
⁸² Les modèles proposés font k prédictions pour chacun des 110 textes, soit k prédictions pour chaque résumé et k prédictions pour chaque discussion.

information montre le nombre de prédictions par résumé ou par discussion qui sont nécessaires pour trouver le nombre maximal d'énoncés de thèse principale. La raison pour laquelle ces mesures top-k sont particulièrement intéressantes est que notre objectif est de réduire l'espace de recherche. Autrement dit, plutôt que de lire l'entièreté d'un résumé ou d'une discussion pour identifier les énoncés de thèse principale, on se demande s'il est possible de trouver les énoncés de thèse principale parmi un petit nombre k d'énoncés, et est-ce que ces suggestions peuvent être ordonnées par vraisemblance? En effet, puisque l'ensemble P_{k+1} des prédictions à $k+1$ inclus toujours l'ensemble P_k des prédictions à k tel que $P_k \subset P_{k+1}$, alors on peut interpréter la précision $k+1$ comme une moyenne arithmétique pondérée entre la précision de l'ensemble P_k et la précision de la disjonction entre l'ensemble P_k et l'ensemble P_{k+1} . Autrement dit, lorsque la précision décroît en fonction de k , les prédictions qui sont ajoutées à $k+1$ ont moins de chances d'être des énoncés de thèse principale, ce qui permet d'ordonner les prédictions par vraisemblance. Enfin, la F-mesure top-k résume les mesures précédentes, car elle constitue la moyenne harmonique de la précision top-k et du rappel top-k.

9.4 Les résultats

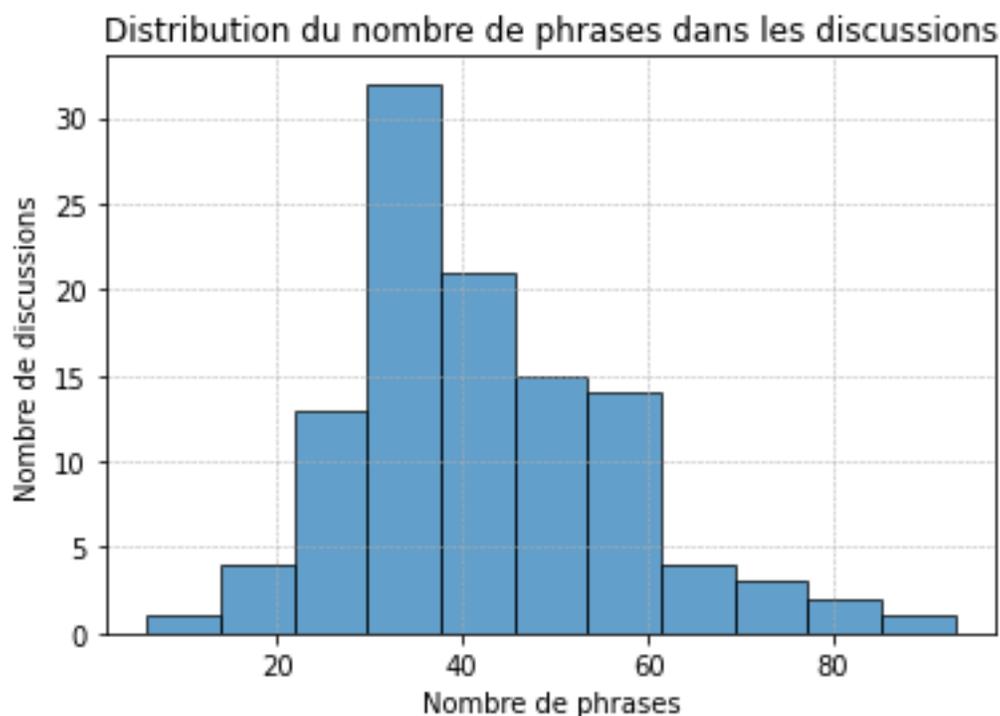
Dans le corpus AbstrCT+, 110 des 669 résumés (~16%) contiennent au moins une thèse principale. Plus précisément, 99 résumés contiennent une seule thèse principale, 9 en contiennent 2 et 2 résumés contiennent 3 thèses principales. Au total, les 110 résumés contiennent 123 thèses principales. De plus, les 110 résumés contiennent 1 370 phrases, soit un minimum de 6 phrases, un maximum de 21 phrases et une moyenne de ~12,4 phrases. À la figure 9.1 suivante, on voit le nombre de phrases contenues dans chacun des 110 résumés.

Figure 9.1 Nombre de phrases contenues dans les résumés



Chacun des 110 résumés comportant au moins une thèse principale est accompagné d'une macrostructure de discussion. Les 110 discussions contiennent 4 612 phrases, soit un minimum de 6 phrases, un maximum de 93 phrases et une moyenne de ~41,9 phrases. À la figure 9.2 suivante, on voit le nombre de phrases contenues dans les discussions.

Figure 9.2 Nombre de phrases contenues dans les discussions



9.4.1 Les résultats de l'identification de la thèse principale dans le corps du texte à l'aide de l'énonciation de la thèse principale dans le résumé

Pour chacune des 110 discussions, nous avons identifié les énoncés de thèse principale à l'aide d'une mesure de similarité entre les thèses principales déjà identifiées dans les résumés et toutes les phrases de la discussion. L'énoncé de la discussion ayant la mesure de similarité cosinus maximale est identifié comme thèse principale. Afin d'évaluer cette méthode, les thèses principales de la section du résumé ainsi que celles de la section discussion sont présentées pour 12 articles (~10% des 110 résumés). Les 12 articles choisis au hasard sont explorés de manière à justifier l'usage de la similarité cosinus dans l'identification des thèses principales.⁸³ En appendice A sont présentées toutes les thèses principales identifiées à partir de la mesure maximale cosinus entre les phrases provenant de discussions et les thèses principales identifiées dans les résumés.

⁸³ Les cas évalués proviennent du sous-ensemble de textes dont le résumé comporte au moins une thèse principale. Ils correspondent aux quatre (4) premiers et quatre (4) derniers textes de ce sous-ensemble. Est ajoutée une série de quatre (4) textes prise au hasard pour un total de douze (12) textes.

La thèse principale identifiée dans le résumé du premier article est « A combination of mitoxantrone plus prednisone is preferable to prednisone alone for reduction of pain in men with metastatic, hormone-resistant, prostate cancer ». La phrase de la section discussion la plus similaire à la thèse principale est « The Canadian study of mitoxantrone plus prednisone versus prednisone alone for men with symptomatic hormone resistant prostate cancer has provided a valuable database for analyzing HQL during treatment » (Osoba *et al.*, 1999).⁸⁴ Malheureusement, cette phrase semble exprimer une thèse subordonnée.

La thèse principale identifiée dans le résumé du second article est « In nonrandomized studies involving patients with different tumor types including non-small-cell lung cancer (NSCLC), ATP infusion appeared to inhibit loss of weight and deterioration of quality of life (QOL) and performance status ». La phrase de la section discussion la plus similaire à la thèse principale est « In conclusion, our results demonstrate that ATP infusion has marked beneficial effects on the QOL of patients with advanced NSCLC » (Agteresch *et al.*, 2000).⁸⁵ Cette phrase semble effectivement exprimer une thèse principale.

La thèse principale identifiée dans le résumé du troisième article est « Oral carbonic anhydrase inhibitors used to treat glaucoma have significant systemic side effects ». La phrase de la section discussion la plus similaire à celle exprimant la thèse principale est « In addition, CA inhibition was below the level known to cause side effects, thus demonstrating the systemic safety of topical brinzolamide » (March et Ochsner, 2000).⁸⁶ Bien que cette dernière exprime effectivement une thèse principale soutenue par les auteurs, précisons que celle-ci diffère de celle exprimée par la phrase identifiée par les annotateurs du corpus AbstrCT. La thèse principale identifiée par les annotateurs joue le rôle de questionnement ou de motivation pour l'étude, c.-à-d. que l'étude clinique d'un inhibiteur de l'anhydrase carbonique (le brinzolamide) pour le traitement du glaucome est justifiée par le manque de traitement sans effet secondaire. L'hypothèse sous-jacente est que le brinzolamide produit moins d'effets secondaires systémiques par rapport à d'autres traitements, ce qui est effectivement démontré par cette étude clinique.

La thèse principale identifiée dans le résumé du quatrième article est « Brimonidine is a viable alternative to timolol for first-line therapy in glaucoma and ocular hypertension ». La phrase de la section discussion

⁸⁴ HCL est l'acronyme de *health-related quality of life*.

⁸⁵ ATP est l'acronyme de adenosine 5'-triphosphate.

⁸⁶ CA est l'acronyme de *carbonic anhydrase*.

la plus similaire à la thèse principale est « The findings of this 4-month study demonstrate that the clinical effectiveness of brimonidine 0.2% twice daily is equivalent to that of timolol 0.5% twice daily as a first-line therapy for glaucoma and ocular hypertension and has less chronotropic effect on the heart » (Javitt et Schiffman, 2000). Notons que cette phrase semble exprimer une thèse principale.

La thèse principale identifiée dans le résumé du cinquième article est « The universality of effects across levels of expectation indicates a potential to utilize this stress reduction approach as complementary therapy in oncologic practice ». La phrase de la section discussion la plus similaire à la thèse principale est « This finding has promising implications for wide applicability in clinical practice because even if a patient might be doubtful of the benefits of stress reduction techniques, a strong potential for positive psychosocial outcome still exists » (Henderson *et al.*, 2012). Cette phrase semble constituer une thèse principale. Elle réfère au segment de texte qui la précède où les auteurs expliquent que l'anticipation d'un effet positif à la suite du traitement est fortement corrélée à de meilleurs résultats pour les approches concurrentes. Leur conclusion est que ce n'est pas le cas avec leur approche.

La thèse principale identifiée dans le résumé du sixième article est « Both procedures seem acceptable as standard reconstructions after distal gastrectomy with regard to postoperative QOL and dysfunction ». La phrase de la section discussion la plus similaire à la thèse principale est « We found no significant difference between R-Y and B-I procedures in terms of overall postoperative dysfunction » (Takiguchi *et al.*, 2012).⁸⁷ Cette phrase semble exprimer une thèse principale.

La thèse principale identifiée dans le résumé du septième article est « Lifestyle intervention research is feasible in a cooperative group setting ». La phrase de la section discussion la plus similaire à la thèse principale est « Our study also demonstrated the feasibility of conducting lifestyle research in a cooperative group setting » (Ligibel *et al.*, 2012). Cette phrase semble effectivement exprimer une thèse principale.

La thèse principale identifiée dans le résumé du huitième article est « Multimodality therapies for cachexia ideally should be introduced within a context of "best supportive care" that includes optimal symptom management and careful psychosocial counseling ». La phrase de la section discussion la plus similaire à

⁸⁷ B-I et R-Y sont les acronymes respectifs de Billroth I et Roux-en-Y, lesquels sont des reconstructions suivant une gastrectomie distale.

la thèse principale est « Our study also demonstrates that multimodality therapies for cachexia must not have significant side effects and should ideally be introduced within a context of the “best supportive care”, which includes optimal symptom management and careful psychosocial counseling » (Macciò *et al.*, 2012). Notons que cette phrase semble exprimer une thèse principale.

La thèse principale identifiée dans le résumé du neuvième article est « The choice of reconstruction after gastrectomy and the significance of remaining reservoir function is a matter of controversy ». La phrase de la section discussion la plus similaire à la thèse principale est « We have found the gastric reservoir function in partially gastrectomized patients to be of identical significance and magnitude as in patients who had a gastric substitute after total gastrectomy » (Svedlund *et al.*, 1997). Spécifions que la phrase sélectionnée par les annotateurs comme exprimant la thèse principale joue plutôt le rôle d'un questionnement à laquelle l'étude clinique tente de répondre. La phrase la plus similaire dans la section discussion est effectivement une thèse importante selon laquelle les effets sur la fonction de réservoir gastrique des traitements étudiés sont similaires.⁸⁸

La thèse principale identifiée dans le résumé du dixième article est « The use of the Pain Education Program by nurses should be seriously considered on oncology units ». La phrase de la section discussion la plus similaire à la thèse principale est « Because of the pain relieving effects, the use of such an intervention should be seriously considered on oncology units » (de Wit *et al.*, 1997). Cette phrase est composée d'une thèse principale normative en ce que le protocole d'intervention faisant l'objet d'une étude clinique est dit devoir être sérieusement considéré en raison des effets démontrés en termes de réduction de la souffrance.⁸⁹

La thèse principale identifiée dans le résumé du onzième article est « No treatment-related death was observed ». La phrase de la section discussion la plus similaire à la thèse principale est « The response rate and survival observed do not allow us to observe an increase in anthracycline cytotoxicity by Ionidamine

⁸⁸ Notons que les auteurs étudient également d'autres effets ayant un impact sur la qualité de vie, notamment les troubles de l'alimentation, indigestion et diarrhée. À ce propos, les auteurs privilégient une procédure de gastrectomie subtotale lorsque cela est possible, ce qui constitue une autre thèse importante.

⁸⁹ Spécifions que le contenu de la thèse principale est explicité plus en détail dans les phrases subséquentes de cette section. Par exemple, la phrase « Therefore, nurses should be trained as pain counselors and the Pain Education Program should be implemented and evaluated in nursing practice » explicite comment les auteurs conçoivent « considérer sérieusement » le protocole d'intervention, notamment au moyen de l'enseignement et de l'évaluation du personnel infirmier.

as occurs in breast carcinoma » (Barone *et al.*, 1998). La phrase de la section discussion la plus similaire à la phrase sélectionnée par les annotateurs est effectivement une des thèses principales soutenues par les auteurs. Une de leur hypothèse de départ est qu'une nouvelle combinaison de substances (etoposide, epirebicine, cisplatine et lonidamine) est efficace dans le traitement du carcinome gastrique et que la toxicité de ce traitement est minimale. Leur étude montre que ce traitement, relativement efficace, ne se révèle pas toxique (contrairement à ce qu'on observe pour le traitement d'autres types de cancer), mais que cette absence de toxicité peut s'expliquer par un taux de réponse au traitement ainsi qu'un taux de survie insuffisant.

La thèse principale identifiée dans le résumé du douzième article est « Thus 'Casodex' 150 mg monotherapy is an option for patients with M1 prostate cancer for whom surgical or medical castration is not indicated or is not acceptable ». La phrase de la section discussion la plus similaire à la thèse principale est « This supports the choice of 'Casodex' 150 mg monotherapy as an option for patients with M1 prostate cancer for whom surgical or medical castration is not indicated or is not acceptable » (Tyrrell *et al.*, 1998). Cette phrase semble effectivement exprimer une thèse principale.

En somme, la méthode d'identification des énoncés de thèse principale dans la discussion à partir de l'énoncé de la thèse principale dans le résumé semble donner de bons résultats. Les exemples présentés indiquent que la section discussion est un emplacement privilégié d'expression de la thèse principale puisque cette macrostructure semble contenir l'expression d'au moins une thèse principale, du moins pour 11 des 12 exemples présentés. L'exploration de ces cas montre que les articles du corpus AbstrCT+ peuvent contenir plus d'une thèse principale, mais que toutes les thèses principales ne sont pas nécessairement exprimées dans le résumé. Or, la présence d'une thèse principale dans le résumé n'implique pas nécessairement un même nombre d'énoncés de thèse principale dans la discussion. D'une part, s'il est possible qu'une même thèse principale soit énoncée de plusieurs manières dans la discussion, cela est moins probable dans le résumé puisque le format de ce dernier est habituellement restreint. D'autre part, l'investigation de la question de recherche peut mener à plusieurs solutions et, ce faisant, aboutir à plusieurs thèses principales. Néanmoins, si les thèses principales exprimées dans le résumé constituent un sous-ensemble de celles exprimées dans le corps du texte ou dans la discussion, alors on peut concevoir l'expression de ce sous-ensemble dans le résumé comme l'indice d'un rôle plus important par rapport aux autres thèses principales. Autrement, les cas évalués montrent que les annotateurs du corpus AbstrCT considèrent parfois les questions ou les motivations de recherche exprimées dans le

résumé comme étant des thèses principales. Enfin, la similitude cosinus semble être une mesure efficace d'identification de l'énoncé d'une thèse principale dans la section discussion à partir de l'énoncé d'une question de recherche ou d'une thèse principale dans le résumé.

9.4.2 Les résultats de l'identification de la thèse principale à l'aide de propriétés syntaxiques ou sémantiques

Afin de favoriser l'évaluation de nos résultats, nous procédons d'abord à l'identification d'un niveau de base pour comparaison ultérieure. Ce dernier correspond à la moyenne des résultats de l'application multiple d'un modèle aléatoire simple sur nos données. Ce modèle fait aléatoirement des prédictions pour chaque texte comportant au moins une thèse principale en fonction du nombre k de prédictions. Autrement dit, les prédictions du modèle de base sont faites au hasard. Les valeurs k investiguées vont de 1 à 5. Le niveau de base est constitué de la moyenne des valeurs de précision top- k (P - tk), de rappel top- k (R - tk) et de F-mesure top- k ($F1$ - tk) qui sont calculées à chaque application du modèle sur les données. Le nombre total d'applications du modèle aléatoire sur nos données est de 1 000. Le tableau 9.5 suivant montre le niveau de base pour l'identification de l'énoncé de thèse principale dans le résumé. Dans ce tableau, « vp » correspond à la moyenne des cas vrais positifs identifiés lors des 1 000 applications du modèle aléatoire sur les données.

Tableau 9.5 Précision top- k (P - tk), rappel top- k (R - tk) et F-mesure top- k ($F1$ - tk) d'un modèle de base pour l'identification de l'énoncé de thèse principale dans le résumé

k	vp	P - tk	R - tk	$F1$ - tk
1	10,19	0,093	0,083	0,087
2	19,54	0,089	0,159	0,114
3	28,18	0,085	0,229	0,124
4	35,70	0,081	0,290	0,127
5	42,77	0,078	0,348	0,127

On observe une grande différence entre les résultats du modèle de base et le modèle parfait précédemment présenté. Alors que le rappel maximal de 1 était atteint à $k=3$, le modèle de base atteint un rappel de $\sim 0,35$ à $k=5$, c.-à-d. que seulement 35% des énoncés de thèse principale sont identifiés. La précision pour les valeurs k de 1 à 5 décroît de 0,09 à 0,08, c.-à-d. qu'entre 8% et 9% de ce qui est prédit comme thèse principale est réellement un énoncé de thèse principal. Dans le tableau 9.6 suivant, on voit le niveau de base pour l'identification de l'énoncé de thèse principale dans la discussion.

Tableau 9.6 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) d'un modèle de base pour l'identification de l'énoncé de thèse principale dans la discussion

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	3,31	0,030	0,027	0,029
2	6,65	0,030	0,055	0,039
3	9,50	0,029	0,078	0,042
4	12,44	0,028	0,102	0,044
5	15,44	0,028	0,127	0,046

On remarque que la performance du modèle aléatoire appliqué sur le résumé est supérieure à celle du modèle appliqué sur la discussion. Par exemple, le rappel du modèle de base appliqué sur la discussion à $k=5$ est de $\sim 0,13$, c.-à-d. que seulement 13% des énoncés de thèse principale sont identifiés. La précision pour les valeurs k de 1 à 5 décroît de 0,03 à 0,028, c.-à-d. qu'entre 2,8% et 3% de ce qui est prédit comme thèse principale est réellement un énoncé de thèse principal. Évidemment, les chances de trouver une thèse principale par hasard dans le résumé sont supérieures à celles de trouver une thèse principale par hasard dans la discussion, car le nombre d'énoncés de thèse principale est constant entre les résumés et les discussions, alors que la moyenne du nombre de phrases dans le résumé est inférieure à la moyenne du nombre de phrases dans la discussion.

9.4.2.1 Les résultats de l'identification de la thèse principale à l'aide du titre

Dans cette section, nous présentons les résultats de la détection de l'énoncé de la thèse principale à l'aide d'une mesure de similarité maximale entre le titre et les phrases de la macrostructure dans laquelle la thèse principale est recherchée. Les modèles de représentation textuelle utilisés sont DBOW et « text-embedding-ada-002 ». Les macrostructures dans lesquelles les thèses principales sont recherchées sont le résumé et la discussion.

9.4.2.1.1 Les résultats des modèles fondés par le titre appliqués au résumé

Le tableau 9.7 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen du titre et de la représentation textuelle DBOW (R_T_DBOW). La performance du modèle R_T_DBOW est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$). La similarité cosinus est calculée entre la représentation vectorielle du titre et celle de chacune des phrases du résumé. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales.

Tableau 9.7 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_T_DBOW

<i>k</i>	<i>vp</i>	<i>P-tk</i>	<i>R-tk</i>	<i>F1-tk</i>
1	11	0,100	0,089	0,094
2	32	0,145	0,260	0,187
3	50	0,152	0,407	0,221
4	61	0,139	0,496	0,217
5	73	0,133	0,593	0,217

On remarque que la performance de R_T_DBOW est à peine supérieure au niveau de base lorsque $k=1$. Par exemple, la précision est $\sim 1,08$ fois supérieure au modèle de base à $k=1$. Cependant, la précision est croissante pour $k=2$ et $k=3$, puis décroissante pour $k=4$ et $k=5$. Par conséquent, la valeur k n'est pas l'indice d'un ordre de vraisemblance pour les prédictions. Néanmoins, la performance de R_T_DBOW est largement plus performante que le niveau de base pour les valeurs supérieures de k (2 à 5), notamment le rappel qui se termine avec une performance $\sim 1,70$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,221$, est observée lorsque $k=3$. Bref, les premières prédictions du modèle ne sont pas très précises, mais cela s'améliore lorsque k augmente pour éventuellement aboutir à une valeur de rappel intéressante.

Le tableau 9.8 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen du titre et de la représentation textuelle d'OpenAI (R_T_OAI). La performance du modèle R_T_OAI est présentée en termes de précision top-k (*P-tk*), rappel top-k (*R-tk*) et F-mesure top-k (*F1-tk*). La similarité cosinus est calculée entre la représentation vectorielle du titre et celle de chacune des phrases du résumé. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales. Par rapport au niveau de base, la performance F1 de R_T_OAI est $\sim 1,18$ fois supérieure lorsque $k=1$.

Tableau 9.8 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_T_OAI

<i>k</i>	<i>vp</i>	<i>P-tk</i>	<i>R-tk</i>	<i>F1-tk</i>
1	12	0,109	0,098	0,103
2	23	0,105	0,187	0,134
3	40	0,121	0,325	0,177
4	62	0,141	0,504	0,220
5	72	0,131	0,585	0,214

La performance de R_T_OAI est légèrement supérieure au niveau de base lorsque $k=1$, notamment la précision qui est $\sim 1,17$ fois supérieure au modèle de base à $k=1$. Cependant, la précision décroît à $k=2$, puis est croissante pour $k=3$ et $k=4$, pour enfin être décroissante à $k=5$. Par conséquent, la valeur k n'est pas l'indice d'un ordre de vraisemblance pour les prédictions. Néanmoins, la performance de R_T_OAI est significativement plus performante que le niveau de base pour les valeurs supérieures de k (3 à 5), notamment le rappel qui se termine avec une performance $\sim 1,68$ fois supérieure au modèle de base lorsque $k=5$. La valeur maximale de la F-mesure, atteignant $\sim 0,220$, est observée lorsque $k=4$. Bref, ce modèle se comporte de manière similaire à la version DBOW.

9.4.2.1.2 Les résultats des modèles fondés par le titre appliqués à la discussion

Le tableau 9.9 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen du titre et de la représentation textuelle DBOW (D_T_DBOW). La performance du modèle D_T_DBOW est présentée en termes de précision top- k ($P-tk$), rappel top- k ($R-tk$) et F-mesure top- k ($F1-tk$). La similarité cosinus est calculée entre la représentation vectorielle du titre et celle de chacune des phrases de la discussion. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales.

Tableau 9.9 Précision top- k ($P-tk$), rappel top- k ($R-tk$) et F-mesure top- k ($F1-tk$) du modèle D_T_DBOW

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	21	0,191	0,172	0,181
2	37	0,168	0,303	0,216
3	45	0,136	0,369	0,199
4	51	0,116	0,418	0,181
5	55	0,100	0,451	0,164

On remarque que la performance de D_T_DBOW est globalement supérieure au niveau de base. La précision est $\sim 6,37$ fois supérieure à $k=1$. Cette mesure est décroissante en fonction des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions en fonction de k . Ce modèle termine avec une valeur de rappel $\sim 3,55$ fois supérieure au modèle de base lorsque $k=5$. La valeur maximale de la F-mesure, atteignant $\sim 0,216$, est observée lorsque $k=2$. En bref, ce modèle performe significativement mieux que le hasard.

Le tableau 9.10 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen du titre et de la représentation textuelle d'OpenAI (D_T_OAI). La performance du modèle D_T_OAI est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$). La similarité cosinus est calculée entre la représentation vectorielle du titre et celle de chacune des phrases de la discussion. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales.

Tableau 9.10 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) du modèle D_T_OAI

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	14	0,127	0,115	0,121
2	26	0,118	0,213	0,152
3	30	0,091	0,246	0,133
4	40	0,091	0,328	0,142
5	48	0,087	0,393	0,143

On remarque que la performance de D_T_OAI est globalement supérieure au niveau de base. La précision est $\sim 4,23$ fois supérieure au modèle de base à $k=1$. Les valeurs de cette mesure sont décroissantes en fonction des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions en fonction de k . Ce modèle termine avec une valeur de rappel $\sim 3,09$ fois supérieure au modèle de base lorsque $k=5$. La valeur maximale de la F-mesure, atteignant $\sim 0,152$, est observée lorsque $k=2$. Bref, ce modèle se comporte de manière similaire à la version DBOW, mais est significativement moins performant.

En somme, tous les modèles de détection de la thèse principale à l'aide du titre performant mieux ou légèrement mieux que le niveau de base. On remarque que les meilleures performances s'obtiennent lorsque la détection de la thèse principale s'effectue sur la représentation DBOW plutôt que OpenAI. Enfin, les modèles dont la performance F1 se démarque le plus sont R_T_DBOW ($\sim 0,221$ à $k=3$) pour le résumé et D_T_DBOW ($\sim 0,216$ à $k=2$) pour la discussion.

9.4.2.2 Les résultats de l'identification de la thèse principale à l'aide du résumé

Dans cette section, nous présentons les résultats de la détection de l'énoncé de la thèse principale à l'aide d'une mesure de similarité maximale entre le résumé et chacune des phrases de la macrostructure dans laquelle la thèse principale est recherchée. Les méthodes de représentation textuelle utilisées sont DBOW et OpenAI. Lorsque la représentation textuelle est DBOW, le résumé est représenté par un vecteur faisant

la moyenne des vecteurs qui représentent une à une les phrases du résumé. Lorsque la représentation est « text-embedding-ada-002 », le résumé est représenté par un vecteur propre. Les macrostructures dans lesquelles les thèses principales sont recherchées sont le résumé et la discussion.

9.4.2.2.1 Les résultats des modèles fondés par le résumé appliqués au résumé

Le tableau 9.11 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen du résumé et de la représentation textuelle DBOW (R_R_DBOW). La performance du modèle R_R_DBOW est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$). La similarité cosinus est calculée entre la représentation vectorielle du résumé et celle de chacune des phrases du résumé. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales.

Tableau 9.11 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) du modèle R_R_DBOW

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	6	0,055	0,049	0,052
2	12	0,055	0,098	0,070
3	19	0,058	0,154	0,084
4	26	0,059	0,211	0,092
5	38	0,069	0,309	0,113

On note que la performance du modèle R_R_DBOW est globalement inférieure à celle du modèle de base. Les valeurs de précision, de rappel et de F-mesure sont inférieures à celles du niveau de base pour toutes les valeurs de k . Puisque R_R_DBOW utilise la similarité maximale entre les vecteurs de phrase du résumé et la moyenne de ces vecteurs comme prédicteur de thèse principale, et que cette approche fait moins bien que le hasard, on peut inférer que les énoncés de thèse principale ne sont pas les plus similaires à la moyenne vectorielle du résumé et qui correspond au centre statistique du résumé dans l'espace de caractéristiques. Autrement dit, les énoncés de thèse principale occupent un positionnement qui diverge du centre statistique du résumé dans l'espace de représentation DBOW. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,113$, est observée lorsque $k=5$.

Le tableau 9.12 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen du résumé et de la représentation textuelle d'OpenAI (R_R_OAI). La performance du modèle R_R_OAI est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et

F-mesure top-k ($F1-tk$). La similarité cosinus est calculée entre la représentation vectorielle du résumé et celle de chacune des phrases du résumé. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales.

Tableau 9.12 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) du modèle R_R_OAI

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	21	0,191	0,171	0,180
2	42	0,191	0,341	0,245
3	58	0,176	0,472	0,256
4	75	0,170	0,610	0,266
5	84	0,153	0,683	0,250

Le modèle R_R_OAI est globalement plus performant que le modèle de base. Ce modèle est $\sim 2,05$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes en fonction de l'augmentation des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions de ce modèle. Ce modèle termine avec une valeur de rappel $\sim 1,96$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,266$, est observée lorsque $k=4$. En bref, ce modèle se comporte très différemment de la version DBOW. Cela indique que les énoncés de thèse principale occupent un positionnement proximal au centre du résumé dans l'espace de représentation « text-embedding-ada-002 ».

9.4.2.2.2 Les résultats des modèles fondés par le résumé appliqués à la discussion

Le tableau 9.13 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen du résumé et de la représentation textuelle DBOW (D_R_DBOW). La performance du modèle D_R_DBOW est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$). La similarité cosinus est calculée entre la représentation vectorielle du résumé et celle de chacune des phrases de la discussion. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales.

Tableau 9.13 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_R_DBOW

<i>k</i>	<i>vp</i>	<i>P-tk</i>	<i>R-tk</i>	<i>F1-tk</i>
1	13	0,118	0,107	0,112
2	24	0,109	0,197	0,140
3	33	0,100	0,270	0,146
4	37	0,084	0,303	0,132
5	51	0,093	0,418	0,152

Le modèle D_R_DBOW est globalement plus performant que le modèle de base. Ce modèle est $\sim 3,93$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes en fonction de l'augmentation des valeurs de k à l'exception de la dernière valeur où $k=5$, ce qui indique un ordre descendant de vraisemblance pour les prédictions de ce modèle, mais seulement avant $k=5$. Ce modèle termine avec une valeur de rappel $\sim 3,29$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,146$, est observée lorsque $k=3$.

Le tableau 9.14 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen du résumé et de la représentation textuelle d'OpenAI (D_R_OAI). La performance du modèle D_R_OAI est présentée en termes de précision top-k (*P-tk*), rappel top-k (*R-tk*) et F-mesure top-k (*F1-tk*). La similarité cosinus est calculée entre la représentation vectorielle du résumé et celle de chacune des phrases de la discussion. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales. Par rapport au niveau de base, la performance F1 de D_R_OAI est $\sim 5,14$ fois supérieure lorsque $k=1$.

Tableau 9.14 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_R_OAI

<i>k</i>	<i>vp</i>	<i>P-tk</i>	<i>R-tk</i>	<i>F1-tk</i>
1	17	0,155	0,139	0,147
2	28	0,127	0,230	0,164
3	42	0,127	0,344	0,186
4	45	0,102	0,369	0,160
5	52	0,095	0,426	0,155

Le modèle D_R_OAI est globalement plus performant que le modèle de base. Ce modèle est $\sim 5,17$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes en fonction de l'augmentation des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions

de ce modèle. Ce modèle termine avec une valeur de rappel $\sim 3,35$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,186$, est observée lorsque $k=3$.

Outre la performance du modèle de détection des thèses principales dans le résumé à l'aide d'une représentation vectorielle du résumé de type DBOW (R_R_DBOW), tous les modèles performant mieux ou légèrement mieux que le niveau de base. Globalement, les performances des modèles de détection à l'aide du résumé sont meilleures avec la représentation « text-embedding-ada-002 » comparativement à DBOW. Enfin, les modèles dont la performance F1 se démarque le plus sont R_R_OAI ($\sim 0,266$ à $k=4$) pour le résumé et D_R_OAI ($\sim 0,186$ à $k=3$) pour la discussion.

9.4.2.3 Les résultats de l'identification de la thèse principale à l'aide de l'introduction

Dans cette section, nous présentons les résultats de la détection de l'énoncé de la thèse principale à l'aide d'une mesure de similarité maximale entre l'introduction et chacune des phrases de la macrostructure dans laquelle la thèse principale est recherchée. Les méthodes de représentation textuelle utilisées sont DBOW et « text-embedding-ada-002 ». Lorsque la représentation textuelle est DBOW, l'introduction est représentée par un vecteur faisant la moyenne des vecteurs qui représentent une à une les phrases de l'introduction. Lorsque la représentation est « text-embedding-ada-002 », l'introduction est représentée par un vecteur propre à l'introduction. Les macrostructures dans lesquelles les thèses principales sont recherchées sont le résumé et la discussion.

9.4.2.3.1 Les résultats des modèles fondés par l'introduction appliqués au résumé

Le tableau 9.15 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen de l'introduction et de la représentation textuelle DBOW (R_I_DBOW). La performance du modèle R_I_DBOW est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$). La similarité cosinus est calculée entre la représentation vectorielle de l'introduction et celle de chacune des phrases du résumé. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales.

Tableau 9.15 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) du modèle R_I_DBOW

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	35	0,318	0,285	0,300
2	60	0,273	0,488	0,350
3	74	0,224	0,602	0,327
4	83	0,189	0,675	0,295
5	93	0,169	0,756	0,276

On note que la performance du modèle R_I_DBOW est globalement supérieure à celle du modèle de base. Ce modèle est $\sim 3,42$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes en fonction de l'augmentation des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions de ce modèle. Ce modèle termine avec une valeur de rappel $\sim 2,17$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,35$, est observée lorsque $k=2$. En bref, ce modèle performe mieux que le hasard.

Le tableau 9.16 suivant montre la performance du modèle de détection de l'énoncé des énoncés de thèse principale identifiés dans le résumé au moyen de l'introduction et de la représentation textuelle d'OpenAI (R_I_OAI). La performance du modèle R_I_OAI est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$). La similarité cosinus est calculée entre la représentation vectorielle de l'introduction et celle de chacune des phrases du résumé. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales.

Tableau 9.16 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) du modèle R_I_OAI

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	25	0,227	0,203	0,215
2	51	0,232	0,415	0,297
3	68	0,206	0,553	0,300
4	76	0,173	0,618	0,270
5	83	0,151	0,675	0,247

On observe que la performance du modèle R_I_OAI est globalement supérieure à celle du modèle de base. Ce modèle est $\sim 2,44$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes en fonction de l'augmentation des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions de ce modèle. Ce modèle termine avec une valeur de rappel $\sim 1,94$ fois

supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,30$, est observée lorsque $k=3$. En bref, ce modèle performe moins bien que la version utilisant DBOW, mais performe mieux que le hasard.

9.4.2.3.2 Les résultats des modèles fondés par l'introduction appliqués à la discussion

Le tableau 9.17 suivant montre la performance du modèle de détection de l'énoncé des énoncés de thèse principale identifiés dans la discussion au moyen de l'introduction et de la représentation textuelle DBOW (D_I_DBOW). La performance du modèle D_I_DBOW est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$). La similarité cosinus est calculée entre la représentation vectorielle de l'introduction et celle de chacune des phrases de la discussion. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales.

Tableau 9.17 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) du modèle D_I_DBOW

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	28	0,255	0,230	0,241
2	45	0,205	0,369	0,263
3	51	0,155	0,418	0,226
4	57	0,130	0,467	0,203
5	64	0,116	0,525	0,190

La performance du modèle D_I_DBOW est globalement supérieure à celle du modèle de base. Ce modèle est $\sim 8,50$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes en fonction de l'augmentation des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions de ce modèle. Ce modèle termine avec une valeur de rappel $\sim 3,29$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,263$, est observée lorsque $k=2$. Bref, ce modèle performe mieux que le hasard.

Le tableau 9.18 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen de l'introduction et de la représentation textuelle d'OpenAI (D_I_OAI). La performance du modèle D_I_OAI est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$). La similarité cosinus est calculée entre la représentation vectorielle de l'introduction et celle de chacune des phrases de la discussion. Les k phrases ayant une similarité

maximale sont déterminées comme étant les k meilleurs candidats de thèses principales. Par rapport au niveau de base, la performance F1 de D_I_OAI est $\sim 5,44$ fois supérieure lorsque $k=1$.

Tableau 9.18 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_I_OAI

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	18	0,164	0,148	0,155
2	32	0,145	0,262	0,187
3	45	0,136	0,369	0,199
4	51	0,116	0,418	0,181
5	55	0,100	0,451	0,164

La performance du modèle D_I_OAI est globalement supérieure à celle du modèle de base. Ce modèle est $\sim 5,47$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes en fonction de l'augmentation des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions de ce modèle. Ce modèle termine avec une valeur de rappel $\sim 3,55$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,199$, est observée lorsque $k=3$. En bref, ce modèle performe moins bien que la version utilisant DBOW, mais mieux que le hasard.

En somme, tous les modèles de détection de la thèse principale à l'aide de la macrostructure d'introduction performant significativement mieux que le niveau de base. On remarque que la représentation DBOW offre de meilleures performances comparativement à la représentation d'OpenAI. Enfin, les modèles dont la performance F1 se démarque le plus sont R_I_DBOW ($\sim 0,35$ à $k=2$) pour le résumé et D_I_DBOW ($\sim 0,263$ à $k=2$) pour la discussion.

9.4.2.4 Les résultats de l'identification de la thèse principale à l'aide de la discussion

Dans cette section, nous présentons les résultats de la détection de l'énoncé de la thèse principale à l'aide d'une mesure de similarité maximale entre la discussion et chacune des phrases de la macrostructure dans laquelle la thèse principale est recherchée. Les méthodes de représentation textuelle utilisées sont DBOW et « text-embedding-ada-002 ». Lorsque la représentation textuelle est DBOW, la discussion est représentée par un vecteur faisant la moyenne des vecteurs qui représentent une à une les phrases de la discussion. Lorsque la représentation est « text-embedding-ada-002 », la discussion est représentée par

un vecteur propre à la discussion. Les macrostructures dans lesquelles les thèses principales sont recherchées sont le résumé et la discussion.

9.4.2.4.1 Les résultats des modèles fondés par la discussion appliqués au résumé

Le tableau 9.19 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen de la discussion et de la représentation textuelle DBOW (R_D_DBOW). La performance du modèle R_D_DBOW est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$). La similarité cosinus est calculée entre la représentation vectorielle de la discussion et celle de chacune des phrases du résumé. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales.

Tableau 9.19 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) du modèle R_D_DBOW

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	21	0,191	0,171	0,180
2	35	0,159	0,285	0,204
3	46	0,139	0,374	0,203
4	60	0,136	0,488	0,213
5	67	0,122	0,545	0,199

On remarque que la performance du modèle R_D_DBOW est globalement supérieure à celle du modèle de base. Ce modèle est $\sim 2,05$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes en fonction de l'augmentation des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions de ce modèle. Ce modèle termine avec une valeur de rappel $\sim 1,57$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,213$, est observée lorsque $k=4$. En bref, ce modèle performe mieux que le hasard.

Le tableau 9.20 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen de la discussion et de la représentation textuelle d'OpenAI (R_D_OAI). La performance du modèle R_D_OAI est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$). La similarité cosinus est calculée entre la représentation vectorielle de la discussion et celle de chacune des phrases du résumé. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales.

Tableau 9.20 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_D_OAI

<i>k</i>	<i>vp</i>	<i>P-tk</i>	<i>R-tk</i>	<i>F1-tk</i>
1	9	0,082	0,073	0,077
2	24	0,109	0,195	0,140
3	41	0,124	0,333	0,181
4	58	0,132	0,472	0,206
5	66	0,120	0,537	0,196

La performance du modèle R_D_OAI est inférieure à celle du modèle de base pour $k=1$, puis légèrement supérieure pour les autres valeurs de k testées. Lorsque $k=1$, la précision est multipliée par $\sim 0,88$ par rapport au niveau de base. Outre à $k=5$, les valeurs de précision sont croissantes en fonction de l'augmentation des valeurs de k , ce qui indique un ordre ascendant de vraisemblance pour les prédictions de ce modèle. Ce modèle termine avec une valeur de rappel $\sim 1,54$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,206$, est observée lorsque $k=4$. En bref, ce modèle performe mieux que le hasard lorsque $k>1$.

9.4.2.4.2 Les résultats des modèles fondés par la discussion appliqués à la discussion

Le tableau 9.21 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen de la discussion et de la représentation textuelle DBOW (D_D_DBOW). La performance du modèle D_D_DBOW est présentée en termes de précision top-k (*P-tk*), rappel top-k (*R-tk*) et F-mesure top-k (*F1-tk*). La similarité cosinus est calculée entre la représentation vectorielle de la discussion et celle de chacune des phrases de la discussion. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales.

Tableau 9.21 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle D_D_DBOW

<i>k</i>	<i>vp</i>	<i>P-tk</i>	<i>R-tk</i>	<i>F1-tk</i>
1	8	0,073	0,066	0,069
2	20	0,091	0,164	0,117
3	26	0,079	0,213	0,115
4	33	0,075	0,270	0,117
5	44	0,080	0,361	0,131

On note que la performance du modèle D_D_DBOW est globalement supérieure à celle du modèle de base. Ce modèle est $\sim 2,43$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont parfois

croissantes, parfois décroissantes par rapport à l'augmentation des valeurs de k , ce qui n'indique aucun ordre ascendant ou descendant de vraisemblance pour les prédictions de ce modèle. Ce modèle termine avec une valeur de rappel $\sim 2,84$ fois supérieure au modèle de base lorsque $k=5$. La valeur maximale de la F-mesure, atteignant $\sim 0,131$, est observée lorsque $k=5$. En bref, ce modèle performe mieux que le hasard. Enfin, puisque D_D_DBOW utilise la similarité maximale entre les vecteurs de phrase de la discussion et la moyenne de ces vecteurs comme prédicteur de thèse principale, et que cette approche fait mieux que le hasard, on peut inférer que les énoncés de thèse principale dans la discussion occupent un positionnement relativement proximal au centre statistique de la discussion dans l'espace de caractéristiques. Autrement dit, les énoncés de thèse principale dans la discussion occupent un positionnement relativement proximal au centre statistique de la discussion dans l'espace de représentation DBOW.

Le tableau 9.22 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen de la discussion et de la représentation textuelle d'OpenAI (D_D_OAI). La performance du modèle D_D_OAI est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$). La similarité cosinus est calculée entre la représentation vectorielle de la discussion et celle de chacune des phrases de la discussion. Les k phrases ayant une similarité maximale sont déterminées comme étant les k meilleurs candidats de thèses principales.

Tableau 9.22 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) du modèle D_D_OAI

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	11	0,100	0,090	0,095
2	15	0,068	0,123	0,088
3	27	0,082	0,221	0,119
4	31	0,070	0,254	0,110
5	36	0,065	0,295	0,107

La performance du modèle D_D_OAI est globalement supérieure à celle du modèle de base. Ce modèle est $\sim 3,33$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont parfois croissantes, parfois décroissantes par rapport à l'augmentation des valeurs de k , ce qui n'indique aucun ordre ascendant ou descendant de vraisemblance pour les prédictions de ce modèle. Ce modèle termine avec une valeur de rappel $\sim 2,32$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale

de la F-mesure, atteignant $\sim 0,119$, est observée lorsque $k=3$. En bref, ce modèle performe mieux que le hasard.

Outre la performance du modèle de détection des thèses principales dans le résumé à l'aide de la représentation d'OpenAI (R_D_OAI) qui performe mieux que le hasard lorsque $k>1$, tous les modèles de détection au moyen de la discussion performant mieux ou légèrement mieux que le niveau de base. On remarque que la représentation DBOW offre de meilleures performances comparativement à la représentation « text-embedding-ada-002 » pour la détection de thèses principales dans le résumé. Cela est moins clair dans la discussion, car les meilleures performances n'appartiennent pas aux mêmes représentations en fonction du nombre k . Enfin, les modèles dont la performance F1 se démarque le plus sont R_I_DBOW ($\sim 0,35$ pour $k=2$) pour le résumé et D_I_DBOW ($\sim 0,263$ pour $k=2$) pour la discussion.

Somme toute, la comparaison de tous les modèles testés montre que le modèle le plus performant lors de la détection de thèses principales dans le résumé est R_I_DBOW. Les modèles R_I_OAI, R_T_DBOW ou R_D_DBOW arrivent seconds dépendamment de la mesure de performance et du nombre k . En ne tenant compte que de $k=1$, R_D_DBOW est le second modèle le plus performant pour le résumé. Lors de la détection des thèses principales dans la discussion, les deux (2) modèles les plus performants sont respectivement D_I_DBOW et D_T_DBOW. D'une part, DBOW apparaît être la représentation textuelle la plus performante pour la détection des thèses principales. D'autre part, l'introduction semble être la structure ou la macrostructure la plus informative pour effectuer cette tâche, qu'importe le type de représentation utilisé ou la structure dans laquelle on cherche les énoncés de thèse principale. Enfin, la deuxième structure ou macrostructure la plus informative pour la tâche de détection des thèses principales apparaissent spécifiques au type de représentation ainsi qu'à la structure ou la macrostructure dans laquelle on cherche les énoncés de thèse principale. Pour la recherche de thèses principales dans la structure du résumé, la macrostructure de la discussion semble être la deuxième structure la plus informative. Pour la recherche de thèses principales dans la macrostructure de discussion, la structure du titre est la deuxième structure la plus informative.

9.4.3 Les résultats de l'identification de la thèse principale à l'aide de l'emplacement

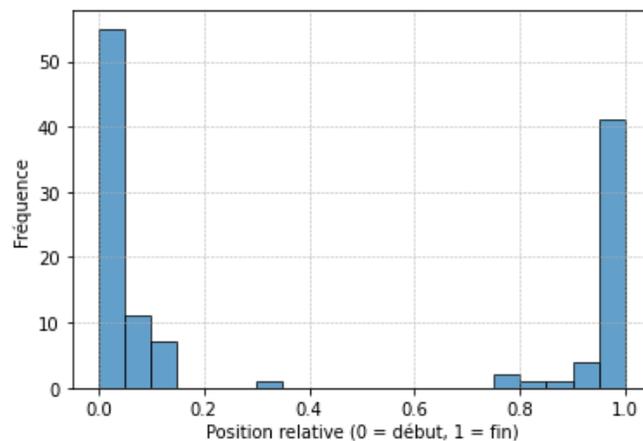
Une analyse de l'emplacement des énoncés de thèse principale dans le résumé ainsi que dans la discussion montre que ceux-ci se positionnent plus fréquemment au début et à la fin de cette structure et

macrostructure. Il est alors possible de développer des modèles de détection des énoncés de thèse principale fondés par cette connaissance.

9.4.3.1 Les résultats des modèles fondés par l'emplacement appliqués au résumé

La figure 9.3 suivante montre la fréquence d'occupation des positions relatives des énoncés de thèse principale dans les résumés. Ces positions sont normalisées de sorte que chaque position des énoncés de thèse principale dans le résumé est divisée par le nombre total de phrases dans le résumé duquel l'énoncé provient. Ainsi, les valeurs adjacentes à 0 correspondent au début du résumé, la valeur 1 correspond à la fin du résumé et toute valeur dans l'intervalle 0-1 correspond à la position relative dans le résumé.

Figure 9.3 Distribution des positions relatives des énoncés de thèse principale dans les résumés



On remarque que la fréquence du positionnement des énoncés de thèse principale est plus élevée au début du résumé. Un modèle simple de détection de thèse principale dans le résumé fondé par le positionnement consiste à déterminer les k premières phrases comme étant les k meilleurs candidats de thèse principale. Le tableau 9.23 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen de la position au début de la structure (nommons ce modèle « R_Début »). La performance du modèle R_Début est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$).

Tableau 9.23 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_Début

<i>k</i>	<i>vp</i>	<i>P-tk</i>	<i>R-tk</i>	<i>F1-tk</i>
1	55	0,500	0,447	0,472
2	71	0,323	0,577	0,414
3	73	0,221	0,593	0,322
4	73	0,166	0,593	0,259
5	74	0,135	0,602	0,220

Autrement, on peut tenir compte du second positionnement le plus fréquent pour l'énoncé de thèse principale, c.-à-d. la fin du résumé. Un modèle simple tenant compte du début et de la fin du résumé est fondé par l'alternance début-fin où les k meilleurs candidats sont déterminés par alternance de la partie entière supérieure des $k/2$ phrases du début (arrondi à l'entier supérieur) et de la partie entière inférieure des $k/2$ phrases de la fin (arrondi à l'entier inférieur) (nommons ce modèle « R_Alt-début-fin »). Le tableau 9.24 suivant montre la performance du modèle R_Alt-début-fin présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$).

Tableau 9.24 Précision top-k (P-tk), rappel top-k (R-tk) et F-mesure top-k (F1-tk) du modèle R_Alt-début

<i>k</i>	<i>vp</i>	<i>P-tk</i>	<i>R-tk</i>	<i>F1-tk</i>
1	55	0,500	0,447	0,472
2	96	0,436	0,780	0,560
3	112	0,339	0,911	0,494
4	116	0,264	0,943	0,412
5	118	0,215	0,959	0,351

Évidemment, le modèle R_Alt-début-fin est plus performant que R_Début pour les valeurs de $k > 1$. La performance de R_Alt-début-fin est globalement supérieure à celle du modèle de base. Ce modèle est $\sim 5,38$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes par rapport à l'augmentation des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions de ce modèle en fonction de k . Ce modèle termine avec une valeur de rappel $\sim 2,76$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,56$, est observée lorsque $k=2$. En bref, ce modèle performe mieux que le hasard.

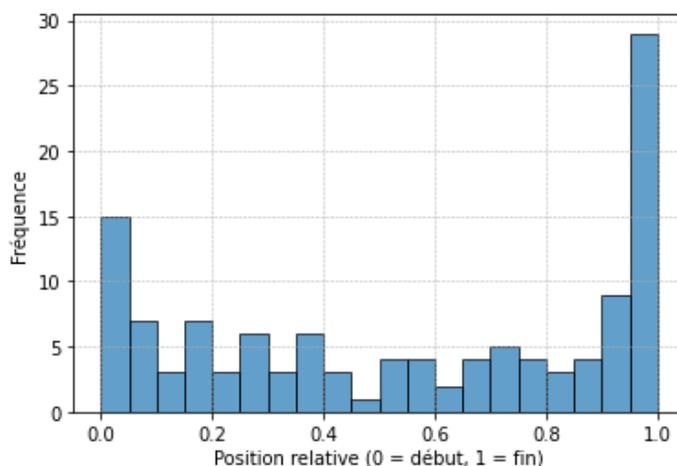
Enfin, spécifions que nos résultats sont cohérents avec d'autres études supportant un positionnement des thèses principales au début et à la fin de résumé. Par exemple, Biber et Gray (2016), dans leur analyse

structurale de rédaction académique, y compris de résumés, suggèrent que les thèses et conclusions principales suivent généralement la présentation du contexte, des détails méthodologiques et des résultats. Swales et Feak (2004), dans leur guide de rédaction universitaire, préconisent également une structure IMRAD où les thèses et conclusions principales se trouvent à la fin, soulignant la prédominance et l'efficacité de ce modèle à transmettre la recherche de manière cohérente. Autrement, Samraj (2005) montre que dans certaines disciplines, les thèses et conclusions principales sont introduites au début du résumé afin d'attirer immédiatement l'attention du lecteur sur les contributions clés de l'étude. De même, Pho (2008) observe que dans de certains domaines, il existe une tendance à présenter les principales affirmations ou découvertes dès le début du résumé. Cette stratégie est utilisée pour souligner dès le départ l'importance de la recherche et ses implications. En somme, le positionnement de la thèse principale à la fin du résumé répond à des objectifs d'efficacité et de cohérence et son positionnement en début du résumé répond à des objectifs rhétoriques.

9.4.3.2 Les résultats des modèles fondés par l'emplacement appliqués à la discussion

La figure 9.4 suivante montre la fréquence d'occupation des positions relatives des énoncés de thèse principale dans les discussions. Ces positions sont normalisées de la même manière que précédemment pour les résumés.

Figure 9.4 Distribution des positions relatives des énoncés de thèse principale dans les discussions



On remarque que les chances de trouver un énoncé de thèse principale à la fin de la discussion sont plus élevées. Un modèle simple de détection de thèse principale fondé par le positionnement dans la discussion consiste à déterminer les k dernières phrases comme étant les k meilleurs candidats de thèse principale.

Le tableau 9.25 suivant montre la performance du modèle de détection des énoncés de thèse principale identifiés dans la discussion au moyen de la position à la fin de la structure (nommons ce modèle « D_Fin »). La performance du modèle D_Fin est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$).

Tableau 9.25 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) du modèle D_Fin

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	16	0,145	0,131	0,138
2	26	0,118	0,213	0,152
3	32	0,097	0,262	0,142
4	34	0,077	0,279	0,121
5	40	0,073	0,328	0,119

Autrement, on peut tenir compte du second positionnement le plus probable pour l'énoncé de thèse principale, c.-à-d. le début de la discussion. Un modèle simple tenant compte du début et de la fin de la discussion est fondé par l'alternance fin-début où les k meilleurs candidats sont déterminés par alternance de la partie entière supérieure des $k/2$ phrases de la fin (arrondi à l'entier supérieur) et de la partie entière inférieure des $k/2$ des phrases du début (arrondies à l'entier inférieur) (nommons ce modèle « D_Alt -fin-début »). Le tableau 9.26 suivant montre la performance du modèle D_Alt -fin-début présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$).

Tableau 9.26 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) du modèle D_Alt-fin-début

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	16	0,145	0,131	0,138
2	28	0,127	0,230	0,164
3	38	0,115	0,311	0,168
4	40	0,091	0,328	0,142
5	46	0,084	0,377	0,137

Le modèle D_Alt-fin-début est plus performant que D_Fin pour les valeurs de $k > 1$. La performance de R_Alt-début-fin est globalement supérieure à celle du modèle de base. Ce modèle est $\sim 4,83$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes par rapport à l'augmentation des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions de ce modèle en fonction de k . Ce modèle termine avec une valeur de rappel $\sim 2,97$ fois supérieure au

modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,168$, est observée lorsque $k=3$. Bref, ce modèle performe mieux que le hasard.

En somme, on note que tous les modèles testés et fondés par le positionnement offrent une performance supérieure au niveau de base. De plus, les modèles fondés par l'alternance sont les plus performants (R_Alt-début-fin et D_Alt -fin-début).

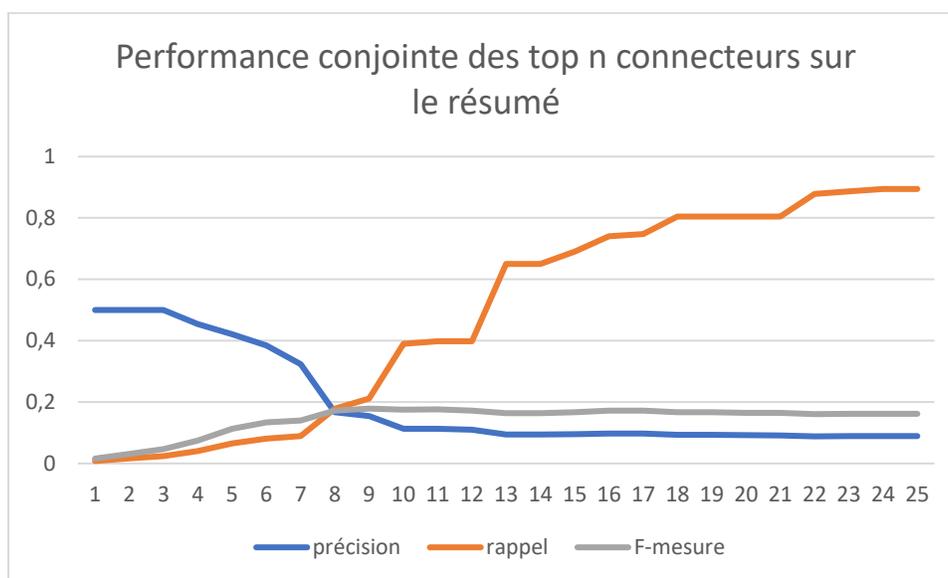
9.4.4 Les résultats de l'identification de la thèse principale à l'aide de connecteurs de discours.

Une analyse de la présence de connecteurs de discours dans les phrases du résumé et de la discussion montre que certains sont plus discriminants pour la tâche d'identification des énoncés de thèse principale. Parmi tous les connecteurs de discours, certains ont une valeur de précision plus élevée que les autres, mais cette valeur n'est généralement pas accompagnée d'une forte valeur de rappel. Cependant, l'utilisation conjointe des connecteurs de discours les plus précis offre possiblement une bonne performance en termes de précision et de rappel. Nous développons donc plusieurs modèles de détection des énoncés de thèse principale fondés par les n connecteurs les plus précis. Cette détection s'effectue à l'aide d'une expression régulière (*regex*) reprenant les n connecteurs séparés par un opérateur de disjonction (« | ») de sorte que n'importe quelle phrase contenant au moins un des n connecteurs est identifiée comme étant un énoncé de thèse principale. En appendice B sont présentés chacun des connecteurs de discours avec sa valeur de précision dans le résumé. En appendice C sont présentés chacun des connecteurs de discours avec sa valeur de précision dans la discussion.

9.4.4.1 Les résultats des modèles fondés par les connecteurs de discours appliqués au résumé

La figure 9.5 suivante montre la précision, le rappel et la F-mesure des top- n connecteurs sélectionnés au moyen de la précision et conjointement appliqués sur le résumé.

Figure 9.5 Précision, rappel et F-mesure des top-*n* connecteurs sélectionnés au moyen de la précision et conjointement appliqués sur le résumé



Les résultats de la F-mesure montrent que les modèles composés de 5 connecteurs ou plus font mieux que le niveau de base. Le nombre optimal de connecteurs de discours est 9 pour une F-mesure de 0,178 (nommons ce modèle « R_CD_9 »). La F-mesure de ce modèle est ~2,16 fois supérieure au niveau de base lorsque $k=1$. La précision est ~1,92 fois supérieure au niveau de base lorsque $k=1$. Les 9 meilleurs connecteurs de discours pour la détection des énoncés de thèse principale dans le résumé sont « *earlier* », « *hence* », « *not only [...] but also* », « *thus* », « *despite* », « *therefore* », « *because* », « *as* », « *plus* ». En se basant sur la typologie de la version 3 du *Penn Discourse Treebank (PDTB-3)*, on remarque que la majorité de ces 9 connecteurs (~44%) expriment une relation de contingence (*because, hence, therefore* et *thus*). Les connecteurs minoritaires expriment des relations temporelles (*as*⁹⁰, *earlier*), des expansions (*not only ... but also*⁹¹, *plus*) ou des comparaisons (*despite*).

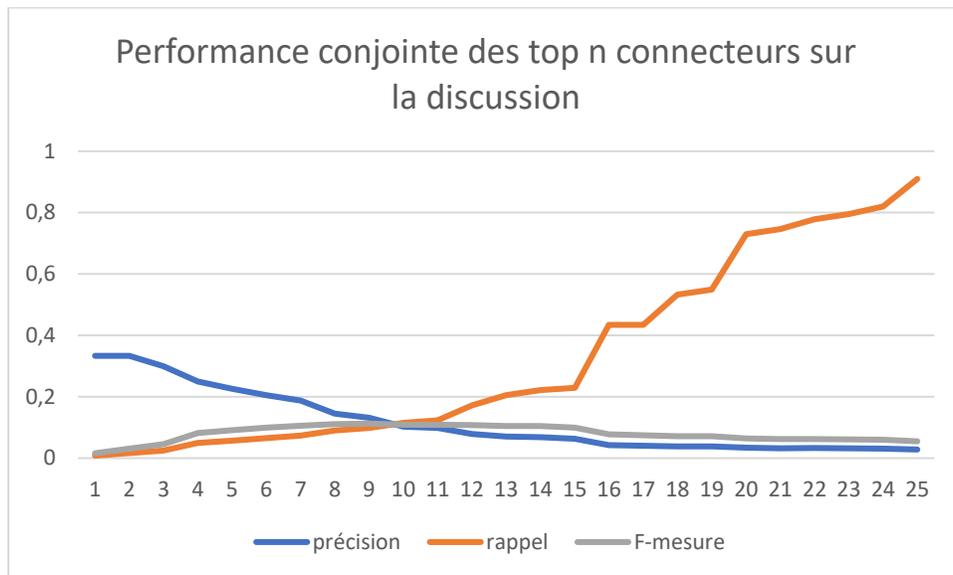
⁹⁰ Le connecteur « *as* » est polysémique et peut exprimer différents types de relation. Dans *PDTB-3*, ce connecteur exprime majoritairement des relations temporelles (~67%). Autrement, ce connecteur peut exprimer des relations de comparaison, de contingence ou d'expansion.

⁹¹ Le connecteur « *not only ... but also* » est polysémique et peut exprimer différents types de relation. Dans *PDTB-3*, ce connecteur exprime majoritairement des relations temporelles (~90%). Autrement, ce connecteur peut exprimer des relations de comparaison.

9.4.4.2 Les résultats des modèles fondés par les connecteurs de discours appliqués à la discussion

La figure 9.6 suivante montre la précision, le rappel et la F-mesure des top-n connecteurs sélectionnés au moyen de la précision et conjointement appliqués sur la discussion.

Figure 9.6 Précision, rappel et F-mesure des top-n connecteurs sélectionnés au moyen de la précision et conjointement appliqués sur la discussion



On remarque que les modèles composés de 2 connecteurs ou plus font mieux que le niveau de base. Le nombre optimal de connecteurs de discours est 9 pour une F-mesure de 0,113 (nommons ce modèle « D_CD_9 »). La F-mesure de ce modèle est $\sim 3,90$ fois supérieure au niveau de base lorsque $k=1$. La précision est $\sim 4,40$ fois supérieure au niveau de base lorsque $k=1$. Les 9 meilleurs connecteurs de discours pour la détection des énoncés de thèse principale dans la discussion sont « *as an alternative* », « *besides* », « *instead of* », « *hence* », « *along with* », « *consequently* », « *even if* », « *earlier* », « *in fact* ». En vertu de la typologie de *PDTB-3*, on observe que la majorité de ces 9 connecteurs ($\sim 56\%$) expriment une relation d'expansion (*along with*, *as an alternative*, *besides*, *in fact*⁹² et *instead of*). Les connecteurs minoritaires expriment des relations de contingence (*consequently*, *hence*), temporelles (*earlier*) ou de comparaison (*even if*).

⁹² Le connecteur « *in fact* » peut exprimer différents types de relation. Dans *PDTB-3*, ce connecteur exprime majoritairement des relations d'expansion ($\sim 87\%$). Autrement, il peut exprimer des relations de comparaison.

En somme, on remarque que les modèles composés de moins de connecteurs offrent une forte valeur de précision et une faible valeur de rappel. Inversement les modèles composés de plus de connecteurs offrent une valeur de précision faible et une valeur de rappel élevée. Cela est attendu puisque les modèles fondés par les connecteurs de discours sont composés de ceux qui sont les plus précis. Pour le résumé et la discussion, le nombre optimal de connecteurs est de 9. Outre deux (2) connecteurs, « *earlier* » et « *hence* », les modèles R_CD_9 et D_CD_9 ne sont pas composés des mêmes connecteurs. Selon la typologie de PDTB-3, « *earlier* » exprime la temporalité et « *hence* » la contingence. Lorsqu'on unie les deux (2) ensembles de connecteurs les plus performants, on obtient un ensemble de 18 connecteurs dont la majorité remplissent la fonction d'expansion (~39%) et, plus spécifiquement, la conjonction⁹³ (~28%), la disjonction⁹⁴ (~6%) et la substitution⁹⁵ (~6%). Le second type de connecteur le plus fréquent exprime la contingence (~33%) et, plus spécifiquement, la relation de causalité entre deux segments de texte, où l'un est présenté comme la cause et l'autre comme l'effet. Parmi les connecteurs exprimant la causalité, la majorité (~28%) relie une cause à un résultat (*consequently, hence, therefore, thus*). Autrement dit, ce type de relateur est associé au résultat ou à l'issue d'un certain événement ou d'une situation, c.-à-d. ce qui se passe en conséquence d'une cause particulière. Il s'agit d'un sous-type de la classe plus large des relateurs de contingence causale qui indiquent une relation de dépendance d'un segment de texte par rapport à un autre.

9.4.5 Les résultats de la combinaison des approches les plus performantes

Les différentes approches de détection des énoncés de thèse principale qui ont été investiguées sont celles fondées par les propriétés syntaxiques ou sémantiques, l'emplacement ainsi que les connecteurs de discours. Pour chacune de ces approches, certains modèles se sont avérés plus performants que d'autres et, subséquemment, nous proposons d'investiguer une approche synthétique où les meilleurs modèles de chacune de ces approches sont utilisés de manière à former des modèles ensemblistes.⁹⁶ Ce modèle est

⁹³ La relation d'expansion conjonctive est utilisée quand deux segments de texte (ou plus) sont mis en relation de manière additive.

⁹⁴ La relation d'expansion disjonctive décrit un choix ou une alternative entre deux segments de texte (ou plus).

⁹⁵ La relation d'expansion substitutive permet à un segment de texte d'agir comme un substitut ou un équivalent pour un autre.

⁹⁶ Plus spécifiquement, un premier modèle ensembliste sera constitué du modèle le plus performant dans chacune des trois approches (propriétés syntaxiques ou sémantiques, emplacement et connecteurs de discours), puis un second modèle ajoutera au premier le second modèle le plus performant de l'approche des propriétés syntaxiques ou sémantiques afin de savoir s'il est possible d'améliorer la performance des modèles ensemblistes en ajoutant un modèle simple relativement performant.

fondé par l'accord interjuge qui consiste à accorder un vote à chacun des modèles les plus performants. Ainsi, chacun des modèles les plus performants identifie les k meilleurs candidats et ceux ayant récoltés le plus de votes sont déterminés comme étant les k meilleurs candidats de thèse principale. Le vote du modèle le plus performant est prépondérant au sens où il détermine un gagnant unique lors d'un vote égalitaire. Spécifions que les modèles fondés par les connecteurs de discours ne contiennent pas de critère permettant de restreindre le nombre de prédictions à une valeur k arbitraire. Puisqu'un tel critère se trouve dans les autres modèles, celui-ci sera utilisé par les modèles fondés par les connecteurs de discours, c.-à-d. que ces derniers n'accorderont leur vote qu'aux candidats déjà proposés par les autres modèles. Donc, le modèle fondé par les connecteurs de discours sélectionne tous les candidats où au moins un des 9 connecteurs est présent, mais seulement parmi les candidats proposés par les modèles fondés par les propriétés syntaxiques ou sémantiques ou l'emplacement.

9.4.5.1 Les résultats du modèle ensembliste appliqué au résumé

Les modèles utilisés pour former un premier modèle ensembliste appliqué au résumé sont R_Alt-début-fin (vote prépondérant), R_I_DBOW et R_CD_9 (nommons ce modèle ensembliste « R_Ensemble_1 »). Au tableau 9.27 suivant, la performance du modèle R_Ensemble_1 est présentée en termes de précision top- k ($P-tk$), rappel top- k ($R-tk$) et F-mesure top- k ($F1-tk$).

Tableau 9.27 Précision top- k ($P-tk$), rappel top- k ($R-tk$) et F-mesure top- k ($F1-tk$) du modèle R_Ensemble_1

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	54	0,491	0,439	0,464
2	97	0,441	0,789	0,566
3	111	0,336	0,902	0,490
4	114	0,259	0,927	0,405
5	115	0,209	0,935	0,342

La performance du modèle R_Ensemble_1 est globalement supérieure à celle du modèle de base. Ce modèle est $\sim 5,28$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes en fonction de l'augmentation des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions de ce modèle. Ce modèle termine avec une valeur de rappel $\sim 2,69$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,566$, est observée lorsque $k=2$. Bref, ce modèle performe mieux que le hasard, mais moins bien que le modèle simple le plus performant soit R_Alt-début-fin.

Les modèles utilisés pour former un deuxième modèle ensembliste appliqué au résumé sont R_Alt-début-fin (vote prépondérant), R_I_DBOW, R_D_DBOW et R_CD_9 (nommons ce modèle ensembliste « R_Ensemble_2 »). Au tableau 9.28 suivant, la performance du modèle R_Ensemble_2 est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$).

Tableau 9.28 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) du modèle R_Ensemble_2

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	48	0,436	0,390	0,412
2	81	0,368	0,659	0,472
3	95	0,288	0,772	0,419
4	99	0,225	0,805	0,352
5	101	0,184	0,821	0,300

La performance du modèle R_Ensemble_2 est globalement supérieure à celle du modèle de base. Ce modèle est ~4,69 fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes en fonction de l'augmentation des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions de ce modèle. Ce modèle termine avec une valeur de rappel ~2,36 fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant ~0,412, est observée lorsque $k=2$. Bref, ce modèle performe mieux que le hasard, mais moins bien que le modèle ensembliste précédent qui n'inclue pas le modèle R_D_DBOW.

9.4.5.2 Les résultats du modèle ensembliste appliqué à la discussion

Les modèles utilisés pour former un premier modèle ensembliste appliqué à la discussion sont D_I_DBOW (vote prépondérant), D_Alt-fin-début et D_CD_9 (nommons ce modèle ensembliste « D_Ensemble_1 »). Au tableau 9.29 suivant, la performance du modèle D_Ensemble_1 est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$).

Tableau 9.29 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) du modèle D_Ensemble_1

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	31	0,282	0,254	0,267
2	46	0,209	0,377	0,269
3	52	0,158	0,426	0,230
4	59	0,134	0,484	0,210
5	65	0,118	0,533	0,193

La performance du modèle D_Ensemble_1 est globalement supérieure à celle du modèle de base. Ce modèle est $\sim 9,4$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes en fonction de l'augmentation des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions de ce modèle. Ce modèle termine avec une valeur de rappel $\sim 4,20$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,269$, est observée lorsque $k=2$. Bref, ce modèle performe mieux que le hasard ainsi que le modèle simple le plus performant soit D_I_DBOW.

Les modèles utilisés pour former le deuxième modèle ensembliste appliqué à la discussion sont D_I_DBOW (vote prépondérant), D_T_DBOW, D_Alt-fin-début et D_CD_9 (nommons ce modèle ensembliste « D_Ensemble_2 »). Au tableau 9.30 suivant, la performance du modèle D_Ensemble_2 est présentée en termes de précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$).

Tableau 9.30 Précision top-k ($P-tk$), rappel top-k ($R-tk$) et F-mesure top-k ($F1-tk$) du modèle D_Ensemble_2

k	vp	$P-tk$	$R-tk$	$F1-tk$
1	31	0,282	0,254	0,267
2	46	0,209	0,377	0,269
3	56	0,170	0,459	0,248
4	62	0,141	0,508	0,221
5	67	0,122	0,549	0,199

La performance du modèle D_Ensemble_2 est globalement supérieure à celle du modèle de base. Ce modèle est $\sim 9,40$ fois plus précis que le niveau de base lorsque $k=1$. Les valeurs de précision sont décroissantes en fonction de l'augmentation des valeurs de k , ce qui indique un ordre descendant de vraisemblance pour les prédictions de ce modèle. Ce modèle termine avec une valeur de rappel $\sim 4,32$ fois supérieure au modèle de base lorsque $k=5$. Enfin, la valeur maximale de la F-mesure, atteignant $\sim 0,269$, est observée lorsque $k=2$. Notons que les valeurs de performance sont identiques entre D_Ensemble_1 et D_Ensemble_2 pour les valeurs k inférieure ou égale à deux (2), mais que pour $k>2$, D_Ensemble_2 performe mieux que D_Ensemble_1. Bref, ce modèle performe mieux que le hasard ainsi que le modèle ensembliste précédent qui n'inclue pas le modèle D_T_DBOW.

En somme, les modèles ensemblistes se comportent très différemment dépendamment de la structure ou de la macrostructure dans lesquelles on les utilise. Dans la structure du résumé, ils performent moins bien

que le modèle simple le plus performant, c.-à-d. celui fondé par l'emplacement. Le modèle D_Alt-début-fin offre les performances les plus élevées pour la détection des énoncés de thèse principale dans le résumé. Dans la macrostructure de la discussion, les modèles ensemblistes offrent de meilleures performances comparativement aux modèles simples. Le modèle ensembliste D_Ensemble_2 composé de quatre (4) modèles simples offre les performances les plus élevées pour la détection des énoncés de thèse principale dans la discussion.

9.5 Conclusion

Dans ce chapitre consacré à la détection des énoncés de thèse principale, une exploration de diverses méthodes a été menée. Les différentes approches ont été classées selon certains critères, allant des propriétés syntaxiques ou sémantiques aux emplacements spécifiques dans les textes, en passant par l'utilisation de connecteurs de discours. Nos expérimentations révèlent que les modèles fondés sur une représentation textuelle de type DBOW sont généralement plus performants que ceux basés sur une représentation de type « text-embedding-ada-002 ». Cette observation montre que la sophistication du modèle de représentation n'est pas un gage de performance.

Aussi, la macrostructure d'introduction se démarque en termes de performance pour la tâche de détection des thèses principales dans le résumé et dans la discussion. Le titre est apparu comme une structure informative, particulièrement lorsqu'elle est utilisée afin d'identifier les énoncés de thèse principale dans la discussion (ainsi que le résumé dans une moindre mesure). De plus, la discussion est apparue comme une macrostructure informative, notamment lorsqu'elle est utilisée pour identifier les énoncés de thèse principale dans le résumé.

L'analyse basée sur l'emplacement a également démontré son efficacité, confirmant que les énoncés de thèse principale ont tendance à se situer à un endroit spécifique soit au début ou à la fin des structures ou macrostructures. Cette approche s'est même révélée la plus performante, au-delà des modèles ensemblistes, pour la détection des énoncés de thèse principale dans le résumé. Ce résultat indique que l'emplacement est une information déterminante pour la tâche qui nous intéresse, et que les autres approches testées fournissent des informations secondaires ou redondantes dans le cas du résumé.

Quant aux connecteurs de discours, ils se sont avérés être des indicateurs précieux, bien que leur utilisation exige une combinaison judicieuse pour maximiser à la fois la précision et le rappel. Cependant,

notons que l'approche des connecteurs de discours pourrait s'avérer moins efficace sur un corpus plus large. La raison est que, bien qu'il soit possible de distinguer les énoncés de thèses (principales ou non) des autres composantes argumentatives à l'aide des connecteurs de discours, il n'est pas évident que les connecteurs de discours soient suffisants pour discriminer les thèses subordonnées des thèses principales. Si cette limite est juste, alors il est possible que la réussite du modèle de détection fondé sur les connecteurs soit un artefact statistique explicable par la petitesse de notre corpus. À cet effet, il serait intéressant de reproduire cette recherche sur un corpus plus large qui n'est pas constitué exclusivement d'articles médicaux. Enfin, il est important de noter que la combinaison de ces différentes approches semble offrir une stratégie optimale pour détecter les énoncés de thèse principale dans la discussion. La combinaison de ces approches, l'utilisation de structure ou de macrostructure spécifique, la prise en compte de la position de la thèse principale, et l'intégration des connecteurs de discours semble maximiser la détection en termes de précision et de rappel. Contrairement au cas de la détection des énoncés de thèse principale dans le résumé, l'emplacement, la structure du titre, la macrostructure de l'introduction, et les connecteurs de discours fournissent des informations complémentaires permettant une maximisation de la performance. Les recherches futures pourraient explorer davantage les modèles ensemblistes, combinant les forces des différentes méthodes présentées ici. De plus, il serait intéressant d'explorer d'autres formes de représentations textuelles et d'examiner leur efficacité relative dans le contexte de cette tâche.

CHAPITRE 10

Conclusion et perspectives

Dans ce chapitre, conclusion et perspectives, nous récapitulons notre parcours de recherche, qui débute par une exploration de l'analyse du discours et des théories de l'argumentation. Nous résumons les modèles computationnels utilisés pour notre analyse des arguments. Nous détaillons nos méthodes pour détecter les contextes de citation dans les structures IMRAD et les thèses principales, en soulignant les résultats de nos études expérimentales. Nous discutons ensuite des limitations et des faiblesses de notre recherche, en offrant une perspective critique sur notre travail. Nous terminons par des propositions pour les recherches futures, suggérant des directions pour affiner nos méthodes et les appliquer à de nouveaux domaines d'étude.

10.1 Introduction

Cette thèse a débuté avec une exploration de l'analyse du discours, se penchant spécifiquement sur les approches externaliste et internaliste, et comment celles-ci influencent notre compréhension du discours exprimé sous la forme matérielle du texte. Nous avons ensuite plongé dans le monde fascinant de l'argumentation, examinant quatre théories majeures, allant des approches rhétoriques antiques aux approches communicationnelles contemporaines. Ensuite, nous nous sommes engagés dans une étude détaillée des modèles computationnels de l'analyse des arguments, en mettant l'accent sur leur applicabilité pratique à travers différentes instanciations computationnelles permettant l'analyse ou la détection de structures et de relations argumentatives. La décomposition de la tâche d'analyse des arguments a montré la nature complexe de ce processus. Nous avons proposé une implémentation de la tâche de détection de certaines composantes et relations argumentatives, en se concentrant sur la représentation vectorielle des textes ainsi que diverses méthodes de détection. Enfin, nous avons posé les bases théoriques de la structure textuelle de l'argumentation dans les *TEXTES Courts de Haut Niveau Théorique* publiés dans les *Revue à Comité de lecture (TECHNETIUM)*. Puis, nous avons formulé des hypothèses concernant la structure argumentaire de ce type de texte. Notamment, nous avons postulé trois niveaux structurels aux *TECHNETIUM* soit un niveau microstructurel où les relations locales sont descriptibles au moyen de la *Rhetorical Structure Theory (RST)*, un niveau macrostructurel où les relations globales sont descriptibles au moyen de la théorie IMRAD, puis un niveau étendu où les structures intratextuelles se relient à celles d'autres textes. Rappelons que notre objectif sous-jacent est d'identifier comment un texte se relie à d'autres textes, non pas de manière bibliographique, mais argumentaire,

notamment en reliant l'objectif principal d'un texte à celui d'autres textes. Or, cet objectif louable implique d'abord une meilleure compréhension de la relation entre l'objectif principal du texte et les différentes microstructures et macrostructures argumentatives. Pour ce faire, nous avons identifié deux (2) composantes argumentaires jouant un rôle crucial pour ce type d'analyse, c.-à-d. les contextes de citation pour leur rôle de relateur envers d'autres textes et les thèses principales en tant qu'objectif principal du TECHNETIUM. Subséquemment, nous avons proposé des méthodes pour identifier ces structures et composantes, lesquelles ont mené à deux études expérimentales.

Lors de la première expérimentation, nous avons investigué différentes méthodes de détection de l'appartenance des contextes de citation aux macrostructures argumentatives IMRAD, utilisant diverses approches de représentation et de classification. Lors de la seconde expérimentation, nous avons approfondi la tâche de détection des énoncés de thèse principale dans le résumé et la discussion, en explorant différentes approches tout en évaluant leur performance. Ces deux tâches sont importantes lors de l'analyse des arguments. D'une part, si la thèse principale est l'objectif principal du texte de haut niveau théorique, alors la tâche d'identification de la tâche principale est un objectif incontournable de la fouille d'argument ou LAATAO. D'autre part, les contextes de citation sont des composantes par lesquelles un texte se relie à d'autres textes. Dans le cas des textes structurés IMRAD, cette relation se subsume au sous-objectif propre à la macrostructure IMRAD de laquelle elle provient. Ce faisant, la tâche d'identification de l'appartenance des contextes de citation aux macrostructures IMRAD est importante pour déterminer la relation intertextuelle. À cet effet, il serait intéressant de voir comment se développe un réseau de relations intertextuelles, non pas composé de toutes les références bibliographiques de manière indiscriminée, mais distinguée en fonction des sous-objectifs IMRAD, notamment l'introduction pour un réseau reliant des questionnements, la méthodologie pour un réseau joignant des aspects méthodologiques, les résultats pour un réseau reliant certaines caractéristiques des résultats et la discussion pour un réseau joignant des aspects de réponses ou, possiblement, des thèses principales.

10.2 Résumé de la méthode

Au chapitre 7, nous avons formulé des hypothèses à propos de la lecture et de l'analyse de texte assistées par ordinateur. Plus spécifiquement, ces hypothèses concernent deux tâches d'analyse d'arguments dans les TECHNETIUM, soit la détection de l'appartenance des contextes de citation aux macrostructures argumentatives IMRAD ainsi que la de détection des énoncés de thèse principale. Notre première hypothèse (H1) soutient que l'identification de l'appartenance des contextes de citation aux

macrostructures argumentatives IMRAD est possible à l'aide de différentes techniques de représentation et de classification. Cette hypothèse donne lieu à deux hypothèses subordonnées soit i) que les techniques de représentation impliquant un grand nombre de paramètres, de type « text-embedding-ada-002 » notamment, sont plus performantes que les autres (H1.1) et ii) que la classification linéaire offre de meilleurs performances (H1.2). Notre seconde hypothèse (H2) soutient qu'il est possible de déterminer les énoncés de thèse principale dans les TECHNETIUM au moyen de diverses méthodes. Parmi ces dernières se trouvent trois types d'indices matériels permettant l'identification des énoncés de thèse principale soit i) l'emplacement (H2.1), ii) les connecteurs de discours (H2.2) et iii) les propriétés syntaxiques ou sémantiques propres à certaines structures ou macrostructures IMRAD (H2.3). Parmi ces structures et macrostructures IMRAD pouvant faciliter l'identification des énoncés de thèse principale se trouvent le titre (H2.3.1), le résumé (H2.3.2), l'introduction (H2.3.3) et la discussion (H2.3.4). De plus, l'hypothèse H2.1 a donné lieu à deux hypothèses subordonnées soit i) que les thèses principales sont susceptibles de se trouver dans une macrostructure particulière (H2.1.1) et ii) que les thèses principales sont susceptibles de se trouver à un emplacement particulier dans cette macrostructure (H2.1.2).

10.2.1 Les modèles computationnels de détection de l'appartenance des contextes de citation aux macrostructures IMRAD

Au chapitre 8, nous avons traduit la première hypothèse (H1) en modèles computationnels de détection de l'appartenance des contextes de citation aux macrostructures argumentatives IMRAD. Un total de 153 modèles sont construits à l'aide de deux composantes, soit une représentation vectorielle des contextes de citation et un classifieur identifiant l'appartenance des contextes de citation aux macrostructures IMRAD. D'une part, 8 types de représentations vectorielles sont utilisés soit BOW (noté « BOW »), DBOW (noté « DBOW »), OpenAI (noté « OAI »), BERT de base (noté « BERT_base »), BERT large (noté « BERT_large »), BERT large affiné sur une tâche de NLI et STS (noté « BERT_nli_stsb »), BERT de base affiné sur plusieurs paires (noté « BERT_mpnet »), et BERT de base affiné sur la tâche de détection de l'appartenance des contextes de citation aux macrostructures IMRAD au moyen de corpus InTeReC (noté « BERT_imrad »). D'autre part, 20 types de classifieurs sont utilisés soit les K-moyennes (noté « KMeans »), le mélange gaussien (noté « Gaussian Mixture»), les k plus proches voisins (noté « K Neighbors Classifier »), les voisins de rayon (noté « Radius Neighbors Classifier »), le plus proche centroïde (noté « Nearest Centroid »), le SVM avec noyau linéaire (notés « SVC_linear »), le SVM avec noyau de fonction de base radiale (notés « SVC_RBF »), la régression logistique (notée « Logistic Regression »), la régression logistique avec validation croisée (notée « Logistic Regression CV »), la régression Ridge (notée « Ridge Classifier »),

la régression Ridge avec validation croisée (notée « Ridge Classifier CV »), la descente de gradient stochastique (notée « SGD Classifier »), la classification passive-agressive (notée « Passive Agressive Classifier »), les arbres de décision (notés « Decision Tree Classifier »), les forêts aléatoires (notées « Random Forest Classifier »), les classifieurs d'arbre supplémentaire (notés « Extra Tree Classifier »), l'amplification de gradient (notée « Gradient Boosting Classifier »), le perceptron (noté « Perceptron »), le perceptron multicouche (noté « MLP Classifier ») et le modèle de classification de séquences BERT (noté « BERT for Sequence »). Ces méthodes de classification sont regroupées en familles soit i) les modèles de classification basée sur la catégorisation, ii) les modèles de classification instanciée, iii) les modèles de classification linéaire, iv) les modèles de classification arborescente, et v) les modèles de classification neuronale. Afin d'évaluer la performance de ces modèles, ceux-ci sont appliqués sur InTeReC qui est un corpus de contextes de citation issus de TECHNETIUM, lesquels sont annotés de leur appartenance aux macrostructures IMRAD. Le corpus InTeReC étant composé d'un sous-ensemble de données déterminées (annotées « introduction », « méthodes », « résultats » ou « discussion ») et d'un sous-ensemble de données indéterminées (annotées « méthodes et résultats » ou « résultats et discussions »), alors il nous a été possible d'utiliser un modèle entraîné sur les données déterminées pour évaluer la performance sur les données indéterminées et tester ainsi la capacité de généralisation de cette approche.

10.2.2 Les modèles computationnels de détection des thèses principales

Au chapitre 9, nous avons traduit la seconde hypothèse (H2) en soixante-quatorze (74) modèles computationnels de détection des énoncés de thèse principale. Systématiquement, la moitié de ces modèles (37) vise la détection des thèses principales dans la structure du résumé et l'autre moitié (37) vise la détection des thèses principales dans la macrostructure de discussion. Pour chaque hypothèse subordonnée à H2, plusieurs modèles sont élaborés. Une technique est élaborée et évaluée pour l'identification des thèses principales dans l'emplacement particulier de la discussion (H2.1.1). Quatre (4) modèles sont construits pour tester l'hypothèse de l'emplacement à l'intérieur de la structure du résumé et de la macrostructure de discussion (H2.1.2) suivant l'inspection du corpus AbstrCT+ (l'emplacement privilégié apparaissant au début ou en alternance début-fin pour le résumé ainsi qu'à la fin ou en alternance fin-début pour la discussion). Cinquante (50) modèles sont créés pour tester l'hypothèse des connecteurs de discours (H2.2) suivant l'inspection du corpus AbstrCT+ (au moyen des n connecteurs les plus précis dans ce corpus entre $n=1$ et $n=25$). Enfin, deux (2) types de représentations vectorielles sont utilisées (basées sur DBOW et « text-embedding-ada-002 ») pour tester les quatre (4) hypothèses subordonnées à H2.3 pour la détection de thèses principales dans le résumé ainsi que dans la discussion.

En somme, seize (16) modèles sont élaborés de manière à postuler comme thèse principale les candidats (du résumé ou de la discussion) les plus similaires (similarité cosinus) au titre (tel que suggéré par H2.3.1), au résumé (tel que suggéré par H2.3.2), à l'introduction (tel que suggéré par H2.3.3) ou à la discussion (tel que suggéré par H2.3.4). Enfin, les meilleurs modèles sont utilisés pour former des modèles ensemblistes où chaque modèle simple vote pour les meilleurs candidats de thèse principale de manière à réaliser un vote majoritaire parmi les prédictions des modèles qui le composent. Chaque modèle vote pour ses k meilleurs candidats de thèse principale⁹⁷, et les k candidats qui reçoivent le plus grand nombre de votes sont sélectionnés comme prédiction finale de l'ensemble. Puisque les meilleurs modèles ne sont pas exactement les mêmes lors de la détection des thèses principales dans le résumé et dans la discussion, alors des modèles ensemblistes distincts sont construits pour cette structure et macrostructure. Plus précisément, deux (2) modèles ensemblistes sont construits pour le résumé et deux (2) autres pour la discussion. Afin d'évaluer la performance de ces soixante-quatorze (74) modèles de détection de thèses principales, ceux-ci sont appliqués sur AbstrCT+ qui est un corpus de TECHNETIUM où les phrases du résumé et les phrases de la discussion sont annotées de leur rôle de thèse principale (ou non).

10.3 La synthèse des résultats

L'application de cent-cinquante-trois (153) modèles de détection de l'appartenance des contextes de citation aux macrostructures IMRAD ainsi que soixante-quatorze (74) modèles de détection des thèses principales sur nos données textuelles a permis d'obtenir des résultats intéressants. Ceux-ci ont été évalués en termes de précision et de rappel afin de déterminer les approches les plus performantes parmi celles investiguées.

10.3.1 La synthèse de la détection de l'appartenance des contextes de citation aux macrostructures

Cette expérience traite de la modélisation de l'identification des macrostructures IMRAD (Introduction, Méthodes, Résultats, Discussion) d'où proviennent des contextes de citation. Pour cela, 80% des données d'un corpus spécifique ont été utilisées pour entraîner des modèles de classification, tandis que les 20% restants ont servi à l'évaluation. Les performances des modèles ont été mesurées à l'aide de différentes métriques comme la F-mesure, la macro-moyenne et la moyenne pondérée. La F-mesure a été calculée comme la moyenne harmonique de la précision et du rappel. La macro-moyenne et la moyenne pondérée

⁹⁷ À l'exception du modèle fondé par les connecteurs de discours qui sélectionne tous les énoncés contenant au moins un connecteur parmi un sous-ensemble d'énoncés composé des candidats proposés par les autres modèles simples.

ont permis d'agrèger les performances des classes IMRAD. Une analyse exploratoire a été réalisée pour identifier les caractéristiques textuelles pertinentes, utilisant la régression logistique pour évaluer l'importance des mots dans différentes sections IMRAD. La majorité des mots spécifiques permettant le classement des contextes de citations apparaissent cohérents avec les fonctions argumentatives des macrostructures auxquelles ils appartiennent. En général, le support de chaque classe semble avoir été suffisant pour la modélisation de cette tâche, à l'exception de la macrostructure « résultats » dont le nombre de cas apparaît suboptimal.

Le modèle BERT a été utilisé pour affiner la représentation vectorielle des données ainsi que pour la classification en portant une attention particulière à la fonction de perte et à la F-mesure moyenne pondérée. Les résultats montrent une amélioration initiale rapide puis une stabilisation des performances, avec des signes de surapprentissage après environ 5 passages du modèle sur les données d'apprentissage (5 époques).

Plusieurs modèles et techniques de représentation vectorielle ont été testés, notamment BOW, DBOW, « text-embedding-ada-002 » d'OpenAI (OAI), et différentes variantes de BERT. Les résultats ont montré que la tâche de représentation vectorielle en amont s'avère déterminante pour la tâche de classification en aval. La technique de représentation qui s'est avérée la plus performante est obtenue en affinant le modèle de classification de séquence BERT sur les données textuelles d'InTeReC (BERT_imrad). Si l'affinage sur InTeReC n'avait pas été possible, alors la représentation vectorielle d'OpenAI sans affinage (AOI) aurait obtenu la performance la plus élevée lorsqu'associée à un SVM à noyaux RBF. La représentation qui s'est avérée la moins performante parmi celles investiguées est la représentation DBOW. Or, il est possible qu'une optimisation des paramètres de ce modèle produise de meilleures performances. Le cas échéant, DBOW apparaîtrait être une représentation capricieuse nécessitant une paramétrie particulière. Dans tous les cas, il serait étonnant qu'une représentation DBOW optimale offre des performances supérieures aux représentations BERT ou OpenAI lorsqu'associée aux classifieurs les plus performants.

Des comparaisons ont été faites entre différents modèles de classification, y compris des classifieurs linéaires, arborescents et neuronaux. Étonnamment, la méthode de représentation la plus performante – la classification de séquence BERT – qui affine son processus de classification en parallèle de l'affinage de la représentation n'est pas la méthode de classification optimale pour cette tâche. Parmi les classifieurs

investigués, les classifieurs les plus performants appartiennent à la famille des classifieurs linéaires et, parmi ces derniers, le plus performant est la régression logistique avec validation croisée.

L'application du modèle de classification de séquence BERT sur les données indéterminées indique un potentiel de généralisation des cas déterminés vers les cas indéterminés. La mesure de rappel maximale observée pour l'application du modèle aux données indéterminées est conséquente avec la mesure de rappel moyenne réelle observée sur l'application du modèle de classification de séquence BERT sur le sous-ensemble d'évaluation des données déterminées.

En somme, les contextes de citation semblent contenir des informations sémantiques ou syntaxiques suffisamment discriminantes pour permettre l'identification de leur appartenance aux macrostructures IMRAD. À cet effet, quatre modèles se distinguent particulièrement pour la résolution de notre problème de classification en raison de leurs performances et caractéristiques uniques. Premièrement, le modèle de classification séquentielle BERT est notable pour sa capacité à créer la représentation BERT_imrad, affichant une performance élevée avec une F1 pondérée d'environ 0,748. Bien qu'efficace et accessible gratuitement, ce modèle nécessite un temps d'entraînement considérable. Deuxièmement, le modèle combinant la régression logistique avec validation croisée et la représentation vectorielle générée par le modèle BERT présente la meilleure performance, avec une F1 pondérée d'environ 0,753. Ce modèle est également gratuit et efficace, mais il requiert un entraînement prolongé et permet de déterminer si le problème est de nature convexe. Troisièmement, un modèle utilisant un SVM avec noyau RBF et la représentation vectorielle d'OpenAI se distingue par une impressionnante F1 pondérée d'environ 0,722. Il est performant, rapide à entraîner, utile pour vérifier la convexité du problème, mais est un système propriétaire et donc moins transparent pour les utilisateurs. Quatrièmement, un modèle combinant un SVM avec noyau RBF et la représentation vectorielle BERT de base offre une performance notable avec une F1 pondérée d'environ 0,711. Ce modèle est non seulement performant et rapide à entraîner, mais également gratuit et utile pour évaluer la convexité du problème de classification. Enfin, notons qu'OpenAI offre une version payante de la classification de séquence qui est adaptée aux diverses représentations offertes par l'entreprise. Étant donné la performance exceptionnelle de la classification de séquence BERT ainsi que de la représentation « text-embedding-ada-002 » non affinée sur notre problème de classification, il est hautement probable qu'une représentation de type « text-embedding-ada-002 » affinée sur nos données offrirait des performances encore plus élevées. Il serait intéressant de tester cette hypothèse, notamment avec le développement rapide de représentations toujours plus complexes (depuis

« text-embedding-ada-002 » est sorti plusieurs versions, notamment « text-embedding-3-large » et « text-embedding-3-small »).

10.3.2 La synthèse de la détection des thèses principales

Cette expérience aborde en détail la détection des énoncés de thèse principale dans les résumés et les discussions de TECHNETIUM, en explorant diverses méthodologies. Les principales approches incluent l'analyse des propriétés syntaxiques ou sémantiques, l'emplacement des énoncés de thèse dans le texte, et l'usage des connecteurs de discours. Notre analyse du corpus AbstrCT montre que seulement 110 des 669 résumés d'articles scientifiques contiennent l'expression d'au moins une thèse principale et, le cas échéant, un maximum de trois (3) énoncés de thèse principale. Ce faisant, la structure du résumé apparaît comme un lieu occasionnel d'expression de thèses principales et un même résumé peut contenir plusieurs énoncés de thèse principale. *A contrario*, il serait étonnant qu'un TECHNETIUM ne contienne aucune thèse principale, et son énonciation dans la macrostructure de discussion est probable. À cet effet, notre analyse d'un échantillon d'énoncés de thèse principale dans la discussion montre que l'identification des énoncés de thèse principale dans la discussion à partir des énoncés de thèse principale dans le résumé est possible au moyen d'une mesure de similarité. Évidemment, tout énoncé de thèse principale dans le résumé n'est pas nécessairement similaire (ou plus similaire) à une expression de thèse principale dans la discussion. Autrement, les énoncés identifiés dans les résumés comme étant des thèses principales ne sont pas nécessairement des thèses principales puisque le processus d'annotation n'est pas infallible. Subséquemment, on ne peut s'attendre à ce que tous les énoncés identifiés comme thèse principale dans la discussion soient *de facto* des thèses principales. Néanmoins, l'accord interannotateur élevé du sous-ensemble de thèses principales issues du résumé permet de croire en la justesse des résultats de ce sous-ensemble. De plus, les résultats obtenus à partir du sous-ensemble de thèses principales issues de la discussion sont justifiés, dans une moindre mesure, à la suite de l'inspection d'un petit échantillon. Enfin, nous avons testé nos hypothèses de recherche à propos de la détection des thèses principales sur les deux (2) sous-ensembles « résumé » et « discussion » afin de contrer les limitations inhérentes à chacun d'eux. Subséquemment, la confirmation (ou l'infirmité) d'une hypothèse sur les deux (2) sous-ensembles est plus crédible que sur un seul de ces sous-ensembles.

Les résultats de l'application des modèles de détection de thèses principales sur nos données textuelles montrent que cette tâche est complexe et difficile. D'une part, l'écart de performance entre un modèle fondé par le hasard et un modèle parfait montre comment notre tâche de classification sur notre jeu de

donnée est difficile. D'autre part, l'écart de performance entre les modèles plus simples et ceux ensemblistes montre la complexité de la tâche en jeu, notamment pour la macrostructure de discussion. Plusieurs caractéristiques apparaissent importantes pour la détection des thèses principales :

La représentation textuelle : Il ressort que les modèles basés sur la représentation textuelle DBOW (Distributed Bag of Words) sont généralement plus performants que ceux basés sur la représentation d'OpenAI. Cette constatation suggère que des modèles de représentation plus complexes ne garantissent pas nécessairement une meilleure performance. Une possible explication est l'usage du lemme (plutôt que le mot lui-même) pour construire la représentation DBOW, lequel accorde un poids plus important à la sémantique d'un mot en se débarrassant de ses inflexions. Néanmoins, il serait intéressant de tester cette hypothèse en explorant d'autres représentations textuelles.

La macrostructure d'introduction : L'introduction se révèle être une macrostructure significative pour la tâche de détection des thèses principales, tant dans les résumés que dans les discussions et ce, peu importe le type de représentation textuelle utilisé. Pour les discussions, cette méthode simple s'avère la plus performante après les modèles ensemblistes plus complexes.

Les autres structures ou macrostructures : Les énoncés de thèse principale entretiennent une relation de similarité envers d'autres structures que l'introduction, mais cela dépend de la provenance structurelle ou macrostructurelle ainsi que de la représentation textuelle utilisée. Les énoncés de thèse principale dans le résumé entretiennent une relation de similarité plus forte envers la macrostructure de discussion, et celles dans la discussion entretiennent une relation de similarité plus forte envers la structure du titre lorsque la représentation vectorielle utilisée est DBOW. Lorsque la représentation « text-embedding-ada-002 » est utilisée, les énoncés de thèses principales dans le résumé ainsi que dans la discussion entretiennent une relation de similarité plus forte envers la structure du résumé. Contrairement à l'introduction qui apparaît être la structure la plus informative pour la tâche qui nous intéresse, ces résultats ne permettent pas d'identifier clairement quelle est la seconde structure la plus informative pour détecter les thèses principales. Néanmoins, en se basant sur la mesure de rappel à $k=5$, un seul des modèles fondés par la similarité ne permet pas de réduire l'espace de recherche au-delà du modèle fondé par le hasard, c.-à-d. le modèle de détection des énoncés de thèse principale identifiés dans le résumé au moyen du résumé et de la représentation textuelle DBOW.

L'emplacement : L'emplacement se révèle être une information importante pour la tâche de détection des thèses principales, tant dans les résumés que dans les discussions. Cette approche démontre son efficacité, soulignant que les thèses principales ont tendance à être positionnées spécifiquement au début ou à la fin des textes. Pour les résumés, cette méthode s'avère la plus performante, surpassant même les modèles ensemblistes plus complexes. Plus précisément, les thèses principales tendent à se trouver plus fréquemment au début de la structure du résumé, et plus fréquemment à la fin de la macrostructure de discussion.

Les connecteurs de discours : Ces derniers servent d'indicateurs utiles pour la détection des énoncés de thèse principale. Il est noté que les connecteurs les plus performants ne sont pas les mêmes pour le résumé et la discussion. Seulement deux (2) connecteurs sont partagés par les thèses principales dans le résumé et dans la discussion soit « *earlier* » et « *hence* ». Lorsqu'on unit les deux (2) ensembles de connecteurs les plus performants, on obtient un ensemble de 16 connecteurs dont la majorité remplit la fonction d'expansion (conjonction, disjonction, etc.) ou de contingence (cause et effet), ce qui est tout à fait cohérent avec la notion de thèse principale. Parmi les modèles que nous avons testés, les modèles fondés par des connecteurs de discours sont les moins performants et cette performance varie significativement dépendamment du nombre de connecteurs utilisés. Ces résultats indiquent que l'approche des connecteurs de discours est suboptimale, lorsqu'utilisée seule du moins, mais cette approche peut être bénéfique lorsque couplée à d'autres approches puisque la présence (ou non) de certains connecteurs de discours peut faire la différence entre une prédiction avérée ou erronée. Toutefois, il serait intéressant d'appliquer ces méthodes sur un plus vaste corpus afin de vérifier si le nombre et l'identité des connecteurs dans les modèles les plus performants ne sont pas un artefact statistique résultant de la petitesse de notre corpus.

Les modèles ensemblistes : La combinaison de différentes méthodes dans un modèle ensembliste offre des performances optimales pour la détection des thèses dans la discussion, intégrant les informations complémentaires des différentes approches. Cependant, pour les résumés, le modèle simple basé sur l'emplacement reste le plus performant. Pour les discussions, l'écart de performance entre les modèles fondés sur une seule propriété et ceux ensemblistes fondés par plusieurs propriétés montre qu'une modélisation efficiente de cette tâche implique la prise en compte de plusieurs caractéristiques propres aux thèses principales. Parmi les propriétés explorées, celles qui apparaissent plus spécifiques aux thèses principales sont, en ordre d'importance, i) une relation de similarité envers certaines structures ou

macrostructures IMRAD (l'introduction et le titre); ii) un emplacement privilégié dans une structure ou une macrostructure (au début et à la fin) et iii) la présence plus marquée de certains connecteurs de discours (ceux exprimant des relations de contingence ou d'expansion notamment). Spécifions que ces caractéristiques sont observées également lors de la détection des énoncés de thèse principale dans la structure du résumé, mais l'emplacement constitue la caractéristique offrant la meilleure performance pour cette structure, et cette performance se dégrade lorsqu'on ajoute des caractéristiques pour former un modèle ensembliste. Spécifions également que la similarité entre la macrostructure d'introduction et la thèse principale est une caractéristique observée, peu importe où elle se trouve (dans le résumé ou la discussion) ou la représentation utilisée (DBOW ou « text-embedding-ada-002 »).

En somme, l'emplacement semble être une information suffisante afin de réduire l'espace de recherche menant à la détection des énoncés de thèse principale dans les résumés. *A contrario*, la réduction de l'espace de recherche pour la détection des énoncés de thèse principale dans les discussions implique la combinaison de plusieurs approches, soit l'utilisation de structure ou de macrostructure spécifique, la prise en compte de la position de la thèse principale, et l'intégration des connecteurs de discours. Les recherches futures pourraient explorer davantage les modèles ensemblistes, combinant les forces des différentes méthodes présentées ici. De plus, il serait intéressant d'explorer d'autres formes de représentations textuelles et d'examiner leur efficacité relative dans le contexte de cette tâche.

10.4 Les faiblesses et les limites de la recherche

Cette étude, se concentrant sur la modélisation pour identifier les contextes de citation dans les structures IMRAD et pour détecter les thèses principales dans les textes, révèle plusieurs faiblesses et limites significatives.

Premièrement, les techniques d'analyse argumentaire utilisées pour ces tâches montrent des performances inférieures à un modèle parfait⁹⁸, mettant en lumière le besoin d'améliorations futures. Cette situation souligne la complexité de l'argumentation, un domaine qui reste difficile à appréhender ou à détecter entièrement par les méthodes informatiques actuelles. En particulier, la détection des thèses principales s'avère complexe, comme en témoigne l'écart entre les performances des modèles simples et ensemblistes, surtout dans les discussions. Bien que les modèles basés sur la représentation textuelle

⁹⁸ Aucun des modèles testés n'a permis d'identifier toutes les thèses principales, ni l'appartenance de tous les contextes de citation aux macrostructures IMRAD.

DBOW semblent supérieurs à ceux basés sur « text-embedding-ada-002 », ils ne garantissent pas forcément une meilleure compréhension de la tâche de détection des thèses principales. En outre, il est important de noter que l'emplacement des thèses principales dans les textes est une information importante, mais que l'efficacité de cette approche est limitée. Pour les résumés, cette méthode est la plus performante, mais elle l'est moins dans le cas des discussions, où combiner différentes approches s'avère nécessaire pour obtenir de meilleurs résultats.

Deuxièmement, la généralisation des modèles et méthodes développés est limitée. Les résultats obtenus à partir d'un corpus spécifique peuvent ne pas s'appliquer universellement à tous les TECHNETIUM, c.-à-d. issus de tous les domaines philosophiques ou scientifiques. D'une part, le corpus InTeReC est composé de contextes de citation issus d'articles scientifiques de divers domaines, mais pas en philosophie. D'autre part, le corpus AbstRCT+ est composé d'articles scientifiques issus exclusivement du domaine pointu de la médecine. Bref, la question de l'applicabilité générale de nos résultats demeure, surtout si l'on considère l'utilisation de corpus spécifiques pour l'entraînement et l'évaluation de nos modèles. Parallèlement, il n'est pas évident que tous les TECHNETIUM suivent une structuration IMRAD. D'une part, il est possible que la grande majorité des TECHNETIUM contiennent des sections textuelles remplissant des fonctions IMRAD. D'autre part, ces sections textuelles remplissant certaines fonctions IMRAD ne sont possiblement pas subdivisées, titrées ou ordonnées de manière uniforme. Subséquemment, il sera donc essentiel de disposer d'un corpus plus large et diversifié, idéalement comprenant des TECHNETIUM de différents domaines scientifiques et philosophiques, suivant explicitement la formule IMRAD ou non. Des analyses transversales à l'intérieur d'un tel corpus permettraient de déterminer si les apprentissages sont transférables d'un domaine à l'autre, ainsi que d'une forme structurale à l'autre.

Troisièmement, concernant l'identification de l'appartenance des contextes de citation aux macrostructures IMRAD, il convient de rappeler que notre corpus InTeReC se concentre uniquement sur un sous-ensemble de contextes de citation, à savoir ceux qui sont simples (ne contenant qu'une seule citation) et directs (qui citent directement de manière non superficielle). L'extension de nos recherches pour inclure des contextes de citation plus complexes et indirects pourrait fournir une évaluation plus complète de la performance de nos approches.

Quatrièmement, le fait que seulement environ 16% des résumés du corpus AbstRCT contiennent des thèses principales est intrigant. Si les thèses identifiées sont effectivement principales, cela ne remet pas

en cause la portée de notre étude. Cependant, vu que la classe des thèses principales est minoritaire dans ce corpus, on peut se demander s'il n'existe pas un nombre significatif de faux négatifs ou de vrais négatifs, ce qui nécessiterait une enquête plus approfondie de ce phénomène. Une telle analyse pourrait être déterminante pour une approche de détection des énoncés de thèse principale dans la discussion au moyen des énoncés de thèse principale dans le résumé identifiés de manière automatique.

Cinquièmement, les limites de la notion de thèse principale méritent d'être soulignées. La thèse principale peut être définie comme une thèse soutenue par des évidences ou des thèses secondaires, sans supporter ni attaquer d'autres thèses. Dans les textes structurés IMRAD, elle peut également être vue comme une solution à une question de recherche. La première définition souligne un aspect structural vertical de la thèse principale. Elle est conséquente avec la théorie communicationnelle hiérarchique de l'argumentation définissant celle-ci en termes d'articulations de raisons qui soutiennent et de raisons qui s'opposent à une thèse principale. La seconde définition souligne un aspect structural horizontal de l'argumentation. Elle est conséquente avec la théorie communicationnelle IMRAD de l'argumentation définissant celle-ci comme un phénomène de communication se réalisant dans une forme linguistique linéaire pouvant prendre une forme IMRAD. Dans le cas des textes de forme IMRAD, ces deux définitions, bien que non contradictoires, présentent des problèmes lorsqu'elles sont prises séparément dans la mesure où chacune pourrait inclure des éléments qui ne sont généralement pas considérés comme des thèses principales. Par exemple, certains arguments d'un discours peuvent supporter une thèse qui ne supporte ni n'attaque d'autres thèses, mais, suivant une attaque en règle (ou non), celle-ci peut être abandonnée en cours de route. Une telle thèse laissée à l'abandon, sans relation de support ou d'attaque envers d'autres thèses, n'est pas une thèse principale, mais plutôt une thèse périphérique. La seconde définition semble aussi inclure des éléments qui ne sont pas considérés habituellement comme des thèses principales, notamment celles répondant à une question de recherche, mais soutenant ou attaquant d'autres thèses par la même occasion. Dans le corpus AbstrRCT, par exemple, Mantovani *et al.* (2002) cherchent à déterminer l'efficacité d'un traitement en phase II de la recherche médicale. Leur réponse est que le traitement investigué est plus ou moins efficace et particulièrement toxique. Or, cette réponse leur permet de conclure que ce traitement ne doit pas faire l'objet d'une recherche de phase III, ce qui constitue une thèse de plus haut niveau qui ne répond pas à la question de recherche, et dont la nature est plus philosophique que scientifique. Bref, une manière de résoudre ces problèmes serait d'attribuer à la thèse principale des propriétés structurales à la fois verticales et linéaires. Ce faisant, la thèse principale serait une thèse qui ne soutient ni n'attaque d'autres thèses, tout en constituant une réponse à une

question de recherche. Subséquemment, nous recommandons une telle définition pour les études futures appliquées sur des textes orientés vers la résolution d'une problématique, notamment les textes structurés IMRAD.

En somme, notre recherche met en lumière diverses contraintes associées à la modélisation des contextes de citation suivant la structure IMRAD et à l'identification des thèses principales. Elle met en évidence la complexité inhérente à l'argumentation et les défis rencontrés par les technologies informatiques actuelles pour appréhender et détecter entièrement ces aspects, en particulier pour la détection des thèses principales. De plus, les résultats soulignent des restrictions concernant la portée de nos conclusions, nécessitant l'utilisation d'un corpus plus étendu pour une application plus large. Enfin, notre étude attire l'attention sur les défis liés à la définition de la thèse principale, soulignant le besoin d'une définition plus détaillée et intégrée. Nous proposons de considérer la thèse principale sous un angle double : a) verticalement, comme un élément clé d'un argument qui ne supporte ni ne conteste d'autres thèses, et b) horizontalement, comme une réponse à une question de recherche spécifique. Cette perspective duale pourrait guider les futures recherches pour une analyse plus précise, en particulier dans les textes axés sur la résolution de problèmes spécifiques.

10.5 Les avenues et recherches ultérieures

Les résultats de notre recherche sur la détection de l'appartenance des contextes de citation aux macrostructures IMRAD et la détection des thèses principales dans les TECHNETIUM offrent plusieurs pistes pour des recherches futures.

Une première avenue est l'amélioration des techniques de modélisation, c.-à-d. développer et tester de nouvelles méthodes pour améliorer la précision et le rappel des modèles existants. Une telle évolution pourrait inclure l'intégration d'approches déductives, telles que l'application de règles spécifiques, aux modèles en place. Conformément aux suggestions de Perier-Camby et coll. (2019), un premier pas consisterait à évaluer et comparer les méthodes de classification ascendantes (*bottom-up*) et descendantes (*top-down*), ces dernières se basant sur des règles ou des connecteurs de discours, particulièrement pour identifier l'appartenance de contextes de citation aux structures IMRAD. Une autre piste prometteuse est la création de modèles hybrides qui combinent les caractéristiques ascendantes de l'apprentissage automatique avec des éléments descendants. Dans cette optique, l'amélioration des modèles existants pourrait être réalisée grâce à l'emploi de techniques avancées d'apprentissage profond

ou l'exploration de nouvelles architectures neuronales, y compris, mais sans s'y limiter, les réseaux neuronaux convolutifs (*Convolutional Neural Network* ou *CNN*), les réseaux neuronaux récurrents (*Recurrent Neural Network* ou *RNN*), les différentes variantes de BERT, les modèles génératifs adversaires (*Adversarial Generative Network* ou *GAN*), les autoencoders variationnels (*Variational AutoEncoder* ou *VAE*), l'apprentissage par renforcement, les réseaux neuronaux capsules (*Capsule neural Network* ou *CapsNet*), les techniques d'augmentation de données, et les approches hybrides. Les *CNN* sont connus pour leur efficacité dans le traitement des images, et peuvent également être appliqués aux données textuelles pour capturer les motifs spatiaux dans les séquences de mots, ce qui pourrait être utile pour identifier les structures et les motifs argumentatifs (Li, 2017). Les *RNN* peuvent être efficaces pour traiter des séquences de données, comme le texte, en tenant compte de leur nature séquentielle et de la dépendance contextuelle entre les mots (Rezaeenour *et al.*, 2023). Il existe plusieurs variantes de BERT que nous n'avons pas testé et qui pourrait s'avérer efficace pour les tâches qui nous intéressent (Chakkarwar *et al.*, 2023). Parmi ces variantes, certaines pourraient être adaptées ou affinées pour mieux détecter les structures spécifiques des textes scientifiques et philosophiques. Les *GAN*, principalement utilisés pour la génération d'images, pourraient être adaptés pour générer des textes ou des structures de texte, aidant à mieux comprendre et modéliser les schémas argumentatifs complexes (Jiang *et al.*, 2023). Les *VAE* pourraient aussi être utilisés pour apprendre des représentations denses et informatives des textes, ce qui pourrait améliorer la classification et la détection des structures de texte (Che *et al.*, 2020). L'apprentissage par renforcement, notamment utilisé pour entraîner ChatGPT, pourrait être utilisé plus spécifiquement pour affiner les modèles de classification ou d'annotation de texte en vue d'une analyse argumentaire (Wang *et al.*, 2022). Les *CapsNet* constituent une architecture relativement nouvelle qui pourrait offrir une meilleure compréhension hiérarchique et spatiale des données textuelles, potentiellement utiles pour analyser la structure des documents (Hinton *et al.*, 1999 ; Jia et Wang, 2022). Les techniques d'augmentation de données, comme la traduction automatique ou la paraphrase pour augmenter le corpus d'entraînement, pourraient permettre aux modèles d'apprendre à partir d'un ensemble de données plus diversifié et d'améliorer potentiellement leur généralisabilité (Nanthini *et al.*, 2023). Enfin, les approches hybrides permettent de combiner différentes architectures et techniques de modélisation pour exploiter les avantages de chaque méthode, notamment pour capturer à la fois la séquentialité et les caractéristiques locales du texte. En principe, il est possible d'améliorer la capacité des modèles à comprendre et à classer avec précision les textes complexes, en tenant compte de leurs structures argumentatives et thématiques spécifiques.

Une seconde avenue est l'exploration de représentations vectorielles alternatives. Il s'agit d'expérimenter avec diverses formes de représentations vectorielles, comme des modèles linguistiques plus complexes ou des variantes de représentations basées sur des plongements de mots, pour évaluer leur efficacité dans la détection des thèses principales et l'appartenance aux macrostructures IMRAD. Quelques exemples de ces représentations et modèles pourraient être basés sur des variantes des modèles basés sur les architectures de type Transformer, qu'elles soient plus récentes et sophistiquées (GPT-3 Turbo ou GPT-4 Turbo par exemple) ou plus anciennes, mais possiblement mieux contextualisées ou spécifiques au domaine investigué (mBERT ou SciBERT par exemple). Il s'agit, donc, de développer ou d'utiliser des représentations textuelles qui sont mieux contextualisées ou spécifiquement entraînées sur des corpus relatifs à des domaines particuliers, comme les sciences ou la philosophie, pour voir si ces représentations améliorent la performance des modèles sur des tâches spécifiques. Une autre approche est d'expérimenter avec des représentations basées sur des graphes pour capturer les relations complexes entre les éléments d'un texte, ce qui peut être particulièrement utile pour analyser des structures argumentatives et des connexions entre différentes parties d'un texte (Etaiwi et Awajan, 2023 ; Lu *et al.*, 2020). Aussi, notons que VAE permet des représentations denses et informatives des données et que CapsNet permet de modéliser la hiérarchie spatiale de caractéristiques dans les données, ce qui pourrait être bénéfique pour comprendre les structures argumentatives complexes. Enfin, précisons que les techniques d'augmentation de données peuvent mener à des représentations mieux contextualisées. Il s'agit donc de tester l'impact de l'augmentation de données sur les représentations vectorielles, comme la génération de textes synthétiques ou la modification des textes existants, pour enrichir le corpus d'entraînement et améliorer la robustesse des modèles. L'objectif de ces explorations est de déterminer si ces techniques avancées ou spécifiques de représentation peuvent améliorer la compréhension et la classification des éléments textuels dans le cadre de tâches telles que la détection des thèses principales et l'analyse des structures IMRAD.

Une troisième avenue est l'extension et la diversification des corpus d'étude. Il s'agit d'utiliser des corpus de texte plus larges et variés pour tester la généralisabilité des modèles. On pourrait inclure des textes de différents domaines académiques, des styles de rédaction variés, et des structures autres que IMRAD pour examiner l'efficacité des modèles dans divers contextes. L'extension et la diversification des corpus impliquent d'intégrer des textes variés provenant de multiples sources et domaines. Voici quelques exemples de types de textes et de sources qui pourraient être inclus dans un corpus élargi. D'abord, des textes académiques de divers domaines, des articles de recherche, des dissertations ou des publications

de divers domaines tels que les sciences sociales, les sciences naturelles, la technologie, l'ingénierie, les mathématiques, la médecine, mais aussi des domaines possiblement moins structurés comme la philosophie, les arts et les humanités. Ensuite, on pourrait intégrer des articles de presse, des blogues, des essais et d'autres formes de littérature non académique pour observer comment les modèles gèrent différents styles de rédaction et structures argumentatives. De plus, on pourrait combiner des textes historiques avec des publications contemporaines pour tester la capacité des modèles à interpréter des variations de langage et de contexte au fil du temps. Aussi, il serait possible d'inclure des textes dans différentes langues pour évaluer la capacité des modèles à comprendre et classifier l'argumentation et les structures textuelles dans un contexte multilingue. Enfin, il serait possible d'élargir le corpus de recherche aux documentations techniques, rapports scientifiques, textes juridiques et législatifs, textes de conférence et présentation, ainsi que les forums en ligne et médias sociaux. L'utilisation des manuels techniques, des rapports de recherche et développement, et d'autres documents techniques permettrait de tester les modèles sur des textes avec un jargon spécialisé et des structures complexes. L'utilisation de documents juridiques, de lois, d'arrêts de cours ou de documents législatifs permettrait d'évaluer la performance des modèles dans la compréhension de ce type de textes avec un langage formel et une structure rigoureuse. L'utilisation de transcriptions de conférences, de présentations ou de discours permettrait d'observer comment les modèles traitent la langue parlée et la structure argumentative dans un contexte moins formel. Enfin, l'utilisation d'extraits de forums en ligne, de commentaires sur les réseaux sociaux et de blogues permettrait de tester la capacité des modèles à comprendre et analyser des textes informels et souvent moins structurés. Bref, en intégrant une gamme aussi large et variée de textes, il est possible de tester la robustesse et la flexibilité des modèles de traitement du langage naturel et de mieux comprendre leurs limites et leurs domaines d'efficacité. Cela permet également d'évaluer la capacité des modèles à généraliser à travers différents types de textes et de structures, et identifier les adaptations nécessaires pour les rendre plus universellement applicables.

Une quatrième avenue consiste en l'analyse des faux positifs et des faux négatifs. Elle vise à identifier et à comprendre les erreurs commises par les modèles, afin de les corriger ou de les minimiser dans les itérations futures. Une telle analyse peut prendre la forme d'apprentissage actif, de techniques d'explication, de revue du processus d'annotation, d'amélioration des caractéristiques, d'ajustement des seuils de décision ou du retour d'information continu. L'apprentissage actif (AA) est une approche pertinente pour améliorer la précision des modèles de traitement automatique du langage, notamment dans le contexte de la réduction des faux positifs et des faux négatifs. AA est une technique

d'apprentissage automatique où le modèle identifie activement les cas où il est le moins certain de ses prédictions, puis demande des éclaircissements ou des étiquettes supplémentaires pour ces cas spécifiques (Prince, 2004). Les zones où le modèle est le moins sûr peuvent indiquer des domaines spécifiques ou des limites de partitionnement nécessitant des améliorations ou des ajustements dans les caractéristiques du modèle. L'interaction subséquente entre les modèles d'apprentissage machine et les experts humains permet une validation et une correction précises des données annotées. Bref, AA peut suggérer l'analyse détaillée de certains cas d'erreurs ou l'étude approfondie de cas spécifiques. À cet effet, les techniques d'explication peuvent aider à mieux comprendre pourquoi un modèle prend certaines décisions. Certaines de ces techniques, *Local Interpretable Model-Agnostic Explanations (LIME)* et *SHapley Additive exPlanations (SHAP)* notamment (Zafar et Khan, 2021), peuvent aider à décomposer les prédictions d'un modèle et identifier les facteurs qui contribuent le plus aux erreurs. Autrement, puisque des erreurs ou des incohérences dans les données annotées peuvent entraîner de faux positifs et des faux négatifs, une revue des processus d'annotation peut être utile afin de vérifier la qualité des annotations utilisées pour l'entraînement des modèles. Aussi, l'amélioration des caractéristiques des modèles permet de modifier ou d'ajouter des caractéristiques (*features*) qui pourraient aider à mieux distinguer les vrais positifs des faux positifs et les vrais négatifs des faux négatifs. Cela pourrait inclure des caractéristiques linguistiques plus fines ou des indicateurs contextuels. Une autre approche est l'ajustement des seuils de décision afin de trouver un équilibre optimal entre la réduction des faux positifs et la minimisation des faux négatifs. Enfin, le retour d'information continu est une technique qui consiste à implémenter un système de rétroaction dans un modèle interactif où les utilisateurs peuvent signaler des erreurs. Par la suite, cette information peut être utilisée pour améliorer continuellement le modèle. En bref, il est possible de réduire significativement le nombre de faux positifs et de faux négatifs, améliorant ainsi la précision et la fiabilité des modèles d'analyse d'arguments.

Une cinquième avenue envisage une exploration plus poussée des modèles ensemblistes, qui fusionnent diverses méthodologies et algorithmes pour renforcer leur efficacité globale. Notre recherche a déjà mis en lumière la pertinence de l'approche ensembliste basée sur la votation, spécialement dans l'identification des thèses principales au sein des discussions. Cependant, le potentiel des modèles ensemblistes ne se limite pas à cela et englobe l'hybridation de modèles simples (comme les SVM, arbres de décision, ou modèles de régression logistique) ou de modèles plus élaborés (comme les CNN, les RNN ou ceux basés sur des Transformers), ainsi que des combinaisons de modèles simples et complexes. Ces mélanges peuvent intégrer des techniques variées telles que le *boosting*, le *bagging* ou le *stacking*. Un

exemple concret d'hybridation consiste à exploiter les réseaux neuronaux pour l'extraction de caractéristiques textuelles, suivie d'une phase de classification à l'aide de modèles traditionnels. Cette méthode s'est révélée efficace pour détecter les contextes de citation relatifs aux structures IMRAD. De plus, les modèles peuvent être entraînés de manière séquentielle (*boosting*) ou parallèle (*bagging*). Autrement, il est possible d'empiler les modèles (*stacking*), les sorties de l'un servant ainsi d'entrées pour l'autre, de manière à synthétiser les informations pertinentes et améliorer la performance globale de la classification. Enfin, d'autres hybridations sont possibles en intégrant des analyses sémantiques ou syntaxiques. D'une part, le couplage de l'analyse sémantique, comme l'analyse des sentiments ou la reconnaissance d'entités nommées, avec des modèles de classification traditionnelle pourrait enrichir la compréhension du contexte et du contenu. D'autre part, la combinaison de l'analyse syntaxique, de manière à structurer le texte, pourrait permettre une meilleure compréhension de la structure argumentative. D'ailleurs, cette approche s'est révélée efficace pour l'identification des thèses principales, notamment l'identification préalable de la macrostructure d'introduction et de discussion. En combinant ces différentes méthodes et modèles, les approches ensemblistes peuvent fournir des résultats plus robustes et précis lorsque différents aspects comme la structure du texte, la sémantique et le contexte doivent être pris en compte.

Une sixième avenue vise à approfondir l'impact des connecteurs de discours dans la détection des thèses principales. Une première étape consiste à réaliser une analyse quantitative pour identifier quels connecteurs de discours sont les plus fréquemment associés aux thèses principales dans un corpus plus diversifié que celui utilisé dans notre étude. Le cas échéant, il serait intéressant de mener des études de cas spécifiques à différents domaines scientifiques et philosophiques, ou encore différents genres (articles, essais, etc.) pour observer comment les connecteurs de discours sont utilisés pour introduire ou conclure les thèses principales. La comparaison entre différents domaines ou styles de rédaction pourrait révéler des modèles d'utilisation de ces connecteurs. Subséquemment, il serait possible de développer des stratégies pour optimiser l'utilisation des connecteurs de discours dans les modèles de détection de thèses. Cela pourrait inclure des techniques pour pondérer différemment certains connecteurs en fonction du contexte ou du type de texte analysé. Ces exemples illustrent comment l'étude de l'impact des connecteurs de discours peut être abordée sous différents angles pour enrichir la compréhension et l'efficacité des modèles de détection de thèses principales.

Une septième avenue est le développement de techniques d'annotation automatique pour faciliter l'expansion des corpus d'entraînement et de test, en particulier pour les thèses principales. Plusieurs techniques à cet effet sont possibles soit les techniques à base de règles, l'apprentissage semi-supervisé, l'annotation collaborative, l'apprentissage profond, le transfert d'apprentissage et le *crowdsourcing*. L'annotation basée sur des règles, comme son nom l'indique, utilise des règles linguistiques spécifiques pour identifier et marquer automatiquement les thèses principales dans les textes. Ces règles peuvent être basées sur la syntaxe, la sémantique, et l'utilisation de certains mots-clés ou phrases qui sont typiquement associés aux thèses principales. L'annotation basée sur l'apprentissage semi-supervisé utilise une petite partie du corpus annotée manuellement, et ces annotations sont ensuite utilisées pour entraîner un modèle qui peut annoter automatiquement le reste du corpus. L'annotation collective implique le développement de plateformes où les humains et les modèles d'IA collaborent pour l'annotation des textes. Les utilisateurs humains peuvent valider ou corriger les annotations proposées par l'IA, améliorant ainsi la précision du modèle au fil du temps. L'annotation basée sur l'apprentissage profond utilise des modèles neuronaux profonds (CNN, RNN, BERT, GPT, etc.) pour identifier les caractéristiques des thèses principales et annoter automatiquement les textes. L'annotation basée sur le transfert de connaissance (*transfer learning*) implique un modèle préentraîné sur une tâche similaire pour l'affiner sur la tâche spécifique d'annotation qui est visée. Enfin, l'annotation par *crowdsourcing* utilise des plateformes spécialisées à cet effet pour collecter des annotations à grande échelle. Ces annotations peuvent ensuite être utilisées pour entraîner ou valider des modèles d'annotation automatique. Ces méthodes et techniques représentent divers moyens par lesquels les techniques d'annotation automatique peuvent être développées et affinées pour améliorer la création et l'expansion des corpus d'entraînement et de test, en se concentrant spécifiquement sur les thèses principales dans divers contextes textuels.

Une huitième avenue est l'intégration des composantes argumentatives explorées dans cette étude pour une analyse plus approfondie des arguments. L'identification des contextes de citation liés aux structures et macrostructures IMRAD ainsi que la détection de thèses principales jouent un rôle crucial en analyse d'argument, notamment pour comprendre les interconnexions entre différents textes. Cette compréhension peut être approfondie au moyen de l'analyse de la provenance et de l'influence d'une idée ou d'un argument, la cartographie des réseaux intertextuels d'arguments, et la revue de littérature automatique.

L'analyse des contextes de citation et des thèses principales dans les textes académiques offre une compréhension approfondie de la portée et de l'impact des idées dans le discours philoscience. Cette approche permet de déterminer où une idée est mentionnée (dans quels textes et dans quelles sections IMRAD), ce qui révèle son rôle et son importance. L'évaluation de la fréquence des contextes des citations où une idée est exprimée aide à comprendre son influence et sa pertinence dans un champ de recherche. Les différentes sections d'un document (introduction, méthodes, résultats, discussion) jouent des rôles distincts, influençant la façon dont les idées y sont intégrées. Parallèlement, l'identification des thèses principales d'un texte permet de mesurer l'importance et la pertinence d'un argument dans un domaine spécifique. L'analyse des citations de thèse principale dans d'autres travaux révèle leur impact dans la communauté scientifique et permet de mesurer l'influence des arguments l'étayant. En principe, cette approche permet une cartographie détaillée de l'évolution des idées et des tendances de recherche dans différents domaines. D'une part, il est alors possible de comparer comment différentes études abordent une question similaire, utilisent des méthodes différentes, arrivent à des résultats divergents ou interprètent les données de manière variée. D'autre part, il est aussi possible de créer un réseau de références croisées, facilitant la navigation et la compréhension des liens thématiques entre les recherches.

La cartographie des réseaux intertextuels d'arguments dans les textes académiques est une approche analytique puissante qui permet de comprendre comment les idées et arguments circulent et interagissent à travers différents travaux. Cette approche implique l'identification des thèses principales et des contextes de citation (ainsi que leur provenance), la cartographie des relations entre ces composantes, la représentation visuelle du réseau d'arguments, puis son interprétation. La première étape consiste à identifier les thèses principales dans les textes et les contextes de citation associés aux structures IMRAD (Introduction, Méthodes, Résultats, et Discussion). Cela implique de reconnaître où et comment une idée est citée ou discutée dans un texte académique. Une fois ces éléments identifiés, il devient possible de tracer des liens entre les différentes thèses et citations. Cette cartographie révèle comment une idée évolue, est reprise, modifiée ou contestée à travers différents documents. En utilisant des outils de visualisation, on peut créer des réseaux intertextuels qui montrent les interconnexions entre divers textes. Chaque nœud dans ce réseau représente une thèse ou une citation, tandis que les liens montrent les relations entre eux. À travers cette cartographie, on peut observer des tendances, comme quels arguments sont les plus influents, quels sujets sont fréquemment discutés, ou comment les idées se propagent à travers la communauté scientifique. D'une part, cette approche permet de mieux comprendre comment les chercheurs interagissent entre eux à travers leurs travaux. Elle montre comment les idées

sont appuyées ou contestées, formant un dialogue académique continu. D'autre part, on peut identifier dans ce type de réseau des zones moins explorées ou des questions restées sans réponse, suggérant de nouvelles directions de recherche. On peut également évaluer l'impact d'un travail de recherche spécifique en mesurant sa présence et son influence dans le réseau d'arguments global. En bref, la cartographie des réseaux intertextuels d'arguments est un outil puissant pour analyser le paysage complexe de la communication scientifique, révélant les dynamiques sous-jacentes de la diffusion et de l'évolution des connaissances.

L'analyse des contextes de citation et des thèses principales, notamment dans le cadre des structures IMRAD, présente un potentiel significatif pour le développement de la génération automatique de revue de littérature (*automatic review article generation*), un sous-domaine émergeant du TALN (Nanba *et al.*, 2000). Cette approche pourrait transformer fondamentalement la manière dont les revues de littérature sont construites, en offrant une méthode structurée et automatisée pour synthétiser les recherches existantes. La rédaction d'une revue de littérature est un processus manuel et intensif, nécessitant la lecture et l'analyse de nombreux articles pour synthétiser les principaux résultats et tendances. En automatisant ce processus, on peut économiser un temps considérable, tout en assurant une couverture exhaustive des travaux pertinents dans un domaine donné. Les structures IMRAD (Introduction, Méthodes, Résultats et Discussion) fournissent un cadre idéal pour structurer automatiquement les revues de littérature. En identifiant les thèses principales et les citations clés dans ces sections, on peut extraire et synthétiser les informations les plus pertinentes de chaque article. L'identification des thèses principales permet de comprendre rapidement les points clés de chaque étude. Cela aide à construire un résumé cohérent qui capture l'essence des recherches sans entrer dans les détails techniques. En examinant comment et où les articles sont cités dans d'autres travaux, il est possible de déterminer leur impact et leur pertinence dans le champ de recherche. Cela aide à identifier les études les plus influentes et les tendances émergentes dans un domaine spécifique. En combinant les informations extraites des thèses principales et des contextes de citation, il est possible de générer automatiquement des synthèses bien organisées. Ces synthèses peuvent offrir une vue d'ensemble complète d'un champ de recherche, mettant en évidence les principaux résultats, les méthodologies utilisées, et les conclusions tirées par différents chercheurs. De plus, cette approche peut également être adaptée pour répondre à des besoins spécifiques. Par exemple, une revue de littérature peut être orientée vers des aspects méthodologiques, des résultats spécifiques, ou des discussions théoriques, en fonction des besoins particuliers. Enfin, un autre avantage est la capacité à mettre à jour les revues de littérature automatiquement. Alors que les revues

traditionnelles deviennent rapidement obsolètes, un système automatisé peut intégrer continuellement de nouvelles recherches, assurant ainsi une pertinence et une actualité constantes. En bref, l'automatisation des revues de littérature via l'analyse des contextes de citation et des thèses principales pourrait révolutionner la manière dont les connaissances académiques sont compilées et partagées, offrant des synthèses plus rapides, plus complètes et constamment à jour des avancées dans divers domaines de recherche.

La neuvième et dernière avenue envisagée est l'intégration des différentes approches de la lecture et analyse de texte assistées par ordinateur (LATAO) soit l'analyse de concept, l'analyse de thème et l'analyse d'argument. Rappelons que cette thèse de doctorat s'inscrit dans un programme de recherche plus vaste en LATAO. Les résultats de notre programme de recherche indiquent de nouvelles pistes ou questions de recherche. Bien que nos résultats aient permis une meilleure compréhension de la LATAO, notamment ses objets d'analyses (concepts, thèmes et arguments), à notre connaissance aucune application n'implique ces trois objets d'analyse dans une même LATAO, car les applications ont tendance à être spécifiquement thématiques, conceptuelles ou argumentaires. L'intégration des différentes approches de la LATAO représente un développement prometteur dans le domaine du TALN. Cette intégration vise à combiner l'analyse de concepts, de thèmes et d'arguments dans un seul cadre computationnel, offrant une compréhension plus holistique et nuancée des textes. La combinaison de l'analyse de concepts, de thèmes et d'arguments implique une plateforme d'analyse de texte multidimensionnelle. Cela signifie que le texte sera examiné sous plusieurs angles simultanément, offrant une vision plus complète de son contenu. Pour atteindre cet objectif, il est nécessaire de développer des modèles computationnels capables de traiter et d'analyser simultanément ces différents aspects du texte. Cela peut impliquer l'utilisation de techniques d'apprentissage profond, d'analyse sémantique ou syntaxique, et de modélisation des données textuelles. L'intégration de différentes méthodes en vue d'une analyse multidimensionnelle ouvre de nouvelles pistes de recherche dans le domaine du TALN. Elle pose de nouvelles questions sur la manière dont les concepts, thèmes et arguments interagissent et se manifestent dans les textes. En combinant ces différentes méthodes d'analyse, il est possible d'obtenir une compréhension plus profonde et plus nuancée des textes. Cela peut être particulièrement bénéfique pour des domaines tels que l'analyse de discours, la recherche d'information, l'éducation, etc. La mise en œuvre d'une telle approche nécessite de surmonter des défis techniques et théoriques, notamment la gestion de la complexité liée à la combinaison de différentes méthodes d'analyse et la nécessité d'une compréhension approfondie de chacun de ces domaines. En bref,

l'intégration de l'analyse de concepts, de thèmes et d'arguments représente une avancée vers une analyse textuelle plus complète, capable de saisir la richesse et la complexité des textes dans leur intégralité.

Enfin, les résultats de notre programme de recherche indiquent de nouvelles pistes ou questions de recherche. Bien que nos résultats aient permis une meilleure compréhension de la LATAO, notamment ses objets d'analyses (concepts, thèmes et arguments), à notre connaissance, aucune application n'implique ces trois objets d'analyse dans une même LATAO (les applications ont tendance à être spécifiquement thématiques, conceptuelles ou argumentaires). Subséquemment, une meilleure compréhension théorique et pratique de ces trois objets d'analyse semble nécessaire afin de permettre leur modélisation computationnelle au sein d'une même chaîne de traitement en vue d'une analyse conjointe.

Les avenues et recherches ultérieures identifiées à partir de notre étude sur la détection des contextes de citation dans les structures IMRAD et la détection des thèses principales dans les TECHNETIUM ouvrent la voie à d'importantes avancées dans le domaine du traitement automatique du langage naturel (TALN). Ces avancées se concentrent principalement sur l'amélioration des techniques de modélisation, l'exploration de nouvelles représentations vectorielles, l'extension et la diversification des corpus d'étude, l'analyse détaillée des faux positifs et des faux négatifs, l'utilisation de modèles ensemblistes, l'étude de l'impact des connecteurs de discours, le développement de techniques d'annotation automatique, l'analyse intégrée des contextes de citation et des thèses principales, et l'intégration des différentes approches de lecture et analyse de texte assistées par ordinateur (LATAO). Ces recherches visent à améliorer la précision et l'efficacité des modèles de TALN, enrichissant ainsi notre compréhension des structures argumentatives et thématiques dans les textes académiques.

CONCLUSION

La synthèse de nos expériences en matière de détection de l'appartenance des contextes de citation aux macrostructures IMRAD et de la détection des énoncés de thèse principale dans les TECHNETIUM offre une confirmation significative de nos hypothèses initiales.

Concernant notre première hypothèse (H1), qui postulait la possibilité d'identifier l'appartenance des contextes de citation aux macrostructures IMRAD par diverses techniques de représentation et de classification, les résultats ont été positifs. L'application de 153 modèles différents, combinant diverses représentations vectorielles et classifieurs, a démontré une capacité notable à distinguer les contextes de citation selon les sections IMRAD. Des modèles comme BERT affinés spécifiquement pour cette tâche ont montré des performances élevées, confirmant ainsi l'efficacité de nos approches de modélisation computationnelle dans l'identification des macrostructures IMRAD. Voici un récapitulatif de nos conclusions pour les sous-hypothèses relatives à l'identification de l'appartenance des contextes de citation aux macrostructures IMRAD, c.-à-d. que les techniques de représentation comportant plus de paramètres comme « text-embedding-ada-002 » sont plus performantes (H1.1) et que les méthodes de classification linéaire sont plus performantes (H1.2):

H1.1 (Représentation Comportant plus de Paramètres) : Cette hypothèse a été partiellement validée. Les techniques de représentation intégrant un grand nombre de paramètres, telles que text-embedding-ada-002 », ont montré une efficacité notable. Cependant, le modèle BERT, bien que possédant moins de paramètres que « text-embedding-ada-002 », mais ayant la particularité d'être affiné sur des données spécifiques à la tâche, a surpassé les autres techniques de représentation, indiquant que la quantité de paramètres n'est pas le seul facteur déterminant la performance. Le processus d'affinage joue donc un rôle tout aussi important.

H1.2 (Classification Linéaire) : Cette hypothèse a été confirmée. Les méthodes de classification linéaire se sont révélées légèrement supérieures, en particulier la régression logistique avec validation croisée et le SVM à noyau RBF.

Quant à notre seconde hypothèse (H2), qui concernait la détection des énoncés de thèse principale dans les TECHNETIUM à travers différents indices matériels, les résultats ont également validé notre postulat.

L'expérimentation avec 74 modèles a révélé l'importance de l'emplacement, des connecteurs de discours et des propriétés syntaxiques ou sémantiques des structures IMRAD dans l'identification des thèses principales. En particulier, l'emplacement s'est avéré être un indicateur puissant, surtout dans les résumés où les thèses principales tendent à se situer au début ou à la fin du texte. Les modèles ensemblistes ont également joué un rôle clé, en particulier pour la détection des thèses principales dans les discussions, où une combinaison d'approches a abouti aux meilleures performances. Cette nécessité de combiner diverses méthodologies pour une détection efficace des thèses principales, particulièrement dans des sections textuelles complexes comme les discussions, souligne la validité et la complexité de notre hypothèse H2. Une validation ultérieure de cette hypothèse avec un corpus plus diversifié serait cependant bénéfique. Voici un récapitulatif de nos conclusions pour les sous-hypothèses relatives à la détection des thèses principales au moyen d'indices matériels particuliers, soit l'emplacement (H2.1), les connecteurs de discours (H2.2) et les propriétés syntaxiques ou sémantiques (H2.3):

H2.1 (Emplacement des Thèses Principales) : Cette hypothèse a été largement confirmée. Nos expériences ont montré que l'emplacement est un indicateur crucial pour identifier les thèses principales. En particulier, pour les résumés, la méthode basée sur l'emplacement s'est avérée être la plus efficace, surpassant même les modèles ensemblistes plus complexes. Cette confirmation souligne l'importance de l'aspect spatial dans la structuration des textes académiques.

H2.2 (Connecteurs de Discours) : Les résultats sont plus nuancés pour cette hypothèse. Bien que les connecteurs de discours aient servi d'indicateurs utiles, les modèles basés uniquement sur ceux-ci n'étaient pas parmi les plus performants. Cela suggère que, bien que pertinents, les connecteurs de discours sont insuffisants seuls pour une détection fiable des thèses principales. Cette hypothèse est donc partiellement confirmée, avec une recommandation d'intégrer les connecteurs de discours avec d'autres méthodologies.

H2.3 (Propriétés Syntaxiques ou Sémantiques) : Cette hypothèse a été en partie confirmée. L'analyse des propriétés syntaxiques et sémantiques de certaines structures ou macrostructures IMRAD a joué un rôle dans la détection des thèses principales, mais cette influence variait selon la structure et la méthode de représentation utilisée. Voici un récapitulatif de nos conclusions pour les sous-hypothèses relatives à la détection des thèses principales au moyen de structures ou de macrostructures particulières, soit le titre (H2.3.1), le résumé (H2.3.2), l'introduction (H2.3.3) et la discussion (H2.3.4):

H2.3.1 (Titre) : Partiellement confirmée. Les thèses principales dans les discussions avaient une relation de similarité notable avec le titre, surtout lorsque la représentation DBOW était utilisée.

H2.3.2 (Résumé) : Partiellement confirmée. Les thèses principales dans les discussions avaient une relation de similarité avec le résumé, surtout avec la représentation « text-embedding-ada-002 ».

H2.3.3 (Introduction) : Confirmée. L'introduction s'est révélée être une macrostructure significative pour la tâche de détection des thèses principales.

H2.3.4 (Discussion) : Partiellement confirmée. La discussion a joué un rôle dans l'identification des thèses principales, mais pas de manière aussi décisive que l'introduction.

En résumé, nos expérimentations ont non seulement confirmé la viabilité de nos hypothèses, mais ont aussi mis en lumière la complexité et la nécessité d'approches nuancées et combinées pour une analyse d'argumentation efficace dans les TECHNETIUM. Ces résultats ouvrent la voie à des recherches futures plus approfondies dans le domaine de la fouille d'argument et de la lecture et analyse argumentaire de texte assistées par ordinateur. Ces avenues comprennent :

L'amélioration des techniques de modélisation : Développer de nouvelles méthodes pour améliorer la précision et le rappel des modèles existants, en explorant des techniques d'apprentissage profond avancées et de nouvelles architectures neuronales.

L'exploration de représentations vectorielles alternatives : Tester différentes formes de représentations vectorielles, notamment des modèles linguistiques plus complexes, pour évaluer leur efficacité dans la détection des thèses principales et l'appartenance aux structures IMRAD.

L'extension et la diversification des corpus : Utiliser des corpus textuels plus larges et variés, provenant de différents domaines académiques et styles de rédaction, pour tester la généralisabilité des modèles.

Le développement de techniques d'annotation : Améliorer les techniques d'annotation pour faciliter la création de corpus d'entraînement et de test.

L'analyse des faux positifs et des faux négatifs : Mener des études spécifiques pour comprendre et minimiser les erreurs des modèles, en utilisant notamment l'apprentissage actif et d'autres techniques d'analyse.

L'utilisation de modèles ensemblistes : Intégrer des approches ensemblistes pour améliorer les performances de détection, en combinant plusieurs modèles ou techniques.

L'étude de l'impact des connecteurs de discours : Approfondir l'étude sur l'efficacité des connecteurs de discours dans la détection des thèses principales.

L'analyse intégrée des composantes argumentatives : Analyser la provenance et l'influence des idées et des arguments, cartographier les réseaux intertextuels d'arguments, et développer des techniques de revue de littérature automatique.

L'intégration des approches de la LATAO : Combiner l'analyse de concepts, de thèmes et d'arguments pour une compréhension plus holistique des textes.

Ces recherches futures permettront d'améliorer la précision et l'efficacité des modèles de TALN, enrichissant ainsi notre compréhension des structures argumentatives, thématiques et conceptuelles dans les textes académiques.

ANNEXE A

Modèle révisé de la RST

Le modèle révisé de la RST que nous avons élaboré dans cette annexe correspond exactement à celui proposé par Stede et coll. (2017 notre trad.) à l'exception des relations d'attribution qui sont inspirées de Stede et coll. (2017) et de Potter (2019). En fait, les principes derrière les définitions des relations d'attribution sont donnés par Potter (2019), mais puisque l'un de ces principes est que l'attribution est souvent utilisée pour remplir la fonction d'une autre relation déjà définie par Stede et coll. (2017), alors la définition déjà définie est partiellement reprise pour définir ces cas spéciaux d'attribution. Subséquemment, à l'exception des relations d'attribution, toutes les définitions sont des traductions en français de celles tirées intégralement du guide d'annotation de Stede et coll. (2017) pour lesquelles nous avons retiré les exemples. Le format des définitions se divise en deux parties. D'abord, le rôle des deux unités jointes par la relation est expliqué. Ensuite, l'effet escompté par l'auteur en appliquant cette relation est explicité. De plus, lorsque des connecteurs rhétoriques peuvent faciliter l'identification d'une relation particulière, certains sont donnés en exemples. Enfin, les définitions peuvent inclure une remarque des théoriciens à propos de certaines particularités des relations. Spécifions que dans les définitions, « N » correspond au nucléus, « S » correspond au satellite, « A » correspond à l'auteur et « L » correspond au lecteur. Les définitions se présentent ainsi :

- N : Description du type et/ou de la fonction du nucléus (N) caractérisant principalement les objectifs de l'auteur (A), et non l'expression dans le texte. Ce champ n'est présent que lorsqu'une telle contrainte existe.
- S : Description du type et/ou de la fonction du satellite (S) caractérisant principalement les objectifs de l'auteur (A), et non l'expression dans le texte. Ce champ n'est présent que lorsqu'une telle contrainte existe.
- N/S : Description de la fonction de la combinaison nucléus/satellite (N/S). Lorsqu'il y a des contraintes ou des tendances sur l'ordre linéaire de N et S, celles-ci seront également données ici. Ce champ n'est présent que lorsqu'une telle contrainte existe.
- Effet : Description de l'effet recherché par l'auteur (A), notamment sur le lecteur (L) en utilisant la relation sous la forme d'un changement « avant-après ».
- Connecteurs typiques : Le cas échéant.
- Remarques : Facultatif.

Les relations sont regroupées en cinq espèces, c.-à-d. les relations pragmatiques, les relations sémantiques, les relations textuelles, les relations multinucléées et les relations d'attribution.

A. Les relations pragmatiques

Arrière-plan (*Background*)

- N/S : Comprendre S permet à L de comprendre plus facilement le contenu de N ; sans l'information d'arrière-plan dans S, il serait difficile de comprendre N. Dans un texte, S préfigure surtout, mais pas toujours N. Un arrière-plan S au début du texte sert souvent à introduire le sujet du texte.
- Effet : La capacité de L à comprendre N est augmentée.
- Connectifs typiques : Rarement signalés par des connecteurs.
- Remarque : Cette relation se présente rarement entre deux EA ; la plupart du temps, elle connectera des segments plus grands. De nombreux éditoriaux commencent par un arrière-plan S, qui sert de point de départ à l'investigation ultérieure de A.

Antithèse (*Antithesis*)

- N : A considère le contenu de N comme plus important ; c'est l'antithèse à laquelle A s'identifie.
- S : par rapport à N, A considère le contenu de S comme moins important. S est considéré comme la thèse à laquelle le A ne s'identifie pas.
- N/S : Les contenus de N et S ne sont pas compatibles - souvent au niveau de leur évaluation. En raison de l'incompatibilité, on ne peut pas avoir le même respect pour N et S. Dans un texte, S précède généralement N.
- Effet : La considération positive de L pour N est augmentée.
- Connecteurs typiques : mais ; négatif - plutôt ; négatif - à la place de ; ...
- Remarque : Des relations similaires sont Concession et Contraste ; voir aussi la section D. Un sous-groupe d'antithèse est constitué de cas où N sert à corriger une hypothèse énoncée dans S.

Concession (*Concession*)

- N : A considère le contenu de N comme plus important.
 - S : Par rapport à N, A considère le contenu de S comme moins important, mais il ne conteste pas que S est vrai.
 - N/S : A concède S et confirme implicitement que S et N ne sont généralement pas compatibles ; dans le cas présent, cependant, ils sont compatibles, et N est souligné.
 - Effet : La considération positive de L pour N est augmentée.
 - Connecteurs typiques : quoique ; mais; toujours; malgré; ...
-

Évidence (*Evidence*)

- N : Une déclaration/thèse/affirmation subjective, que L pourrait ne pas accepter ou ne pas considérer comme suffisamment importante ou positive.
 - S : Une déclaration que L est susceptible d'accepter ; habituellement une description « objective » d'un fait.
 - N/S : Comprendre S permet à L d'accepter plus facilement N ou de partager le point de vue particulier de A.
 - Effet : La croyance de L en N est augmentée.
 - Connecteurs typiques : Connecteurs causaux.
 - Remarque : La preuve relie souvent un segment S plus grand à un N plus court (la thèse).
-

Raison (*Raison*)

- N : Une déclaration/thèse/affirmation subjective, que L pourrait ne pas accepter ou ne pas considérer comme suffisamment importante ou positive.

- S : une déclaration/thèse/affirmation subjective.
 - N/S : comprendre S permet à L d'accepter plus facilement N ou de partager le point de vue particulier de A.
 - Effet : la croyance de R en N est augmentée.
 - Connecteurs typiques : connecteurs causaux.
 - Remarque : La raison est plus précise que la preuve. La différence réside dans le fait que S est présenté par A comme "objectif" (Preuve) ou constitue également une déclaration subjective elle-même (Raison).
-

Raison-N (*Raison-N*)

- N : une déclaration/thèse/affirmation subjective.
 - S : Une déclaration/thèse/affirmation subjective, que L pourrait ne pas accepter ou ne pas considérer comme suffisamment importante ou positive.
 - N/S : comprendre N permet à L d'accepter plus facilement S ou de partager le point de vue particulier de A.
 - Effet : la croyance de L en S est augmentée.
 - Connecteurs typiques : connecteurs causaux.
 - Remarque : Cette relation est parallèle à Reason. Elle s'applique lorsque pour la fonction texte, la raison est plus importante que la cause. Dans le genre des éditoriaux, cela sera relativement rare.
-

Justifier (*Justify*)

- N : Une déclaration/thèse/affirmation subjective, que L pourrait ne pas accepter ou ne pas considérer comme suffisamment importante ou positive.
- S : énoncé d'une attitude fondamentale (par exemple, politique, morale) de la personne qui agit.

- N/S : comprendre S permet à L d'accepter plus facilement N ou de partager le point de vue particulier de A.
 - Effet : la volonté de L d'accepter le droit de A de présenter N est augmentée.
 - Connecteurs typiques : connecteurs causaux.
-

Évaluation-S (*Evaluation-S*)

- N : Description d'un état des choses ou énoncé subjectif (mais pas du point de vue de A).
 - S : une évaluation subjective (positive/négative, souhaitable/indésirable) du point de vue de A.
 - N/S : S évalue N.
 - Effet : L reconnaît que S évalue N et reconnaît la valeur qu'il attribue.
 - Connectifs typiques : Rarement signalés par des connecteurs.
 - Remarque : Habituellement, le segment « évaluation » suit le segment « évalué ». Mais parfois l'ordre est le contraire.
-

Évaluation-N (*Evaluation-N*)

- N : une évaluation subjective (positive/négative, souhaitable/indésirable) du point de vue de A.
 - S : Description d'un état des choses ou énoncé subjectif (mais pas du point de vue de A).
 - N/S : N évalue S.
 - Effet : L reconnaît que N évalue S et reconnaît la valeur qu'il attribue.
 - Connectifs typiques : Rarement signalés par des connecteurs.
 - Remarque : Cette relation est parallèle à Evaluation-S. Décider entre les deux dépend uniquement de juger de l'importance relative des segments pour le texte.
-

Motivation (*Motivation*)

- N : une action à effectuer par L.
 - N/S : S présente une raison d'effectuer l'action décrite dans N.
 - Effet : le désir de L d'effectuer une action en N est augmenté.
 - Connecteurs typiques : connecteurs causaux.
 - Remarque : Dans le genre éditorial, cette relation est rare. Parfois, A encourage métaphoriquement tous les L à faire une certaine action, comme dans l'exemple donné ici.
-

Habilitation (*Enablement*)

- N : une action à effectuer par L.
 - N/S : comprendre S permet à L d'effectuer plus facilement l'action décrite dans N.
 - Effet : la capacité potentielle de L à effectuer l'action dans N augmente.
 - Connecteurs typiques :
 - Remarque : Dans le genre éditorial, cette relation est rare.
-

B. Les relations sémantiques

Circonstance (*Circumstance*)

- S : un événement ou un état qui se produisent réellement ou se sont produits (c'est-à-dire, pas un état hypothétique).
- N/S : S caractérise un cadre dans lequel N doit être interprété, comme sa position temporelle ou locative.
- Effet : L reconnaît que S fournit le cadre d'interprétation de N.
- Connecteurs typiques : as ; quand; alors que; ... pour un cadre temporel.

Condition (*Condition*)

- S : Une situation hypothétique, future ou autrement irréaliste.
- N/S : La réalisation de N dépend de la réalisation de S.
- Effet : L reconnaît comment la réalisation de N dépend de la réalisation de S.
- Connecteurs typiques : si.. alors ; au cas où; ...

Sinon (*Otherwise*)

- N : Une situation hypothétique, future ou autrement irréaliste.
- S : Une situation hypothétique, future ou autrement irréaliste.
- N/S : la réalisation de N empêche la réalisation de S.
- Effet : L reconnaît la relation de dépendance de prévention entre la réalisation de N et la réalisation de S.
- Connecteurs typiques : sinon ; ...

Sauf si (*Unless*)

- N : Une situation hypothétique, future ou autrement non réalisée.
 - S : Une situation hypothétique, future ou autrement non réalisée.
 - N/S : S détermine la réalisation de N : N n'est réalisé que si S n'est pas réalisé.
 - Effet : L reconnaît que N est réalisé à condition que S ne soit pas réalisé.
 - Connecteurs typiques : sauf si ; ...
-

Élaboration (*Elaboration*)

- N/S : S fournit des détails ou plus d'informations sur l'état des choses décrit dans N (mais pas sur une seule entité mentionnée dans N ; voir E-ÉLABORATION ci-dessous). N précède S dans le texte. Les relations typiques entre N et S sont *set::element*, *whole::part*, *abstraction:instance*, *procedure::step*.
 - Effet : L reconnaît que S fournit des détails supplémentaires sur l'état des choses dans N. L identifie l'élément du sujet pour lequel des détails sont fournis.
 - Connecteurs typiques : en particulier ; Par exemple ; ...
-

E-Élaboration (*E-Elaboration*)

- N/S : S fournit des détails ou plus d'informations sur une seule entité mentionnée dans N. N précède S dans le texte.
 - Effet : L reconnaît S comme fournissant des détails supplémentaires pour une entité dans N. L identifie l'élément de l'objet pour lequel des détails sont fournis.
 - Connectifs typiques : Rarement signalés par des connecteurs.
 - Remarque : Dans l'exemple, la clause *as (S)* ne renseigne pas sur l'activité 'sens', mais sur la 'classe', c'est-à-dire une entité mentionnée dans N.
-

Interprétation (*Interpretation*)

- N/S : S déplace le contenu de N vers un cadre conceptuel différent. Cela n'implique pas une évaluation de N (ou l'évaluation n'a qu'une importance secondaire). N précède S dans le texte.
- Effet : L reconnaît que S relie N à un cadre d'idées non impliquées dans la connaissance présentée dans N lui-même.
- Connecteurs typiques : ainsi ;

- Remarque : Chaque fois que S fournit principalement une évaluation (sur l'échelle positive/négative) de N, l'évaluation doit être utilisée plutôt que l'interprétation.
-

Moyens (*Means*)

- N : Une activité ou une action.
 - N/S : S fournit des informations qui rendent la réalisation/exécution de N plus probable ou simple (par exemple, un instrument).
 - Effet : L reconnaît que la méthode ou l'instrument dans S tend à rendre la réalisation de N plus probable.
 - Connecteurs typiques : ainsi ; ...
-

Cause (*Cause*)

- N : Un état ou un événement dans le monde.
 - S : Un état ou un événement dans le monde.
 - N/S : l'état/l'événement dans N est causé par l'état/l'événement dans S.
 - Effet : L reconnaît S comme cause de N.
 - Connecteurs typiques : parce que ; depuis; donc; ...
-

Résultat (*Result*)

- N : Un état ou un événement dans le monde.
- S : Un état ou un événement dans le monde.
- N/S : l'état/l'événement dans S est causé par l'état/l'événement dans N.
- Effet : L reconnaît N comme cause de S.

- Connecteurs typiques : parce que ; depuis; donc; ...
 - Remarque : Cette relation est parallèle à Cause. Décider entre les deux dépend uniquement de juger de l'importance relative des segments pour le texte.
-

But (*Purpose*)

- N : Une activité ou une action.
 - S : Une situation hypothétique, future ou autrement non réalisée.
 - N/S : S est réalisé par la réalisation/l'exécution de N.
 - Effet : L reconnaît que l'activité dans N est initiée afin de réaliser S.
 - Connecteurs typiques : afin de ; pour; ...
 - Remarque : Il existe une relation causale au sens large. La différence avec les relations Cause/Résultat est qu'avec But, S est clairement marqué comme hypothétique/non réalisé, et représente l'intention ou le but de la personne agissante.
-

Solution (*Solutionhood*)

- S : Le contenu de S peut être considéré comme un problème.
 - N/S : Le contenu de N peut être considéré comme une solution au problème N. N précède généralement S dans le texte.
 - Effet : L reconnaît N comme une solution au problème présenté dans S.
 - Connectifs typiques : Rarement signalés par des connecteurs.
-

C. Les relations textuelles

Préparation (*Preparation*)

- N/S : S précède N dans le texte. S oriente L vers le sujet de N.
 - Effet : L est plus prêt, intéressé ou orienté pour lire N.
 - Connectifs typiques : Rarement signalés par des connecteurs ; parfois par un colon.
 - Remarque : Cette relation est à utiliser lorsque S ne sert pas à un objectif plus fort que celui de définir le sujet pour N, ou lorsqu'il consiste en une formule d'introduction. S ne doit contenir qu'un minimum d'informations en soi.
-

Reformulation (*Restatement*)

- N/S : N précède S dans le texte. S reprend les informations données en N, en utilisant une formulation différente. N et S sont de taille à peu près égale.
 - Effet : R reconnaît S comme une reformulation de N.
 - Connectifs typiques : en d'autres termes ; ...
-

Résumé (*Summary*)

- N : N se compose de plusieurs EDU.
 - N/S : S succède à N dans le texte et répète l'information donnée dans N, mais sous une forme plus courte.
 - Effet : L reconnaît S comme une reformulation plus courte de N.
 - Connecteurs typiques : en bref ; ...
-

D. Les relations multinucléées

Contraste (*Contrast*)

- N : Exactement deux noyaux. Les deux sont d'égale importance pour les objectifs de A. Le contenu est comparable, mais pas identique. Ils diffèrent par des aspects importants pour A.
 - Effet : L reconnaît la comparabilité et la ou les différences résultant de la comparaison effectuée.
 - Connecteurs typiques : d'autre part ; encore; mais; ...
-

Séquence (*Sequence*)

- N : Les noyaux décrivent des états de choses qui se produisent dans un ordre temporel particulier.
 - Effet : L reconnaît les relations de succession entre les noyaux.
 - Connecteurs typiques : puis ; avant; après; ...
-

Liste (*List*)

- N : Les noyaux fournissent des informations qui peuvent être reconnues comme liées, énumérées. Ils contribuent tous à la fonction texte de la même manière.
 - Effet : L reconnaît la comparabilité des éléments liés.
 - Connecteurs typiques : virgule, énumération, ...
 - Remarque : Voir Section 4.6 pour des conseils sur la gestion des contextes elliptiques (comme dans l'exemple).
-

Conjonction (*Conjonction*)

- N : Les noyaux fournissent des informations qui peuvent être reconnues comme liées, énumérées. Ils contribuent tous à la fonction du texte de la même manière et sont liés par des conjonctions de coordination.
 - Effet : L reconnaît que les éléments liés sont conjoints.
 - Connecteurs typiques : et ; ou ; ...
 - Remarque : Les fonctions de Liste et de Conjonction sont identiques. Lorsque la condition de surface pour la conjonction est vérifiée, cette relation doit être utilisée.
-

Articulation (*Joint*)

- N : Les noyaux fournissent différents types d'informations, qui ne sont pas du même type ; pourtant ils ne sont pas dans une relation sémantique ou pragmatique clairement identifiable, ni ne forment une énumération. Pourtant, il existe un lien cohérent, car ils contribuent de manière similaire à la fonction globale du texte.
 - Effet : L reconnaît que chaque noyau apporte son propre message, qui sert cependant le même objectif général du texte.
 - Connecteurs typiques : connecteurs additifs tels qu'en plus ; aussi.
 - Remarque : Joint doit être utilisé lorsqu'une relation multinucléaire est nécessaire (du point de vue global du texte), mais qu'aucune des relations spécifiques n'est applicable.
-

E. Les relations d'attribution

A-Justifier (*Attribution as Justify*)

- N : Une déclaration/thèse/affirmation subjective, que L pourrait ne pas accepter ou ne pas considérer comme suffisamment importante ou positive.
- S : Énoncé d'une attitude fondamentale (par exemple, politique, morale) de la personne qui agit.
- N/S : Comprendre S permet à L d'accepter plus facilement N ou de partager le point de vue particulier de A.
- Effet : La volonté de L d'accepter le droit de A de présenter N est augmentée.
- Connecteurs typiques : Connecteurs causaux.
- Remarque : La fonction de ces attributions n'est pas principalement de donner du crédit aux sources, mais de fournir une autorité pour des affirmations pour lesquelles A manque d'expertise. C'est-à-dire que A s'appuie sur une autorité empruntée.

A-Évaluation (*Attribution as Evaluation*)

- N : Description d'un état des choses ou énoncé subjectif (mais pas du point de vue de A).
- S : Une évaluation subjective (positive/négative, souhaitable/indésirable) du point de vue de A.
- N/S : S évalue N.
- Effet : L reconnaît que S évalue N et reconnaît la valeur qu'il attribue.
- Connectifs typiques : Rarement signalés par des connecteurs.
- Remarque :

A-Interprétation (*Attribution as Interpretation*)

- N/S : S déplace le contenu de N vers un cadre conceptuel différent. Cela n'implique pas une évaluation de N (ou l'évaluation n'a qu'une importance secondaire). N précède S dans le texte.

- Effet : L reconnaît que S relie N à un cadre d'idées non impliquées dans la connaissance présentée dans N lui-même.
 - Connecteurs typiques : ainsi ;
 - Remarque :
-

A-Cause (*Attribution as Cause*)

- N : Un état ou un événement dans le monde.
 - S : Un état ou un événement dans le monde.
 - N/S : l'état/l'événement dans N est causé par l'état/l'événement dans S.
 - Effet : L reconnaît S comme cause de N.
 - Connecteurs typiques : parce que ; depuis; donc; ...
 - Remarque : Similaire à A-Évaluation, sauf que la saillance rhétorique est sur le prédicat d'attribution.
-

A-Élaboration (*Attribution as Elaboration*)

- N/S : S fournit des détails ou plus d'informations sur une seule entité mentionnée dans N. N précède S dans le texte.
 - Effet : L reconnaît S comme fournissant des détails supplémentaires pour une entité dans N. L identifie l'élément de l'objet pour lequel des détails sont fournis.
 - Connectifs typiques : Rarement signalés par des connecteurs.
 - Remarque : Les attributions en tant qu'élaborations incluent des états cognitifs, tels que penser et croire.
-

A-Évidence (*Attribution as Evidence*)

- N : Une déclaration/thèse/affirmation subjective, que L pourrait ne pas accepter ou ne pas considérer comme suffisamment importante ou positive.
 - S : Une déclaration que L est susceptible d'accepter ; habituellement une description « objective » d'un fait.
 - N/S : Comprendre S permet à L d'accepter plus facilement N ou de partager le point de vue particulier de A.
 - Effet : La croyance de L en N est augmentée.
 - Connecteurs typiques : Connecteurs causaux.
 - Remarque :
-

Attribution (*Attribution as Attribution*)

- N : Quelque chose qui est dit.
 - S : Quelqu'un ou groupe à qui attribuer ce qui est dit.
 - N/S : Permet à L d'attribuer N à S.
 - Effet : S apparaît être A de N
 - Remarque : Les attributions ne sont parfois que des attributions. Le cas échéant leur fonction en est une d'articulation textuelle.
-

A-FConcession (*Cognitive States as (Faux) Concession*)

- N : A considère le contenu de N comme plus important, mais A croit que L est plus susceptible d'accepter N sous forme de croyance plutôt que sous forme d'affirmation.
- S : S est un état cognitif attribuable à A portant sur N donnant à N l'apparence d'une croyance ou d'une opinion.
- N/S : A apparaît concéder N.
- Effet : La considération positive de L pour N est augmentée.

- Remarque : Les expressions d'états cognitifs sont parfois utilisées pour modérer une affirmation avec une indication d'incertitude. Cela implique une délimitation apparente de la revendication comme une simple question d'opinion. Mais l'effet recherché n'est pas de concéder l'incertitude sur la réclamation, mais de sembler concéder l'incertitude afin de modérer la livraison de la réclamation, avec l'effet recherché d'assurer l'acceptation.
-

ANNEXE B

PDTB3 : Connecteur explicite et leur sens

Tiré de Prasad *et al.* (2019), le tableau suivant montre les types de connecteur explicite (*explicit connective*) présentés en ordre alphabétique. La deuxième colonne du tableau – *count* – donne le nombre d'occurrence du type de connecteur dans le corpus PDTB3. La troisième colonne – *sense* – donne le sens du type au moyen de trois niveaux d'interprétation qui sont séparés par un point dans le nom du type. Le premier niveau contient quatre (4) grands types de haut niveau soit « *contingency* », « *expansion* », « *temporal* », et « *comparaison* ». La contingence (*contingency*) est une catégorie regroupant les connecteurs qui expriment une relation de cause à effet ou de raison. Ils indiquent que l'occurrence d'un événement ou d'une situation dépend d'un autre. Les connecteurs typiques de cette catégorie incluent des mots comme « *because* » (parce que), « *therefore* » (donc), « *thus* » (ainsi), indiquant une relation de causalité ou d'implication logique entre les idées. L'expansion (*expansion*) est une catégorie regroupant les connecteurs utilisés pour ajouter des informations, donner des exemples, clarifier, ou étendre sur un point déjà mentionné. Ils peuvent aussi servir à mettre en place des relations d'addition, d'alternative, ou de conjonction. Des exemples typiques comprennent « *and* » (et), « *additionally* » (de plus), « *for instance* » (par exemple). Ils sont essentiels pour enrichir ou compléter l'information présentée. Le temporel (*temporal*) est une catégorie incluant les connecteurs qui établissent une relation temporelle entre les événements ou les situations. Ils peuvent indiquer quand quelque chose se produit, sa durée, ou sa séquence par rapport à d'autres événements. Des exemples courants sont « *before* » (avant), « *after* » (après), « *meanwhile* » (pendant ce temps). Ces connecteurs sont importants pour comprendre l'ordre et la temporalité des événements dans le texte. La comparaison (*comparison*) est une catégorie regroupant les connecteurs qui établissent une relation de comparaison ou de contraste entre deux éléments ou idées. Ils peuvent souligner des similitudes, des différences, ou des oppositions. Des exemples typiques de cette catégorie sont « *like* » (comme), « *similarly* » (de même), « *on the contrary* » (au contraire). Ils sont utilisés pour évaluer ou mettre en relation des éléments en les comparant. Enfin, on remarque qu'un même connecteur peut se retrouver dans plusieurs types distincts et, le cas échéant, le contexte est nécessaire pour identifier le type. Notons que les connecteurs discontinus sont indiqués par un symbole « + » entre leurs parties.

Explicit_connective	Count	Sense
about	2	Contingency.Cause.Reason
accordingly	5	Contingency.Cause.Result
additionally	7	Expansion.Conjunction
after	56	Temporal.Asynchronous.Succession
after	56	Contingency.Cause.Reason
after	533	Temporal.Asynchronous.Succession
afterward	6	Temporal.Asynchronous.Precedence
afterwards	5	Temporal.Asynchronous.Precedence
albeit	1	Comparison.Concession.Arg2-as-denier
along with	2	Expansion.Conjunction
also	1	Temporal.Synchronous
also	1736	Expansion.Conjunction
alternatively	4	Expansion.Disjunction
alternatively	2	Expansion.Substitution.Arg2-as-subst
although	206	Comparison.Concession.Arg1-as-denier
although	105	Comparison.Concession.Arg2-as-denier
although	14	Comparison.Contrast
although	2	Expansion.Exception.Arg2-as-excpt
although	1	Temporal.Synchronous Comparison.Contrast
and	2	Comparison.Concession+SpeechAct.Arg2-as-denier+SpeechAct
and	16	Comparison.Contrast
and	1	Contingency.Cause+SpeechAct.Result+SpeechAct
and	4	Contingency.Cause.Reason
and	12	Contingency.Cause.Result Expansion.Conjunction
and	5	Contingency.Cause.Result
and	22	Contingency.Condition.Arg1-as-cond
and	128	Contingency.Purpose.Arg2-as-goal
and	6189	Expansion.Conjunction
and	4	Expansion.Level-of-detail.Arg2-as-detail
and	19	Expansion.Manner.Arg2-as-inanner
and then	1	Expansion.Disjunction
and/or	2	Expansion.Conjunction Expansion.Disjunction
as	7	Comparison.Concession.Arg1-as-denier
as	5	Comparison.Contrast
as	65	Comparison.Similarity
as	3	Contingency.Cause+Belief.Reason+Belief
as	180	Contingency.Cause.Reason
as	4	Expansion.Instantiation.Arg2-as-instance
as	8	Expansion.Level-of-detail.Arg1-as-detail
as	3	Expansion.Manner.Arg2-as-manner

as	23	Temporal.Asynchronous.Succession Contingency.Cause.Reason
as	3	Temporal.Asynchronous.Succession
as	2	Temporal.Synchronous Comparison.Contrast
as	1	Temporal.Synchronous Comparison.Similarity
as	5	Temporal.Synchronous Contingency.Cause+Belief.Reason+Belief
as	146	Temporal.Synchronous Contingency.Cause.Reason
as	383	Temporal.Synchronous
as a result	78	Contingency.Cause.Result
as an alternative	2	Expansion.Disjunction
as if	4	Comparison.Concession.Arg2-as-denier
as if	10	Comparison.Similarity Expansion.Manner.Arg2-as-manner
as if	1	Expansion.Instantiation.Arg1-as-instance
as if	1	Expansion.Manner.Arg2-as-manner
as long as	13	Contingency.Condition.Arg2-as-cond
as long as	7	Temporal.Synchronous Contingency.Condition.Arg2-as-cond
as long as	4	Temporal.Synchronous
as much as	2	Comparison.Concession.Arg1-as-denier
as much as	1	Expansion.Conjunction
as much as	1	Expansion.Substitution.Arg2-as-subst
as soon as	11	Temporal.Asynchronous.Succession
as soon as	9	Temporal.Synchronous
as though	2	Comparison.Similarity Expansion.Manner.Arg2-as-manner
as though	1	Comparison.Similarity
as though	2	Expansion.Level-of-detail.Arg2-as-detail
as well	6	Comparison.Similarity
as well	12	Expansion.Conjunction
as well as	7	Expansion.Conjunction
at the same time	65	Temporal.Synchronous
because	2	Contingency.Cause+Belief.Reason+Belief
because	833	Contingency.Cause.Reason
because	2	Contingency.Condition+SpeechAct
because of	12	Contingency.Cause.Reason
before	309	Temporal.Asynchronous.Precedence
before	2	Temporal.Asynchronous.Succession
before and after	1	Temporal.Asynchronous.Precedence Temporal.Asynchronous.Succession
besides	19	Expansion.Conjunction
beyond	1	Expansion.Conjunction
both+and	6	Expansion.Conjunction
but	2	Comparison.Concession+SpeechAct.Arg2-as-denier+SpeechAct
but	3063	Comparison.Concession.Arg2-as-denier
but	618	Comparison.Contrast

but	1	Contingency.Cause+SpeechAct.Reason+SpeechAct
but	1	Contingency.Cause.Reason Comparison.Concession.Arg2-as-denier
but	42	Expansion.Conjunction
but	3	Expansion.Exception.Arg2-as-excpt
but	1	Temporal.Synchronous Comparison.Contrast
but also	1	Expansion.Conjunction
but then	3	Comparison.Concession.Arg2-as-denier
but then again	1	Comparison.Concession.Arg2-as-denier
by	1	Comparison.Concession.Arg1-as-denier
by	3	Contingency.Cause+Belief.Reason+Belief Expansion.Manner.Arg2-as-manner
by	95	Contingency.Cause.Reason Expansion.Manner.Arg2-as-manner
by	10	Contingency.Cause.Reason
by	32	Contingency.Condition.Arg2-as-cond Expansion.Manner.Arg2-as-manner
by	2	Contingency.Condition.Arg2-as-cond
by	115	Contingency.Purpose.Arg1-as-goal Expansion.Manner.Arg2-as-manner
by	2	Expansion.Level-of-detail.Arg2-as-detail
by	174	Expansion.Manner.Arg2-as-manner
by	1	Temporal.Synchronous
by comparison	11	Comparison.Contrast
by contrast	28	Comparison.Contrast
by then	1	Temporal.Asynchronous.Succession Contingency.Cause.Reason
by then	6	Temporal.Asynchronous.Succession
consequently	10	Contingency.Cause.Result
conversely	2	Comparison.Contrast
depending on	3	Contingency.Condition.Arg2-as-cond
depending upon	1	Contingency.Condition.Arg2-as-cond
despite	9	Comparison.Concession.Arg1-as-denier
due to	1	Contingency.Cause.Reason
earlier	15	Temporal.Asynchronous.Succession
either+or	2	Contingency.Negative-condition.Arg1-as-negCond
either+or	36	Expansion.Disjunction
else	1	Contingency.Negative-condition.Arg1-as-negCond
even after	8	Temporal.Asynchronous.Succession Comparison.Concession.Arg1-as-denier
even as	2	Comparison.Concession.Arg1-as-denier
even as	11	Temporal.Synchronous Comparison.Concession.Arg1-as-denier
even before	14	Temporal.Asynchronous.Precedence Comparison.Concession.Arg1-as-denier
even before then	1	Temporal.Asynchronous.Succession Comparison.Concession.Arg2-as-denier
even if	87	Comparison.Concession.Arg1-as-denier
even so	9	Comparison.Concession.Arg2-as-denier
even then	2	Temporal.Asynchronous.Precedence Comparison.Concession.Arg2-as-denier
even though	69	Comparison.Concession.Arg1-as-denier

even though	26	Comparison.Concession.Arg2-as-denier
even when	8	Comparison.Concession.Arg1-as-denier
even when	4	Temporal.Asynchronous.Succession Comparison.Concession.Arg1-as-denier
even when	2	Temporal.Synchronous Comparison.Concession.Arg1-as-denier
even while	3	Temporal.Synchronous Comparison.Concession.Arg1-as-denier
even with	2	Comparison.Concession.Arg1-as-denier
except	12	Expansion.Exception.Arg2-as-ecpt
finally	18	Expansion.Conjunction
finally	1	Temporal.Asynchronous.Precedence Contingency.Cause.Result
finally	13	Temporal.Asynchronous.Precedence
for	1	Comparison.Concession.Arg1-as-denier
for	34	Contingency.Cause.Reason
for	1	Contingency.Cause.Result
for	7	Contingency.Condition.Arg2-as-cond
for	16	Contingency.Purpose.Arg2-as-goal
for	1	Expansion.Level-of-detail.Arg2-as-detail
for example	200	Expansion.Instantiation.Arg2-as-instance
for instance	98	Expansion.Instantiation.Arg2-as-instance
from	2	Contingency.Cause+Belief.Reason+Belief
from	1	Contingency.Cause.Reason Contingency.Condition.Arg2-as-cond
from	1	Contingency.Cause.Reason Expansion.Manner.Arg2-as-manner
from	2	Contingency.Cause.Reason
from	1	Expansion.Substitution.Arg1-as-subst
further	7	Expansion.Conjunction
furthermore	12	Expansion.Conjunction
given	3	Contingency.Cause+Belief.Reason+Belief
given	6	Contingency.Cause.Reason
hence	5	Contingency.Cause.Result
however	5	Comparison.Concession.Arg1-as-denier
however	390	Comparison.Concession.Arg2-as-denier
however	95	Comparison.Contrast
however	1	Temporal.Synchronous Comparison.Contrast
if	1	Comparison.Concession+SpeechAct.Arg2-as-denier+SpeechAct
if	6	Comparison.Concession.Arg1-as-denier
if	3	Comparison.Concession.Arg2-as-denier
if	2	Comparison.Contrast
if	56	Contingency.Condition+SpeechAct
if	1	Contingency.Condition.Arg2-as-cond Expansion.Level-of-detail.Arg2-as-detail
if	1084	Contingency.Condition.Arg2-as-cond
if	1	Temporal.Synchronous Contingency.Condition.Arg2-as-cond
if and when	2	Contingency.Condition.Arg2-as-cond

if and when	1	Temporal.Synchronous Contingency.Condition.Arg2-as-cond
if only	1	Comparison.Concession.Arg2-as-denier
if only	4	Contingency.Condition.Arg2-as-cond
if only	1	Contingency.Purpose.Arg2-as-goal
if + then	1	Contingency.Condition+SpeechAct
if + then	37	Contingency.Condition.Arg2-as-cond
in	1	Contingency.Cause+Belief.Reason+Belief
in	3	Contingency.Cause.Reason Expansion.Manner.Arg2-as-manner
in	1	Contingency.Cause.Reason
in	2	Contingency.Condition.Arg2-as-cond Expansion.Manner.Arg2-as-manner
in	8	Contingency.Condition.Arg2-as-cond
in	2	Contingency.Purpose.Arg2-as-goal
in	1	Expansion.Instantiation.Arg1-as-instance
in	17	Expansion.Level-of-detail.Arg1-as-detail
in	2	Expansion.Level-of-detail.Arg2-as-detail
in	13	Expansion.Manner.Arg2-as-manner
in	1	Temporal.Synchronous Contingency.Purpose.Arg2-as-goal
in	1	Temporal.Synchronous Expansion.Level-of-detail.Arg1-as-detail
in	18	Temporal.Synchronous
in addition	165	Expansion.Conjunction
in any case	3	Comparison.Concession.Arg2-as-denier
in case	6	Contingency.Condition.Arg2-as-cond
in contrast	12	Comparison.Contrast
in fact	4	Comparison.Concession.Arg2-as-denier
in fact	7	Comparison.Contrast
in fact	36	Expansion.Conjunction
in fact	3	Expansion.Instantiation.Arg2-as-instance
in fact	1	Expansion.Level-of-detail.Arg1-as-detail
in fact	34	Expansion.Level-of-detail.Arg2-as-detail
in order	4	Contingency.Condition.Arg2-as-cond
in order	51	Contingency.Purpose.Arg2-as-goal
in other words	17	Expansion.Equivalence
in particular	6	Expansion.Instantiation.Arg2-as-instance
in particular	9	Expansion.Level-of-detail.Arg2-as-detail
in short	4	Expansion.Level-of-detail.Arg1-as-detail
in sum	2	Expansion.Level-of-detail.Arg1-as-detail
in that	1	Expansion.Level-of-detail.Arg2-as-detail
in the end	1	Comparison.Concession.Arg2-as-denier
in the end	1	Comparison.Contrast
in the end	2	Contingency.Cause.Result
in the end	1	Expansion.Conjunction

in the end	2	Expansion.Level-of-detail.Arg1-as-detail
in the end	1	Expansion.Level-of-detail.Arg2-as-detail
in the end	3	Temporal.Asynchronous.Precedence
in the meantime	2	Temporal.Asynchronous.Succession
in the meantime	1	Temporal.Synchronous Comparison.Contrast
in the meantime	11	Temporal.Synchronous
in the meanwhile	1	Temporal.Synchronous
indeed	1	Comparison.Concession.Arg2-as-denier
indeed	4	Contingency.Caused-Belief.Reason+Belief
indeed	1	Contingency.Cause.Reason
indeed	1	Contingency.Cause.Result Expansion.Conjunction
indeed	54	Expansion.Conjunction
indeed	2	Expansion.Equivalence
indeed	2	Expansion.Instantiation.Arg2-as-instance
indeed	1	Expansion.Level-of-detail.Arg1-as-detail
indeed	37	Expansion.Level-of-detail.Arg2-as-detail
insofar as	1	Contingency.Cause.Reason
insofar as	1	Expansion.Level-of-detail.Arg2-as-detail
instead	112	Expansion.Substitution.Arg2-as-subst
instead of	43	Expansion.Substitution.Arg1-as-subst
later	92	Temporal.Asynchronous.Precedence
later on	2	Temporal.Asynchronous.Precedence
lest	2	Contingency.Negative-condition.Arg1-as-negCond
like	1	Comparison.Contrast
like	3	Comparison.Similarity
like	1	Expansion.Instantiation.Arg2-as-instance
likewise	8	Expansion.Conjunction
meantime	2	Temporal.Synchronous
meanwhile	2	Comparison.Concession.Arg2-as-denier
meanwhile	7	Comparison.Contrast
meanwhile	27	Expansion.Conjunction
meanwhile	1	Temporal.Synchronous Comparison.Concession.Arg2-as-denier
meanwhile	33	Temporal.Synchronous Comparison.Contrast
meanwhile	2	Temporal.Synchronous Comparison.Similarity
meanwhile	120	Temporal.Synchronous
more accurately	1	Expansion.Substitution.Arg2-as-subst
moreover	103	Expansion.Conjunction
much less	3	Expansion.Conjunction
neither+nor	1	Comparison.Contrast
neither+nor	2	Expansion.Conjunction
nevertheless	32	Comparison.Concession.Arg2-as-denier

nevertheless	12	Comparison.Contrast
next	1	Temporal.Asynchronous.Precedence Expansion.Conjunction
next	4	Temporal.Asynchronous.Precedence
no matter	8	Comparison.Concession.Arg1-as-denier
nonetheless	25	Comparison.Concession.Arg2-as-denier
nonetheless	2	Comparison.Contrast
nor	1	Comparison.Concession.Arg2-as-denier
nor	31	Expansion.Conjunction
nor	1	Expansion.Disjunction
not just+but	1	Expansion.Conjunction
not just+but+also	1	Expansion.Conjunction
not only	5	Expansion.Conjunction
not only+also	1	Expansion.Conjunction
not only because of	1	Contingency.Cause.Reason
not only-f but	1	Comparison.Concession.Arg2-as-denier
not only-f but	18	Expansion.Conjunction
not only-f but also	1	Comparison.Contrast
not only-f but also	9	Expansion.Conjunction
not so much as	1	Expansion.Substitution.Arg2-as-subst
now that	10	Contingency.Cause.Reason
now that	1	Temporal.Asynchronous.Precedence Contingency.Cause.Reason
now that	4	Temporal.Asynchronous.Succession Contingency.Cause.Reason
now that	4	Temporal.Synchronous Contingency.Cause.Reason
now that	2	Temporal.Synchronous
on	1	Contingency.Cause.Reason
on the contrary	4	Comparison.Contrast
on the one hand-f on the other	1	Comparison.Contrast
on the one hand-fon the other hand	1	Comparison.Concession.Arg2-as-denier
on the one hand-fon the other hand	1	Comparison.Contrast
on the other hand	4	Comparison.Concession.Arg2-as-denier
on the other hand	32	Comparison.Contrast
once	4	Contingency.Condition.Arg2-as-cond
once	7	Temporal.Asynchronous.Succession Contingency.Cause.Reason
once	3	Temporal.Asynchronous.Succession Contingency.Condition.Arg2-as-cond
once	70	Temporal.Asynchronous.Succession
only	2	Comparison.Concession.Arg2-as-denier
only	1	Comparison.Contrast
only	3	Expansion.Exception.Arg2-as-excpt
only	1	Expansion.Level-of-detail.Arg2-as-detail
only if	13	Contingency.Condition.Arg2-as-cond
or	11	Comparison.Concession+SpeechAct.Arg2-as-denier+SpeechAct

or	1	Comparison.Concession.Arg2-as-denier
or	2	Contingency.Condition+SpeechAct
or	7	Contingency.Negative-condition.Arg1-as-negCond
or	71	Expansion.Conjunction
or	258	Expansion.Disjunction
or	6	Expansion.Equivalence
or otherwise	2	Expansion.Disjunction
otherwise	4	Contingency.Negative-condition.Arg1-as-negCond
otherwise	15	Expansion.Exception.Arg1-as-excpt
plus	1	Expansion.Conjunction
previously	1	Temporal.Asynchronous.Succession Comparison.Contrast
previously	53	Temporal.Asynchronous.Succession
rather	17	Expansion.Substitution.Arg2-as-subst
rather than	40	Expansion.Substitution.Arg1-as-subst
regardless	2	Comparison.Concession.Arg2-as-denier
regardless of	6	Comparison.Concession.Arg1-as-denier
separately	72	Expansion.Conjunction
separately	2	Temporal.Synchronous Expansion.Conjunction
similarly	18	Comparison.Similarity
simultaneously	6	Temporal.Synchronous
since	96	Contingency.Cause.Reason
since	10	Temporal.Asynchronous.Precedence
since	11	Temporal.Asynchronous.Succession Contingency.Cause.Reason
since	83	Temporal.Asynchronous.Succession
since before	1	Temporal.Asynchronous.Succession
so	1	Contingency.Caused-Belief.Result+Belief
so	222	Contingency.Cause.Result
so	44	Contingency.Purpose.Arg2-as-goal
so as	3	Contingency.Purpose.Arg2-as-goal
so long as	4	Contingency.Condition.Arg2-as-cond
so much as	1	Expansion.Substitution.Arg2-as-subst
so that	10	Contingency.Cause.Result
so that	21	Contingency.Purpose.Arg2-as-goal
specifically	10	Expansion.Level-of-detail.Arg2-as-detail
still	115	Comparison.Concession.Arg2-as-denier
still	75	Comparison.Contrast
still	2	Temporal.Asynchronous.Precedence
still	1	Temporal.Synchronous
subsequently	3	Temporal.Asynchronous.Precedence
such as	2	Expansion.Instantiation.Arg2-as-instance
that is	2	Expansion.Equivalence

that is	2	Expansion.Level-of-detail.Arg2-as-detail
then	1	Contingency.Cause.Result Expansion.Conjunction
then	7	Contingency.Cause.Result
then	1	Contingency.Condition.Arg1-as-cond
then	11	Expansion.Conjunction
then	5	Temporal.Asynchronous.Precedence Contingency.Cause.Result
then	310	Temporal.Asynchronous.Precedence
then	4	Temporal.Synchronous
thereafter	11	Temporal.Asynchronous.Precedence
thereby	93	Contingency.Cause.Result Expansion.Manner.Arg1-as-manner
therefore	26	Contingency.Cause.Result
though	91	Comparison.Concession.Arg1-as-denier
though	128	Comparison.Concession.Arg2-as-denier
though	16	Comparison.Contrast
thus	1	Contingency.Caused-Belief.Result+Belief
thus	111	Contingency.Cause.Result
till	1	Contingency.Negative-condition.Arg2-as-negCond
till	4	Temporal.Asynchronous.Precedence
ultimately	1	Contingency.Cause.Reason
ultimately	1	Expansion.Conjunction
ultimately	1	Temporal.Asynchronous.Precedence Contingency.Cause.Result
ultimately	15	Temporal.Asynchronous.Precedence
unless	98	Contingency.Negative-condition.Arg2-as-negCond
until	17	Contingency.Condition.Arg2-as-cond
until	143	Temporal.Asynchronous.Precedence
until	2	Temporal.Asynchronous.Succession Contingency.Condition.Arg2-as-cond
until	7	Temporal.Asynchronous.Succession
upon	1	Temporal.Asynchronous.Succession Contingency.Cause.Reason
upon	2	Temporal.Synchronous Contingency.Cause.Reason
upon	2	Temporal.Synchronous
whatever	4	Comparison.Concession.Arg1-as-denier
when	3	Comparison.Concession.Arg1-as-denier
when	1	Comparison.Concession.Arg2-as-denier
when	1	Comparison.Contrast
when	21	Contingency.Cause.Reason
when	12	Contingency.Condition+SpeechAct
when	2	Contingency.Condition.Arg2-as-cond Expansion.Level-of-detail.Arg2-as-detail
when	116	Contingency.Condition.Arg2-as-cond
when	2	Expansion.Manner.Arg2-as-manner
when	1	Temporal.Asynchronous.Precedence Contingency.Condition.Arg2-as-cond
when	4	Temporal.Asynchronous.Precedence

when	1	Temporal.Asynchronous.Succession Contingency.Cause+Belief.Reason+Belief
when	65	Temporal.Asynchronous.Succession Contingency.Cause.Reason
when	1	Temporal.Asynchronous.Succession Contingency.Cause.Result
when	1	Temporal.Asynchronous.Succession Contingency.Condition+SpeechAct
when	11	Temporal.Asynchronous.Succession Contingency.Condition.Arg2-as-cond
when	160	Temporal.Asynchronous.Succession
when	1	Temporal.Synchronous Comparison.Contrast
when	2	Temporal.Synchronous Contingency.Cause+Belief.Reason+Belief
when	90	Temporal.Synchronous Contingency.Cause.Reason
when	2	Temporal.Synchronous Contingency.Cause.Result
when	1	Temporal.Synchronous Contingency.Condition+SpeechAct
when	65	Temporal.Synchronous Contingency.Condition.Arg2-as-cond
when	4	Temporal.Synchronous Expansion.Level-of-detail.Arg2-as-detail
when	509	Temporal.Synchronous
when and if	1	Temporal.Asynchronous.Succession Contingency.Condition.Arg2-as-cond
whenever	9	Contingency.Condition.Arg2-as-cond
where	2	Contingency.Condition.Arg2-as-cond
whereas	5	Comparison.Contrast
whether	7	Comparison.Concession.Arg1-as-denier
while	203	Comparison.Concession.Arg1-as-denier
while	2	Comparison.Concession.Arg2-as-denier
while	140	Comparison.Contrast
while	1	Comparison.Similarity
while	43	Expansion.Conjunction
while	27	Temporal.Synchronous Comparison.Concession.Arg1-as-denier
while	5	Temporal.Synchronous Comparison.Concession.Arg2-as-denier
while	197	Temporal.Synchronous Comparison.Contrast
while	22	Temporal.Synchronous Expansion.Conjunction
while	163	Temporal.Synchronous
with	2	Comparison.Concession.Arg1-as-denier
with	1	Comparison.Contrast
with	5	Contingency.Cause+Belief.Reason+Belief Expansion.Level-of-detail.Arg2-as-detail
with	5	Contingency.Cause+Belief.Reason+Belief
with	3	Contingency.Cause.Reason Expansion.Level-of-detail.Arg2-as-detail
with	109	Contingency.Cause.Reason
with	2	Contingency.Condition.Arg2-as-cond
with	41	Expansion.Conjunction
with	7	Expansion.Instantiation.Arg2-as-instance
with	111	Expansion.Level-of-detail.Arg2-as-detail
with	6	Expansion.Manner.Arg2-as-manner
with	5	Temporal.Synchronous

without	19	Comparison.Concession.Arg2-as-denier
without	1	Contingency.Cause.Reason
without	1	Contingency.Cause.Result
without	9	Contingency.Negative-condition.Arg2-as-negCond
without	2	Expansion.Level-of-detail.Arg2-as-detail
without	62	Expansion.Manner.Arg2-as-manner
yet	96	Comparison.Concession.Arg2-as-denier
yet	4	Comparison.Contrast
yet	2	Expansion.Conjunction

APPENDICE A

Les thèses principales

Le tableau suivant montre les énoncés de thèse principale dans les résumés, lesquels sont chacun associé à un énoncé plus similaire dans la discussion. De plus, chaque énoncé est associé à un identifiant provenant de l'éditeur PubMed (PMID).

PMID	Thèse principale dans le résumé	Thèse principale dans la discussion
11788911	For these reasons, this regimen could not be recommended for a phase III randomized study.	For these reasons this regimen could not be recommended for a phase III randomized study.
11790211	The safety and efficacy of LAC for colon cancer are unknown, and the nature and magnitude of any quality-of-life (QOL) benefit resulting from LAC for colon cancer is also unknown.	Our findings demonstrate that compared with open colectomy, LAC for colon cancer results in statistically sigpatient selection.
16487438	The impact of prolonged subclinical hyperthyroidism on quality of life is unclear.	Most studies on quality of life in subclinical hyperthyroidism have been performed in patients with endogenous subclinical hyperthyroidism.
17664467	For patients with advanced gastric or gastroesophageal cancer (AGGEC) providing clinical benefit with improved palliation is highly desirable.	14 Although chemotherapy can prolong survival and improve QOL in patients with gastric cancer, the precise magnitude of the benefit is difficult to estimate.
20351334	Therefore, VFL may be another option in the second-line treatment of patients with advanced NSCLC.	Therefore, VFL may be another option in the second-line treatment of patients with advanced NSCLC.
23545101	Bacillus Calmette-Guérin and intravesical chemotherapy represent viable adjuvant options for intermediate risk nonmuscle invasive bladder cancer.	These results are in line with those of a recent randomised trial comparing the short term efficacy of gemcitabine with BCG on 80 intermediate-risk NMIBC patients [15].
9093725	The extent to which chemotherapy may relieve tumour-related symptoms, improve quality of life and prolong survival in patients with gastric cancer is not known in spite of the extensive use of this treatment modality.	This study strongly indicates that chemotherapy can prolong survival and improve quality of life in a proportion of the patients with gastric cancer, but the precise magnitude of the survival benefit is difficult to estimate.
9143576	The choice of reconstruction after gastrectomy and the significance of remaining reservoir function is a matter of controversy.	We have found the gastric reservoir function in partially gastrectomized patients to be of identical significance and magnitude as in patients who had a gastric substitute after total gastrectomy [12].
10682964	Oral carbonic anhydrase inhibitors used to treat glaucoma have significant systemic side effects.	In addition, CA inhibition was below the level known to cause side effects, thus demonstrating the systemic safety of topical brinzolamide.
10877373	Brimonidine is a viable alternative to timolol for first-line therapy in glaucoma and ocular hypertension.	The findings of this 4-month study demonstrate that the clinical effectiveness of brimonidine 0.2% twice daily is equivalent to that of timolol 0.5% twice daily as a first-line therapy for

		glaucoma and ocular hypertension and has less chronotropic effect on the heart.
19476406	Fixed combinations of 0.2% brimonidine-0.5% timolol and 2% dorzolamide-0.5% timolol are used to lower intraocular pressure (IOP).	Both brimonidine-timolol and dorzolamide-timolol effectively lowered IOP when added to an ongoing PGA or used as monotherapy in patients with glaucoma or OHT.
20142538	Combination therapy can be used to treat patients for whom monotherapy does not provide sufficient IOP reduction.	Although cost-effective treatment can be achieved for many patients with monotherapy, the combination should be considered when 2 drugs are needed to achieve the therapeutic goal.
20142538	The simplicity, efficacy, and tolerability of FCLT contribute to its utility in clinical practice.	The simplicity of FCLT coupled with the efficacy and tolerability of the formulation contribute to its utility in clinical practice, particularly among patients who require substantial IOP reductions.
21035866	To test the hypothesis that neodymium:yttrium-aluminum-garnet (Nd:YAG) laser peripheral iridotomy (LPI) significantly reduces the incidence of conversion from pigment dispersion syndrome (PDS) with ocular hypertension (OHT) to pigmentary glaucoma (PG).	This study suggests that there is little benefit of Nd:YAG LPI in preventing conversion from PDS with OHT to PG in patients of the type studied.
8628544	Latanoprost has the potential for becoming a new first-line treatment for glaucoma.	Latanoprost offers several potential advantages over currently available medications for glaucoma therapy.
12721238	This form of exercise can be an important component of supportive care for these patients.	This form of exercise can be an important component of supportive care for these patients.
27629859	In patients with chronic hepatitis B, tenofovir disoproxil fumarate (TDF) plus pegylated interferon (PEG-IFN) for 48-weeks results in higher rates of hepatitis B surface antigen (HBsAg) loss than either monotherapy.	In patients with HBeAg-positive and HBeAg-negative CHB without advanced liver disease, combination therapy with PEG-IFN plus TDF for 48 weeks had the most profound impact on HBsAg decline and resulted in higher rates of HBsAg loss compared with TDF or PEGIFN monotherapy, or PEG-IFN plus TDF for 16 weeks.
28291960	Our results suggest that ologen may be a useful alternative to MMC in JOAG.	Thus, our results suggest that ologen may be a useful alternative to MMC in JOAG, but confirmation of this will require further studies.
28454526	In addition, changing medication from PT to NPT might improve subjective satisfaction and tear break up time.	In addition, changing from preservative containing 0.0015% Tafluprost to preservative-free 0.0015% Tafluprost improved subjective satisfaction and tear break up time.
28603112	It is concluded that Safoof akseer e jigar possesses as effective a therapeutic value in treating hepatitis B as allopathic medicine.	It is concluded that Safoof akseer e Jigar possesses a therapeutic value in treating hepatitis B infection/disease and its associated symptoms.
28911991	Cataract surgery at earlier stages by phacoemulsification may be more beneficial than undergoing manual SICS.	Cataract surgery at earlier stages by phacoemulsification would be more beneficial than undergoing manual SICS.
29133641	This reiterates the role of amnion as a safe and effective bleb modulator.	This suggests its possible protective role in saving the already doubly compromised conjunctiva due

		to surgical and MMC insult, from any additional surgical insult.
29316213	Valsartan/cilnidipine could not significantly suppress HMBPS compared with valsartan/hydrochlorothiazide.	These findings indicate that a valsartan/cilnidipine combination could control HMBPS level similar to a valsartan/HCTZ combination.
29329305	The efficacy of switching to tenofovir disoproxil fumarate (TDF) monotherapy from lamivudine (LAM) plus adefovir dipivoxil (ADV) combination therapy (stable switching) in patients with LAM-resistant chronic hepatitis B (CHB) and undetectable hepatitis B virus (HBV) DNA is not clear.	In this study, we showed that stable switching from LAM+ADV to TDF was safe and effective in patients with LAM-resistant CHB and undetectable HBV DNA.
29375225	Spironolactone could be used as the fourth-line therapy in patients with resistant hypertension.	There are some studies on spironolactone in the treatment of resistant hypertension.
29428751	Regular consumption of this product may be used as a potential functional food.	Hence, daily consumption of fermented milk with <i>L. lactis</i> NRRL B-50571 may be used as a potential functional food.
29428785	Increasing evidence suggests that iron-containing multiple micronutrient may reduce the risk of pregnancy-induced hypertension including gestational hypertension or preeclampsia.	Thus, the association of antioxidant vitamins with hypertension during pregnancy has been inconsistent between studies.
29452204	It remains unresolved how children with immune-tolerant CHB should be treated.	However, whether these immune-tolerant children are indicated for antiviral treatment remains controversial.
29507099	Available data indicate that dietary sodium (as salt) relates directly to blood pressure (BP).	Mechanisms and pathways to explain the independent and interacting effects of DASH diet and Na on BP remain to be elucidated.
29510792	Carbonic anhydrase inhibition may constitute a potential target for drug therapy in patients with sleep apnea and comorbid hypertension.	Our findings suggest that CA activity influences BP regulation and sleep-disordered breathing in patients with OSA.
29553852	Hypertension is a generally accepted atherogenic risk factor.	One of the traditional risk factors most associated with c-IMT is BP (13).
29553852	By lifestyle and medical intervention to control these risk factors may prevent progression of c-IMT in old-aged cohort with hypertension.	In the present study on longitudinal evolution of c-IMT in the oldaged cohort with hypertension over 4 years, we observed a significant progression in c-IMT z-scores which associated with the CV risk factors.
29565948	Sodium nitrite has been reported to be effective in reducing chronic peripheral pain.	This study demonstrates that SR-nitrite eliminates the headaches and dizziness observed with an immediate release formulation of NaNO ₂ , and the potential of SR-nitrite to alleviate pain in patients with diabetic neuropathy with few side effects.
29571976	Recently, there has been an increasing interest in the influence of antioxidant vitamins on the efficacy of oral hypoglycemic therapy in type 2 diabetic patients (T2DM).	Therefore, we designed this study to investigate the possible contribution of vitamins C and/or E supplementation on the efficacy of oral hypoglycemic therapy (Metformin) in T2DM male patients.

29577550	Methionine aminopeptidase 2 (MetAP2) inhibition has been shown to result in significant weight loss and improved glucose control.	These changes suggest that patients treated with ZGN-1061 may experience improvement in glucose control and significant reduction in body weight as observed in pre-clinical studies of ZGN-1061 in obese mice and in clinical trials of other MetAP2 inhibitors.
29588137	Low serum vitamin D has been associated with metabolic syndrome and Non-alcoholic fatty liver disease (NAFLD).	Furthermore, the severity of vitamin D deficiency has been associated with NAFLD severity in histopathology and patients with lower vitamin D had more advanced liver steatosis and fibrosis [20,21].
29588138	Cardiovascular complications are one of main cause of increased mortality and morbidity among Diabetes Mellitus (DM) patients.	We found that supplementation of EPA is very efficient in the reduction of Met and Cys which may attribute to the prevention of vascular complications in the T2DM patients.
29588138	Altered metabolism of sulphur amino acids in diabetes reflected as increases in concentration of methionine and cysteine/cystine in the blood which known as a markers of Cardiovascular Diseases (CVD).	The aggregate of results revealed that impaired insulin action is associated with increases in blood level of methionine and cysteine/cystine [10].
29605574	Multiple studies have evaluated the hypoglycemic effect of cinnamon in patients with diabetes mellitus (DM) type II, with conflicting results.	Different mechanisms have been suggested for the observed hypoglycemic effects of cinnamon [40].
29620663	Evidence has shown that long-term sodium reduction can not only reduce blood pressure, but also provide cardiovascular benefits.	Numerous clinical studies have demonstrated the benefits of reducing BP in ISH patients.
29643323	Treatment-related quality of life (QOL) is an important aspect of diabetes management.	While our original data suggested that linagliptin provides clinically meaningful improvement in glycemic control without any unacceptable side effects and with a low risk of hypoglycemia [18], this sub-analysis revealed the positive influence of linagliptin on treatment-related QOL, which is another important aspect of diabetes management.
29652567	KAMUT khorasan is an ancient grain with widely acclaimed health benefits.	High selenium levels are a widely reported characteristic of ancient khorasan wheat.
29654388	The immunosuppressive drug rapamycin may influence insulin sensitivity in insulin-responsive tissues.	A reduction in the need for exogenous insulin may derive from an increase in beta-cell secretion or an improvement in insulin sensitivity.
29654849	Fracture risk is increased in individuals with type 2 diabetes (T2D).	While increased bone resorption can be interpreted as an increase in fracture risk, these changes in CTX during intensified antidiabetic treatment may also reflect a normalisation of bone remodelling.
15139072	Research suggests that stress-reduction programs tailored to the cancer setting help patients cope with the effects of treatment and improve their quality of life.	Although research into the efficacy and mechanisms of yoga is in its beginning stages, the findings reported to date are supportive ¹³ and, along with our finding of improved sleep, suggest that the health effects of yoga in cancer patients should be explored further.

16524633	Morbidity is an important issue in cancer research.	Recording and reporting the morbidity of cancer treatment is an important issue, which can be approached in various ways.
18485819	Debilitating hot flushes and sleep dysfunction often affect survivors of breast cancer, most notably in those taking anti-oestrogen medications.	Despite this belief, she reported no effect on the quantity or intensity of her hot flushes.
18485819	Conventional treatments have been only partially effective in diminishing these issues, and some have serious risks.	Severe complications occurred in only 1.7 of 1000 procedures.
18485819	Long-term relief of symptoms has the potential to improve overall quality of life and increase compliance with anti-oestrogen medications for breast cancer.	Despite this belief, she reported no effect on the quantity or intensity of her hot flushes.
20733132	Therefore, we recommend no additional clinical use of TCG in this population.	More than 10,000 patients have been enrolled onto intergroup trials that evaluated anthracyclines, topotecan, and gemcitabine incorporated either in triple-drug regimens, within sequential doublets, or as sequential single-agent therapy.
21901389	The universality of effects across levels of expectation indicates a potential to utilize this stress reduction approach as complementary therapy in oncologic practice.	This finding has promising implications for wide applicability in clinical practice because even if a patient might be doubtful of the benefits of stress reduction techniques, a strong potential for positive psychosocial outcome still exists.
21993852	Both procedures seem acceptable as standard reconstructions after distal gastrectomy with regard to postoperative QOL and dysfunction.	We found no significant difference between R-Y and B-I procedures in terms of overall postoperative dysfunction.
23632471	Incorporating patients' perspectives into a dietary intervention after gastrectomy for gastric cancer may contribute to improved patient outcomes and quality care.	Incorporating patient perspectives into a dietary intervention contributes to better patient outcomes and quality care.
23768153	VR to assist breast cancer patients in the return to work process is an important component of cancer survivorship plans.	VR to assist breast cancer patients in the return to work process is therefore an important component of cancer survivorship plans.
23840308	Delirium is a serious and common postoperative complication, especially in frail elderly patients.	Within this study, geriatric liaison intervention for the prevention of postoperative delirium in frail patients in a general oncological surgical population has not proven to be effective.
24007947	Analgesia and early quality of recovery may be improved by epidural analgesia.	Epidural analgesia does not improve survival but has been proposed as beneficial in abdominal surgery because of improved analgesia, reduced respiratory complications and earlier return of bowel function [10,12,30,31].
24019545	Tivozanib is a potent and selective tyrosine kinase inhibitor of vascular endothelial growth factor receptor 1 (VEGFR1), -2, and -3.	Inhibiting the VEGF pathway has been proposed as the primary mechanism for tumor inhibition for clear cell RCC. ²⁷ Tivozanib is characterized from preclinical data as a highly potent, selective inhibitor of VEGFR, with less inhibition of other kinases, including c-KIT and platelet-derived growth factor receptor beta, ¹²⁻¹⁴ and a long half-life compared with sorafenib.

24028441	Non-randomised studies have suggested that the postoperative complications of (Campos LS, Limberger LF, Stein AT, Kalil AN) laparoscopic radical hysterectomy are similar to those in abdominal radical hysterectomy.	Laparoscopic radical hysterectomy had lower pain scores compared with abdominal radical hysterectomy in this series.
24040301	Postcancer fatigue is a frequently occurring problem, impairing quality of life.	Altogether, the results of our study could not confirm the hypothesis that metabolic and volumetric traits observed in CFS patients are a reflection of fatigue in general and are, therefore, also of importance for patients suffering from postcancer fatigue.
8683634	Combined radiation therapy and chemotherapy after surgery, compared with postsurgical radiation therapy alone, has been shown to improve disease-free survival and overall survival significantly among patients with poor-prognosis (i.e., advanced stage disease or metastasis to regional lymph nodes) resectable rectal cancer.	For patients with poor-prognosis resectable rectal cancer, additional modest improvements in disease control and in overall survival may be achieved by optimizing the use of 5-FU together with radiation therapy {2223}.
9643663	Thus 'Casodex' 150 mg monotherapy is an option for patients with M1 prostate cancer for whom surgical or medical castration is not indicated or is not acceptable.	This supports the choice of 'Casodex' 150 mg monotherapy as an option for patients with M1 prostate cancer for whom surgical or medical castration is not indicated or is not acceptable.
10561201	A combination of mitoxantrone plus prednisone is preferable to prednisone alone for reduction of pain in men with metastatic, hormone-resistant, prostate cancer.	The Canadian study of mitoxantrone plus prednisone versus prednisone alone for men with symptomatic hormoneresistant prostate cancer has provided a valuable database for analyzing HQL during treatment.
10675381	In nonrandomized studies involving patients with different tumor types including non-small-cell lung cancer (NSCLC), ATP infusion appeared to inhibit loss of weight and deterioration of quality of life (QOL) and performance status.	In conclusion, our results demonstrate that ATP infusion has marked beneficial effects on the QOL of patients with advanced NSCLC.
10880548	Menopausal symptoms (e.g., hot flashes, vaginal dryness, and stress urinary incontinence) are very common in breast cancer survivors and cannot be managed with standard estrogen replacement therapy (ERT) in these patients.	Vaginal dryness was one of the strongest predictors of sexual dysfunction in both breast cancer survivors and postmenopausal women.
11034245	Intravenous 5-fluorouracil (5-FU) and leucovorin has been a mainstay of chemotherapy, but oral 5-FU derivatives have been shown a comparable antitumor activity.	12 Bajetta et al. ¹³ compared toxicity and therapeutic efficacy between intravenous 5-FU (500 mg/m ²) and intravenous doxifluridine (4000 mg/m ²).
11181777	The combination of etoposide plus cisplatin (EP) is considered to be standard therapy for small-cell lung cancer (SCLC).	In our study, the standard group received the EP combination every 4 weeks.
12122096	The role of adjuvant chemotherapy in postmenopausal patients with lymph node-negative breast cancer is controversial.	IBCSG Trial IX is the largest randomized trial comparing chemotherapy followed by tamoxifen versus tamoxifen alone for postmenopausal women with lymph node-negative breast cancer.

12953086	Despite considerable improvement in the treatment of advanced ovarian cancer, the optimization of efficacy and tolerability remains an important issue.	Hence, the TC combination should be considered an important option for the first-line treatment of patients with advanced ovarian cancer.
15472910	The study underscores the value of undertaking studies in areas of disease prevalence and the necessity of selecting appropriate outcome measures.	Our study also shows the importance of undertaking clinical trials in areas of disease prevalence.
15547181	Thus, docetaxel-carboplatin represents an alternative first-line chemotherapy regimen for patients with newly diagnosed ovarian cancer.	In conclusion, treatment with docetaxel-carboplatin should be viewed as an alternative to treatment with paclitaxel-carboplatin for newly diagnosed stage Ic-IV ovarian cancer.
15659490	The role of high-dose chemotherapy (HDCT) in metastatic breast cancer remains controversial.	In conclusion, the role of HDCT in metastatic breast cancer remains controversial.
15753541	Thalidomide, which is an inhibitor of TNF-alpha synthesis, may represent a novel and rational approach to the treatment of cancer cachexia.	Thalidomide has also recently been shown to inhibit lipopolysaccharide mediated induction of cyclooxygenase 2 and prostaglandin E2 which may represent an alternative pathway by which it can promote an anti-inflammatory response.
16179098	Standard chemotherapy regimens are generally considered too toxic for elderly patients.	The regimen was very well tolerated, and the authors suggested that elderly patients may benefit from its minimal toxicity.
16446322	For malignant pleural mesothelioma (MPM) patients with a poor prognosis, maintaining health-related quality of life (HRQOL) is important.	In palliative situations where a disease, such as MPM, is rapidly fatal, this should be regarded as an important outcome.
16446322	This study provides important information about the HRQOL of chemotherapy-treated MPM patients.	Given that survival was increased significantly in the cisplatin/raltitrexed arm, and that there was stabilization on global HRQOL, these findings provide useful insights to help treat symptoms and understand the impact of cisplatin/raltitrexed in selected MPM patients.
16598426	Experimental and early clinical investigations have demonstrated encouraging results for estramustine in the treatment of malignant glioma.	In addition, some clinical evaluation [15] has given support to evaluate estramustine in a controlled clinical setting.
16670385	Sentinel lymph node biopsy in women with operable breast cancer is routinely used in some countries for staging the axilla despite limited data from randomized trials on morbidity and mortality outcomes.	In conclusion, this study has shown that sentinel lymph node biopsy is a safe and effective alternative to routine axillary dissection for nodal staging in early-stage breast cancer.
16934909	Targeted therapies may improve chemotherapy effectiveness without worsening toxicity.	The use of second- and third-line therapies for patients with NSCLC is associated with benefits in both overall survival and slowed rate of decline in quality of life.
16944295	The impact of treatment on health-related quality of life (HRQoL) is an important consideration in the adjuvant treatment of operable breast cancer.	These are the first published results from a longitudinal follow-up of the impact of 5 years of adjuvant aromatase inhibitor therapy on HRQoL.

16981945	There is no consensus on an accepted intervention that significantly reduces its severity.	Only nine studies showed a potential benefit in preventing or reducing the severity of mucositis, although no single intervention could be strongly recommended on the currently available evidence.
16981945	There is a paucity of high-level evidence on potentially useful interventions and a continued need for new and innovative research, incorporating quality-of-life measurements, in patients experiencing radiation-induced mucositis.	There is currently a paucity of high-level evidence on potentially useful interventions and consequently a continued need for new and innovative research that should also incorporate quality-of-life assessments in patients experiencing radiation-induced mucositis.
17090981	Quality of life is an important issue in pancreatic cancer patients.	We reviewed the results or the toxicity of treatment of unresectable pancreatic cancer, described in several aforementioned clinical trials.
17251532	The efficacy and safety of uterine-artery embolization, as compared with standard surgical methods, for the treatment of symptomatic uterine fibroids remain uncertain.	Two other randomized, controlled trials compared uterine-artery embolization with hysterectomy.
17324657	The ADA MNT protocol for radiation oncology is a useful guide to the level of nutrition support required.	The ADA MNT for radiation oncology patients is a useful guide to the level of nutrition support required.
17664468	Therapy of patients with advanced gastric or gastroesophageal junction cancer should provide symptom relief and improve quality of life (QOL) because most patients are symptomatic at baseline.	The V325 trial is the largest randomized, controlled trial that formally and systematically assessed QOL in patients with advanced gastric cancer or gastroesophageal junction.
17912635	Relapse after completing adjuvant tamoxifen therapy is a persistent threat for women with hormone-responsive breast cancer.	These findings support the concept that distant micrometastases that have survived 5 years of tamoxifen therapy remain highly estrogen-sensitive and responsive to extended adjuvant letrozole treatment.
18064563	Women with breast cancer may have an even higher risk of depression particularly in a postmenopausal or estrogen deficiency state.	An antidepressant should be considered for early stage breast cancer patients with depressive symptoms who are receiving adjuvant treatment.
18093738	Methods are needed to reduce professional time required while achieving sustained intervention effectiveness.	Future efforts that combine patient and provider targeted methods may improve these broader long-term outcomes [2,15].
18285602	Dermatitis is a frequent adverse effect of adjuvant breast radiotherapy.	In 2002, Vicini et al [13] reported the first clinical use of breast IMRT in a prospective series and suggested a reduced occurrence of acute skin adverse effects.
18285602	It is more likely in full-breasted women and when the radiation is distributed nonhomogeneously in the breast.	CT simulation, 3D dose distribution treatment planning systems, and treatment unit with multileaf collimation are now becoming standard equipment in radiation oncology, such that breast IMRT should be offered to patients receiving adjuvant breast RT instead of the standard wedge technique.

18809612	Hot flashes are a significant problem for many breast cancer survivors.	The mechanism of action of hypnosis to reduce hot flashes is unknown, and this question was not addressed in this study.
18809612	Hot flashes can cause discomfort, disrupted sleep, anxiety, and decreased quality of life.	It has been established previously that women with unrelieved hot flashes suffer negative emotional consequences, such as anxiety, depression, interrupted sleep, and decreased quality of life.
19194996	There were no significant effects of the intervention on mood, QOL or healthcare utilisation.	We found no difference between the intervention group and control group in terms of general healthcare utilisation or unscheduled cancer-related appointments; however, participants in both conditions reported a decreased use of general healthcare facilities in the 6 months following completion of treatment.
19199105	More efficient treatment protocol is needed for clinical practice.	More efficient treatment protocol is needed for clinical practice.
19687337	Exercise training to improve cardiovascular fitness should be considered in the management of lymphoma patients.	33,34 Exercise training may be considered as a supportive care intervention for lymphoma patients receiving chemotherapy or off treatments.
19720909	Differences in chemotherapy scheduling, pre-existing morbidity, and toxicity are important in individualizing therapy.	Differences in scheduling are also important in determining which regimen is best for an individual patient, but unfortunately, all four regimens require multiple days of therapy per cycle.
19858379	On the basis of our trial and the meta-analysis, GEM-CAP should be considered as one of the standard first-line options in locally advanced and metastatic pancreatic cancer.	On the basis of our trial and the meta-analysis, GEM-CAP should be considered as one of the standard first-line options in locally advanced and metastatic pancreatic cancer.
20881891	Safe and effective interventions are needed to alleviate sexual dysfunction.	Safe and effective interventions are needed to alleviate sexual dysfunction.
20885302	Systematic assessment of depression is needed in hospice patients.	Thus, the addition of the systematic assessment of depression as an add-on to usual care probably had a greater effect because it is not a symptom on which hospice staff normally focus their efforts.
21236344	Article from the Special issue on Targeted Inhibitors.	Therefore, COX-2 has become a therapeutic target.
21237636	Hence, nurse-led telephone follow-up seems an appropriate way to reduce clinic visits and represents an accepted alternative strategy.	Hence, nurse-led telephone follow-up seems an appropriate way to reduce clinic visits and represents an accepted alternative strategy.
21444867	Despite its efficacy in reducing locoregional recurrence, EBRT should be avoided in patients with low- and intermediate-risk EC.	Combined with the 15-year outcome results of the PORTEC-1 trial, it is clear that EBRT should be avoided in patients with low- and intermediaterisk EC.
21538678	Insomnia is a common complaint among cancer survivors.	In fact, the SHUTi effect sizes for insomnia severity (d 5 2.24), sleep efficiency (d 5 1.05), SOL (d 5 0.83), and WASO (d 5 0.72) were comparable to those found in a study of faceto-face group CBT-I for cancer survivors (ISI d 5

		1.78; sleep efficiency d 5 1.33; SOL d 5 0.95; WASO d 5 1.30) [25].
21751205	Anemia is an expected consequence of intensive chemotherapy regimens administered to patients with acute leukemia.	In this study using epoetin alpha during intensive induction and consolidation chemotherapy (hyper-CVAD or augmented BFM) in patients with newly diagnosed ALL, LL, or BL, we observed a modest but statistically significant decrease in the number of PRBC units transfused per week compared with the no epoetin alpha arm.
21751295	The generalizability of palliative care intervention research is often limited by high rates of study attrition.	Research involving patients with advanced cancer is often plagued by high levels of attrition.
21871978	The postoperative clinical superiority of the interposition of jejunum reconstruction (INT) to Roux-en-Y reconstruction (RY) after total gastrectomy has not been clarified.	13,21 Iwahashi et al ²¹ showed that a double tract was not superior to RY after total gastrectomy through a randomized prospective study.
21871978	The safer and simpler RY method may be a more suitable reconstruction method than INT after total gastrectomy.	RY reconstruction seems to be beneficial after total gastrectomy.
22113257	Lifestyle intervention research is feasible in a cooperative group setting.	Our study also demonstrated the feasibility of conducting lifestyle research in a cooperative group setting.
22198049	Multimodality therapies for cachexia ideally should be introduced within a context of "best supportive care" that includes optimal symptom management and careful psychosocial counseling.	Our study also demonstrates that multimodality therapies for cachexia must not have significant side effects and should ideally be introduced within a context of the "best supportive care", which includes optimal symptom management and careful psychosocial counseling.
22282373	Cisplatin reduces plasma ghrelin levels through the 5-hydroxytryptamine (5-HT) receptor.	It has been reported that acute gastrointestinal disorders caused by cisplatin involve 5-HT secretion from the enterochromaffin cells in association with 5-HT ₃ receptors.
22336221	Patients completed the Lung Cancer Symptom Scale (LCSS) at baseline, after each cycle, and post-discontinuation.	Some patients probably had worsening symptoms that were not reported, either because of missed LCSS assessments or because symptoms occurred after all scheduled LCSS assessments were completed.
22340966	Hormone replacement therapy is often either undesirable or contraindicated.	Additionally, we did not control for all potential confounding factors that could have an effect on menopausal symptoms, such as the use of drugs to manage HFNS and the use of adjuvant hormone therapy.
22340966	Safe, effective non-hormonal treatments are needed.	The cost of delivering the treatment needs to be assessed, with a need for comparison with non-hormonal drugs.
22430268	MBSR is recommended to support survivors of breast cancer.	The end of hospital treatment is known as a critical point for psychological problems ⁴ and is a time when people are more likely to seek the help of MBSR at an integrative center like The Haven.
22494567	Maintenance treatment with erlotinib appears to be a useful option for the management of	Erlotinib maintenance treatment thus represents a novel treatment option in Asian patients with

	Asian patients with advanced NSCLC without progression after first-line chemotherapy.	advanced NSCLC who have not progressed after first-line chemotherapy.
22825506	The incidence and development of cancer are closely related to dysfunction of immune function.	Besides inducing antigen-specific cytotoxic T lymphocytes, DC can also directly or indirectly influence proliferation of B cells and activate humoral immunoresponse ¹⁵ .
23254324	Rigorous standard operating procedures for data collection are required.	Piloting of data collection was an important aim of this study.
23466360	Incorporating MRI-based dose-volume histogram analysis into the treatment planning process may reduce acute urinary morbidity after brachytherapy.	McLaughlin et al. (10) studied visualization of the prostate using MRI and the potential implications for treatment planning.
23525453	Malignant pleural effusion (MPE) is a common complication of advanced non-small cell lung cancer (NSCLC).	Malignant pleural effusion (MPE) is seen in more than 50% of patients presenting with NSCLC (1,2).
23525453	Bevacizumab, a humanized monoclonal antibody against vascular endothelial growth factor (VEGF), has been shown to be efficient in suppressing the accumulation of pleural fluid.	The ability of VEGF to promote vascular permeability renders it a key instigator of MPE (28,29).
23525453	We propose that VEGF expression levels in MPE could serve as a prognostic marker for bevacizumab therapy.	Therefore, treatment with an antibody specific to VEGF in patients with high levels of VEGF in the pleural effusion suggests that VEGF can be considered a biomarker for bevacizumab treatment of MPE (18,51,52).
9414057	The use of the Pain Education Program by nurses should be seriously considered on oncology units.	Because of the pain relieving effects, the use of such an intervention should be seriously considered on oncology units.

APPENDICE B

Les connecteurs de discours contenus dans les énoncés de thèse principale du résumé

Le tableau suivant montre tous les connecteurs de discours contenus dans les énoncés de thèse principale du résumé. Pour chaque connecteur, la capacité de prédiction de l'appartenance d'un énoncé au statut de thèse principale dans le résumé est donnée en termes de précision, de rappel et de F-mesure.

<i>Connecteurs</i>	<i>Nombre de thèses principales</i>	<i>Précision</i>	<i>Rappel</i>	<i>F-mesure</i>
<i>but also</i>	1	0,500	0,008	0,016
<i>earlier</i>	1	0,500	0,008	0,016
<i>hence</i>	1	0,500	0,008	0,016
<i>not only+but also</i>	1	0,500	0,008	0,016
<i>thus</i>	2	0,400	0,016	0,031
<i>despite</i>	3	0,375	0,024	0,046
<i>not only</i>	1	0,333	0,008	0,016
<i>not only+also</i>	1	0,333	0,008	0,016
<i>not only+but</i>	1	0,333	0,008	0,016
<i>therefore</i>	2	0,286	0,016	0,031
<i>because</i>	1	0,125	0,008	0,015
<i>as</i>	11	0,110	0,089	0,099
<i>plus</i>	4	0,105	0,033	0,050
<i>for</i>	30	0,101	0,244	0,143
<i>about</i>	1	0,100	0,008	0,015
<i>also</i>	2	0,087	0,016	0,027
<i>in</i>	49	0,086	0,398	0,141
<i>in addition</i>	1	0,083	0,008	0,015
<i>on</i>	10	0,079	0,081	0,080
<i>after</i>	10	0,079	0,081	0,080
<i>without</i>	2	0,077	0,016	0,027
<i>with</i>	34	0,075	0,276	0,118
<i>when</i>	1	0,071	0,008	0,015
<i>from</i>	6	0,066	0,049	0,056
<i>only</i>	2	0,065	0,016	0,026
<i>and</i>	48	0,064	0,390	0,110
<i>while</i>	1	0,063	0,008	0,014
<i>by</i>	5	0,053	0,041	0,046
<i>but</i>	2	0,049	0,016	0,024
<i>or</i>	8	0,047	0,065	0,055

APPENDICE C

Les connecteurs de discours contenus dans les énoncés de thèse principale de la discussion

Le tableau suivant montre tous les connecteurs de discours contenus dans les énoncés de thèse principale de la discussion. Pour chaque connecteur, la capacité de prédiction de l'appartenance d'un énoncé au statut de thèse principale dans la discussion est donnée en termes de précision, de rappel et de F-mesure.

Connecteurs	Nombre de thèses principales	Précision	Rappel	F-mesure
<i>as an alternative</i>	1	0,333	0,008	0,016
<i>besides</i>	1	0,333	0,008	0,016
<i>instead of</i>	1	0,250	0,008	0,016
<i>hence</i>	3	0,214	0,025	0,044
<i>instead</i>	1	0,167	0,008	0,016
<i>along with</i>	1	0,143	0,008	0,016
<i>consequently</i>	1	0,125	0,008	0,015
<i>even if</i>	1	0,111	0,008	0,015
<i>earlier</i>	2	0,071	0,016	0,027
<i>in fact</i>	1	0,067	0,008	0,015
<i>because of</i>	3	0,065	0,025	0,036
<i>like</i>	1	0,059	0,008	0,014
<i>therefore</i>	6	0,050	0,049	0,050
<i>thus</i>	4	0,046	0,033	0,038
<i>plus</i>	2	0,044	0,016	0,024
<i>either+or</i>	2	0,040	0,016	0,023
<i>for</i>	34	0,038	0,279	0,066
<i>because</i>	5	0,036	0,041	0,038
<i>on</i>	19	0,034	0,156	0,056
<i>in addition</i>	3	0,034	0,025	0,029
<i>with</i>	46	0,032	0,377	0,060
<i>or</i>	15	0,030	0,123	0,048
<i>also</i>	9	0,030	0,074	0,042
<i>from</i>	10	0,029	0,082	0,043
<i>as</i>	18	0,028	0,148	0,048
<i>in</i>	72	0,028	0,590	0,053
<i>where</i>	1	0,026	0,008	0,012
<i>without</i>	2	0,026	0,016	0,020
<i>when</i>	3	0,025	0,025	0,025
<i>while</i>	2	0,025	0,016	0,020

<i>and</i>	57	0,025	0,467	0,047
<i>given</i>	2	0,024	0,016	0,019
<i>such as</i>	3	0,024	0,025	0,024
<i>but</i>	5	0,023	0,041	0,030
<i>still</i>	1	0,023	0,008	0,012
<i>furthermore</i>	1	0,022	0,008	0,012
<i>previously</i>	1	0,021	0,008	0,012
<i>although</i>	4	0,020	0,033	0,025
<i>both+and</i>	3	0,019	0,025	0,022
<i>whether</i>	1	0,019	0,008	0,011
<i>further</i>	2	0,018	0,016	0,017
<i>despite</i>	1	0,018	0,008	0,011
<i>after</i>	5	0,017	0,041	0,024
<i>if</i>	1	0,016	0,008	0,011
<i>by</i>	7	0,015	0,057	0,023
<i>due to</i>	1	0,013	0,008	0,010
<i>only</i>	2	0,010	0,016	0,013
<i>however</i>	2	0,007	0,016	0,010

BIBLIOGRAPHIE

- Abbott, B. (2004). Definiteness and Indefiniteness. Dans L. R. Horn et G. Ward (dir.), *Handbook of Pragmatics*. Blackwell.
- Abeykoon, V., Fox, G., Kim, M., Ekanayake, S., Kamburugamuve, S., Govindarajan, K., Wickramasinghe, P., Perera, N., Widanage, C., Uyar, A., Gunduz, G. et Akkas, S. (2022). Stochastic gradient descent-based support vector machines training optimization on Big Data and HPC frameworks. *Concurrency and Computation: Practice and Experience*, 34(8), e6292. <https://doi.org/10.1002/cpe.6292>
- Agteresch, H. J., Dagnelie, P. C., van der Gaast, A., Stijnen, T. et Wilson, J. H. P. (2000). Randomized Clinical Trial of Adenosine 5'-Triphosphate in Patients With Advanced Non-Small-Cell Lung Cancer. *JNCI: Journal of the National Cancer Institute*, 92(4), 321-328. <https://doi.org/10.1093/jnci/92.4.321>
- Aizerman, M. A., Braverman, E. M. et Rozoner, L. I. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Autom Remote Control*, 25, 821-837.
- Alghamdi, R. et Alfalqi, K. (2015). A Survey of Topic Modeling in Text Mining. *International Journal of Advanced Computer Science and Applications*, 6(1). <https://doi.org/10.14569/IJACSA.2015.060121>
- Al-Khatib, K., Wachsmuth, H., Hagen, M., Köhler, J. et Stein, B. (2016). Cross-Domain Mining of Argumentative Text through Distant Supervision. Dans *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (p. 1395-1404). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N16-1165>
- Allahyari, M., Pouriye, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B. et Kochut, K. (2017, 28 juillet). *Text Summarization Techniques: A Brief Survey*. arXiv. Récupéré le 17 octobre 2022 de <http://arxiv.org/abs/1707.02268>
- Amgoud, L. et Vesic, S. (2014). Rich preference-based argumentation frameworks. *International Journal of Approximate Reasoning*, 55(n° 2), 585-606. <https://doi.org/10.1016/j.ijar.2013.10.010>
- Ammann, H. (1925). *Die menschliche Rede. Sprachphilosophische Untersuchungen* (Moritz Schauenburg). <https://www.abebooks.com/first-edition/menschliche-Rede-Sprachphilosophische-Untersuchungen-Teile-Band/11140750747/bd>
- Anand, P., King, J., Body-Graber, J., Wagner, E., Martell, C., Oard, D. et Resnik, P. (2011). *Believe Me - We Can Do This! Annotating Persuasive Acts in Blog Text*, 5.
- Anderson, J. R. (1990). *The adaptive character of thought* (p. xii, 276). Lawrence Erlbaum Associates, Inc.
- Anderson, M. H. et Lemken, R. K. (2019). An Empirical Assessment of the Influence of March and Simon's Organizations: The Realized Contribution and Unfulfilled Promise of a Masterpiece. *Journal of Management Studies*, 56(8), 1537-1569. <https://doi.org/10.1111/joms.12527>

- Anderson, M. H. et Lemken, R. K. (2023). Citation Context Analysis as a Method for Conducting Rigorous and Impactful Literature Reviews. *Organizational Research Methods*, 26(1), 77-106. <https://doi.org/10.1177/1094428120969905>
- Andrade, C. (2011). How to write a good abstract for a scientific paper or conference presentation. *Indian Journal of Psychiatry*, 53(2), 172-175. <https://doi.org/10.4103/0019-5545.82558>
- Angenot, M. (1985). *Critique de la raison sémiotique: fragment avec pin up*. Presses de l'Université de Montréal.
- Anscombre, J.-C. et Delahaie, J. (2014). Connecteurs et connexion, un état des lieux. *Cahiers de Lexicologie*, 105, 161-180.
- Anscombre, J.-C., Donaire, M. L. et Haillet, P. P. (2013). *Opérateurs discursifs du français*. <https://www.peterlang.com/document/1053015>
- Apel, K. O. (1981). La question d'une fondation ultime de la raison. *Critique*, 413, 895-928.
- Apel, K.-O. (1975). *Die Idee der Sprache in der Tradition des Humanismus von Dante bis Vico* (2., durchges. Aufl. edition). Bouvier.
- Aristote. (2011). *Rhétorique. Livre III* (M. Dufour et A. Wartelle, trad.). <https://www.lesbelleslettres.com/livre/9782251000657/rhetorique-tome-iii-livre-iii>
- Armağan, A. (2013). How to write an introduction section of a scientific article? *Turkish journal of urology*, 39(Suppl 1), 8-9. <https://doi.org/10.5152/tud.2013.046>
- Arthur, D. et Vassilvitskii, S. (2007). k-means++: the advantages of careful seeding. Dans *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (p. 1027-1035). Society for Industrial and Applied Mathematics.
- Aston, G. (1977). Comprehending Value: Aspects of the Structure of Argumentative Discourse. *Comprehending Value: Aspects of the Structure of Argumentative Discourse*, 6(3), 465-509.
- Bakhtine, M. (1978). *Esthétique et théorie du roman* (Gallimard, O. Daria, trad.). <https://www.librairie-gallimard.com/livre/9782070299331-esthetique-et-theorie-du-roman-mikhail-bakhtine/>
- Ball, R. (2009). Scholarly communication in transition: The use of question marks in the titles of scientific articles in medicine, life sciences and physics 1966–2005. *Scientometrics*, 79(3), 667-679. <https://doi.org/10.1007/s11192-007-1984-5>
- Bar-Haim, R., Bhattacharya, I., Dinuzzo, F., Saha, A. et Slonim, N. (2017). Stance Classification of Context-Dependent Claims. Dans *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (p. 251-261). Association for Computational Linguistics. <https://aclanthology.org/E17-1024>
- Barone, C., Corsi, D. C., Pozzo, C., Cassano, A., Fontana, T., Noviello, M. R., Landriscina, M., Colloca, G. et Astone, A. (1998). Treatment of patients with advanced gastric carcinoma with a 5-fluorouracil-

- based on a cisplatin-based regimen: two parallel randomized phase II studies. *Cancer*, 82(8), 1460-1467. [https://doi.org/10.1002/\(sici\)1097-0142\(19980415\)82:8<1460::aid-cnrcr5>3.0.co;2-d](https://doi.org/10.1002/(sici)1097-0142(19980415)82:8<1460::aid-cnrcr5>3.0.co;2-d)
- Barrett, T. (2011). *Criticizing Art: Understanding the Contemporary* (3rd edition). McGraw Hill.
- Bart, C. L. (1998). *Le discours politique*. Presses Universitaires de France - PUF.
- Barthes, R. (1970). L'ancienne rhétorique. *Communications*, 16(1), 172-223. <https://doi.org/10.3406/comm.1970.1236>
- Basile, T. M. A., Mauro, N. D., Esposito, F., Ferilli, S. et Vergari, A. (2018). Density Estimators for Positive-Unlabeled Learning. Dans *New Frontiers in Mining Complex Patterns: 6th International Workshop, NFMCP 2017, Held in Conjunction with ECML-PKDD 2017, Skopje, Macedonia, September 18-22, 2017, Revised Selected Papers* (p. 49-64). Springer International Publishing. <https://www.research.ed.ac.uk/en/publications/density-estimators-for-positive-unlabeled-learning>
- Beardsley, M. C. (1950). *Practical logic*. Prentice-Hall.
- Bekker, J. et Davis, J. (2020). Learning from positive and unlabeled data: a survey. *Machine Learning*, 109(4), 719-760. <https://doi.org/10.1007/s10994-020-05877-5>
- Ben Ayed, A., Biskri, I. et Guy Meunier, J. (2020). Vector Space Modeling based Evaluation of Automatically Generated Text Summaries. *International Journal on Natural Language Computing*, 9(3), 43-52. <https://doi.org/10.5121/ijnlc.2020.9303>
- Bench-Capon, T. J. M. et Dunne, P. E. (2007). Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10), 619-641. <https://doi.org/10.1016/j.artint.2007.05.001>
- Benhadid, I., Meunier, J.-G., Hamidi, S., Remaki, Z. et Nyongwa, M. (1998). Étude expérimentale comparative des méthodes statistiques pour la classification des données textuelles. Dans *JADT, Nice, 1998* (p. 661). <https://archipel.uqam.ca/6928/>
- Benoît, C. (1846). *Essai historique sur les premiers manuels d'invention oratoire jusqu'à Aristote*. Joubert.
- Benveniste, É. (1966). *Problèmes de linguistique générale*. Gallimard.
- Berger, P. L. et Luckmann, T. (1966). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge* (Penguin Books).
- Berkson, J. (1944). Application of the Logistic Function to Bio-Assay. *Journal of the American Statistical Association*, 39(227), 357-365. <https://doi.org/10.2307/2280041>
- Bertin, M. et Atanassova, I. (2018). InTeReC: In-text Reference Corpus for Applying Natural Language Processing to Bibliometrics (vol. 2080, p. 54-62). Proceedings of the 7th International Workshop on Bibliometric-enhanced Information Retrieval (BIR 2018) co-located with the 40th European Conference on Information Retrieval (ECIR 2018), Grenoble, France, March 26, 2018, CEUR-WS.org. <http://ceur-ws.org/Vol-2080/paper6.pdf>

- Bertin, M., Atanassova, I., Larivière, V. et Gingras, Y. (2015). Mapping the linguistic context of citations. *Bulletin of the Association for Information Science and Technology*, 41(2), 26-29. <https://doi.org/10.1002/bult.2015.1720410208>
- Bertoldi, N., Lareau, F., Pence, C. H. et Malaterre, C. (2023). A quantitative window on the history of statistics: topic-modelling 120 years of Biometrika. *Digital Scholarship in the Humanities*.
- Bhatia, V. K. (2013). *Analysing Genre: Language Use in Professional Settings*. Routledge. <https://doi.org/10.4324/9781315844992>
- Bhatnagar, V., Kanojia, D. et Chebrolu, K. (2022, 14 septembre). *Harnessing Abstractive Summarization for Fact-Checked Claim Detection*. arXiv. <https://doi.org/10.48550/arXiv.2209.04612>
- Bhattacharya, A., Eube, J., Röglin, H. et Schmidt, M. (2019, 2 décembre). *Noisy, Greedy and Not So Greedy k-means++*. arXiv. <https://doi.org/10.48550/arXiv.1912.00653>
- Biber, D. (1989). *A typology of English texts*, 27(1), 3-44. <https://doi.org/10.1515/ling.1989.27.1.3>
- Biber, D., Conrad, S. et Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.
- Biber, D. et Gray, B. (2016). *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511920776>
- Biran, O. et Rambow, O. (2011). Identifying justifications in written dialogs by classifying text as argumentative. *International Journal of Semantic Computing*, 05(04), 363-381. <https://doi.org/10.1142/S1793351X11001328>
- Biro, J. et Siegel, H. (1992). Normativity, argumentation and an epistemic theory of fallacies. Dans F. H. van Eemeren, R. Grootendorst, J. A. Blair et C. A. Willard (dir.), *Argumentation illuminated* (p. 85-103). SICSAT.
- Biskri, I., Achouri, A., Rompré, L., Descoteaux, S. et Bensaber, B. A. (2013). Computer-Assisted Reading: Getting Help from Text Classification and Maximal Association Rules. *Journal of Advances in Information Technology*, 4(4), 157-165. <https://doi.org/10.4304/jait.4.4.157-165>
- Biskri, I. et Meunier, J. G. (2002). SATIM: Système d'Analyse et de Traitement de l'Information Multidimensionnelle. *Proceedings of JADT 2002*, 185-196.
- Blair, J. A. et Johnson, R. H. (1987). The Current State of Informal Logic. *Informal Logic*, 9(2). <https://doi.org/10.22329/il.v9i2.2671>
- Bondarenko, A., Dung, P. M., Kowalski, R. A. et Toni, F. (1997). An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1), 63-101. [https://doi.org/10.1016/S0004-3702\(97\)00015-5](https://doi.org/10.1016/S0004-3702(97)00015-5)
- Bottou, L. (2012). Stochastic Gradient Descent Tricks. Dans G. Montavon, G. B. Orr et K.-R. Müller (dir.), *Neural Networks: Tricks of the Trade: Second Edition* (p. 421-436). Springer. https://doi.org/10.1007/978-3-642-35289-8_25

- Bourdieu, P. (1998). *La Domination masculine*. Seuil. <https://www.seuil.com/ouvrage/la-domination-masculine-pierre-bourdieu/9782020352512>
- Bouziane, A., Bouchiha, D., Doumi, N. et Malki, M. (2015). Question Answering Systems: Survey and Trends. *Procedia Computer Science*, 73, 366-375. <https://doi.org/10.1016/j.procs.2015.12.005>
- Boyd, S. et Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511804441>
- Brachman, R. J. et Levesque, H. J. (1985). *Readings in Knowledge Representation*. Morgan Kaufmann Publishers Inc.
- Brandom, R. (1994). *Making it explicit: reasoning, representing, and discursive commitment*. Harvard University Press.
- Breiman, L. (2001a). Random Forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L. (2001b). Statistical Modeling: The Two Cultures. *Statistical Science*, 16(3), 199-215.
- Breton, F. et Gauthier, G. (2000). *Histoire des théories de l'argumentation*. la Découverte.
- Bross, I. D. et Stermole, D. (1973). Computer-assisted discourse analysis of a jargon. *Computer studies in the humanities and Verbal behavior*, 4(2), 65-76.
- Bruni, E., Tran, N. K. et Baroni, M. (2014). Multimodal Distributional Semantics. *Journal of Artificial Intelligence Research*, 49, 1-47. <https://doi.org/10.1613/jair.4135>
- Buchanan, B. G. et Shortliffe, E. H. (dir.). (1984). *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*. Addison-Wesley.
- Burman, E. et Parker, I. (dir.). (2016). *Discourse Analytic Research: Repertoires and readings of texts in action*. Routledge. <https://doi.org/10.4324/9781315403069>
- Burstein, J. et Marcu, D. (2003). A Machine Learning Approach for Identification of Thesis and Conclusion Statements in Student Essays. *Computers and the Humanities*, 37(4), 455-467.
- Burton-Jones, A. (2014). What have we learned from the Smart Machine? *Information and Organization*, 24(2), 71-105. <https://doi.org/10.1016/j.infoandorg.2014.03.001>
- Busa, R. A. (2004). Frontmatter. Dans S. Schreibman, R. Siemens et J. Unsworth (dir.), *A Companion to Digital Humanities* (p. i-xxviii). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470999875.fmatter>
- Cabrio, E. et Villata, S. (2012a). Combining Textual Entailment and Argumentation Theory for Supporting Online Debates Interactions. Dans *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (p. 208-212). Association for Computational Linguistics. <https://aclanthology.org/P12-2041>

- Cabrio, E. et Villata, S. (2012b). Natural Language Arguments: A Combined Approach. 20th European Conference on Artificial Intelligence (ECAI 2012). <https://hal.inria.fr/hal-00724780>
- Cabrio, E. et Villata, S. (2013). A natural language bipolar argumentation approach to support users in online debate interactions†. *Argument & Computation*, 4(3), 209-230. <https://doi.org/10.1080/19462166.2013.862303>
- Cabrio, E. et Villata, S. (2018). Five Years of Argument Mining: a Data-driven Analysis. Dans *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* (p. 5427-5433). International Joint Conferences on Artificial Intelligence Organization. <https://doi.org/10.24963/ijcai.2018/766>
- Cao, Z., Li, S., Liu, Y., Li, W. et Ji, H. (2015). A Novel Neural Topic Model and Its Supervised Extension. *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1). <https://doi.org/10.1609/aaai.v29i1.9499>
- Chabrol, C. (1973). Grammaires textuelles et structures narratives. Dans C. Chabrol, R. Barthes et S. Alexandrescu (dir.), *Semiotique narrative et textuelle*. Larousse. <http://catalogue.bnf.fr/ark:/12148/cb354386379>
- Chakkarwar, V., Tamane, S. et Thombre, A. (2023). A Review on BERT and Its Implementation in Various NLP Tasks (p. 112-121). International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022), Atlantis Press. https://doi.org/10.2991/978-94-6463-136-4_12
- Chartier, J.-F. et Meunier, J.-G. (2011). Text Mining Methods for Social Representation Analysis in Large Corpora. *Papers on Social Representations*, 20, 37.1-37.47.
- Chartier, J.-F., Pulizzotto, D., Chartrand, L. et Meunier, J.-G. (2019). A data-driven computational semiotics: The semantic vector space of Magritte's artworks. *Semiotica*, 2019. <https://doi.org/10.1515/sem-2018-0120>
- Chartrand, L., Cheung, J. C. K. et Bouguessa, M. (2017, 18 juin). *Detecting Large Concept Extensions for Conceptual Analysis*. arXiv. <https://doi.org/10.48550/arXiv.1706.05723>
- Chartrand, L., Meunier, J.-G., Pulizzotto, D., González, J. L., Chartier, J.-F., Le, N. T., Lareau, F. et Amaya, J. T. (2016). *CoFiH: A heuristic for concept discovery in computer- assisted conceptual analysis*, 11.
- Chartrand, L., Meunier, J.-G., Pulizzotto, D. et Lareau, F. (2018). Un algorithme pour extraire les segments qui expriment un concept – premières expérimentations. *Applied Semiotics/Semiotique appliquee*, (26), 108-125.
- Chaudhari, S. et Shevade, S. (2012). Learning from positive and unlabelled examples using maximum margin clustering. Dans *Proceedings of the 19th international conference on Neural Information Processing - Volume Part III* (p. 465-473). Springer-Verlag. https://doi.org/10.1007/978-3-642-34487-9_56
- Che, L., Yang, X. et Wang, L. (2020). Text feature extraction based on stacked variational autoencoder. *Microprocessors and Microsystems*, 76, 103063. <https://doi.org/10.1016/j.micpro.2020.103063>

- Chen, X., Jia, S. et Xiang, Y. (2020). A review: Knowledge reasoning over knowledge graph. *Expert Systems with Applications*, 141, 112948. <https://doi.org/10.1016/j.eswa.2019.112948>
- Chen, Y. et Liu, L. (2016). Development and research of Topic Detection and Tracking. Dans *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)* (p. 170-173). <https://doi.org/10.1109/ICSESS.2016.7883041>
- Chuang, J., Wilkerson, J. D., Weiss, R., Tingley, D. et Stewart, B. M. (2014). Computer-Assisted Content Analysis: Topic Models for Exploring Multiple Subjective Interpretations. *Advances in Neural Information Processing Systems Workshop on Human-Propelled Machine Learning*. <https://bstewart.scholar.princeton.edu/publications/computer-assisted-content-analysis-topic-models-exploring-multiple-subjective>
- Chuang, S.-L. et Chien, L.-F. (2003). Enriching web taxonomies through subject categorization of query terms from search engine logs. *Decision Support Systems*, 35(1), 113-127. [https://doi.org/10.1016/S0167-9236\(02\)00099-4](https://doi.org/10.1016/S0167-9236(02)00099-4)
- Claesen, M., De Smet, F., Suykens, J. A. K. et De Moor, B. (2015). A Robust Ensemble Approach to Learn From Positive and Unlabeled Data Using SVM Base Models. *Neurocomputing*, 160, 73-84. <https://doi.org/10.1016/j.neucom.2014.10.081>
- Cohen, D. H. (2001). Evaluating Arguments and Making Meta-Arguments. *Informal Logic*, 21(2).
- Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4, 317-370. <https://doi.org/10.1017/S0140525X00009092>
- Cooley, R., Mobasher, B. et Srivastava, J. (1999). Data Preparation for Mining World Wide Web Browsing Patterns. *Knowledge and Information Systems*, 1(1), 5-32. <https://doi.org/10.1007/BF03325089>
- Cope, B., Kalantzis, M., Abd-El-Khalick, F. et Bagley, E. (2013). Science in Writing: Learning Scientific Argument in Principle and Practice. *E-Learning and Digital Media*, 10(4), 420-441. <https://doi.org/10.2304/elea.2013.10.4.420>
- Corston-Oliver, S. H. (1998). Identifying the Linguistic Correlates of Rhetorical Relations. Dans *Discourse Relations and Discourse Markers*. <https://aclanthology.org/W98-0302>
- Cortes, C. et Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/10.1007/BF00994018>
- Cosmides, L. et Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58(1), 1-73. [https://doi.org/10.1016/0010-0277\(95\)00664-8](https://doi.org/10.1016/0010-0277(95)00664-8)
- Cover, T. et Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27. <https://doi.org/10.1109/TIT.1967.1053964>
- Cowling, P. I., Remde, S., Hartley, P., Stewart, W., Stock-Brooks, J. et Woolley, T. (2010). C-Link: Concept Linkage in Knowledge Repositories. *AAAI Spring Symposium: Linked Data Meets Artificial Intelligence*.

- Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S. et Singer, Y. (2006). Online Passive-Aggressive Algorithms. *Journal of Machine Learning Research*, 7, 551-585.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4), 303-314. <https://doi.org/10.1007/BF02551274>
- Danis, J. (2012). *L'analyse conceptuelle de textes assistée par ordinateur (LACTAO) : une expérimentation appliquée au concept d'évolution dans l'oeuvre d'Henri Bergson* [Mémoire de maîtrise en philosophie, UQAM]. <https://oa.mg/work/1488007332>
- Danis, J. et Meunier, J.-G. (2012). CARCAT: Computer-Assisted Reading and Conceptual Analysis of Texts: An experiment applied to the concept of evolution in the work of Henri Bergson. *Digital Studies/Le champ numérique*, 3(1). <https://doi.org/10.16995/dscn.241>
- Danis, J., Meunier, J.-G., Chartier, J.-F., Motasem, A. et Descles, J.-P. (2010). *Classification automatique et stratégie d'annotation appliquées à un concept philosophique: la dimension psychologique du concept de LANGAGE dans l'œuvre de Bergson*.
- Day, R. A. (1989). The origins of the scientific paper: the IMRaD format. *J Am Med Writers Assoc*, 4(2), 16-18.
- Day, R. A. et Gastel, B. (2011). *How to write and publish a scientific paper* (7th ed). Greenwood.
- de Beaugrande, R.-A. et Dressler, W. (1981). *Introduction to Text Linguistics*. Longman.
- de Man, P. (1982). *Allegories of Reading : Figural Language in Rousseau, Nietzsche, Rilke, and Proust* (Yale University Press).
- Denis, F., Gilleron, R. et Letouzey, F. (2005). Learning from positive and unlabeled examples. *Theoretical Computer Science*, 348(1), 70-83. <https://doi.org/10.1016/j.tcs.2005.09.007>
- Derrida, J. (1967a). *De la grammatologie* (Les Editions de Minuit). Editions de Minuit.
- Derrida, J. (1967b). *L'Écriture et la différence*. Éditions du Seuil.
- Derrida, J. (1988). Y a-t-il une langue philosophique? Dans Message Jacques, Roman Joël et Tassin Étienne (dir.), *À quoi pensent les philosophes* (p. 30-36). Editions Autrement.
- de Wit, R., van Dam, F., Zandbelt, L., van Buuren, A., van der Heijden, K., Leenhouts, G. et Loonstra, S. (1997). A pain education program for chronic cancer pain patients: follow-up results from a randomized controlled trial. *Pain*, 73(1), 55-69. [https://doi.org/10.1016/s0304-3959\(97\)00070-5](https://doi.org/10.1016/s0304-3959(97)00070-5)
- Dilthey, W. (1992). *Critique de la raison historique : Introduction aux sciences de l'esprit et autres textes* (S. Mesure, trad.). les Éd. du Cerf. <https://bibliotheques.paris.fr/2021/doc/SYRACUSE/361988/critique-de-la-raison-historique-introduction-aux-sciences-de-l-esprit-et-autres-textes>
- Dor, D. (2003). On newspaper headlines as relevance optimizers. *Journal of Pragmatics*, 35(5), 695-721. [https://doi.org/10.1016/S0378-2166\(02\)00134-0](https://doi.org/10.1016/S0378-2166(02)00134-0)

- Dou, W. et Liu, S. (2016). Topic- and Time-Oriented Visual Text Analysis. *IEEE computer graphics and applications*, 36(4), 8-13. <https://doi.org/10.1109/MCG.2016.73>
- Doyle, L. (1959). Programmed interpretation of text as a basis for information-retrieval systems. Dans *Papers presented at the the March 3-5, 1959, western joint computer conference* (p. 60-63). Association for Computing Machinery. <https://doi.org/10.1145/1457838.1457847>
- Doyle, L. B. (1961a). Information retrieval. *Communications of the ACM*, 4(4), 195. <https://doi.org/10.1145/355578.366528>
- Doyle, L. B. (1961b). Semantic Road Maps for Literature Searchers. *Journal of the ACM*, 8(4), 553-578. <https://doi.org/10.1145/321088.321095>
- Doyle, L. B. (1962). Associated characteristics of words in text. *Communications of the ACM*, 5(4), 223. <https://doi.org/10.1145/366920.366986>
- Doyle, L. B. (1963). The microstatistics of text. *Information Storage and Retrieval*, 1(4), 189-214. [https://doi.org/10.1016/0020-0271\(63\)90017-2](https://doi.org/10.1016/0020-0271(63)90017-2)
- Doyle, L. B. (1965). Is Automatic Classification a Reasonable Application of Statistical Analysis of Text? *Journal of the ACM*, 12(4), 473-489. <https://doi.org/10.1145/321296.321298>
- Duan, X., Liao, M., Zhao, X., Wu, W. et Lv, P. (2019). An Unsupervised Joint Model for Claim Detection. Dans F. Sun, H. Liu et D. Hu (dir.), *Cognitive Systems and Signal Processing* (p. 197-209). Springer. https://doi.org/10.1007/978-981-13-7983-3_18
- Ducrot, O. (1980). *Les Mots du discours* (Minuit). http://www.leseditionsdeminuit.fr/livre-Les_Mots_du_discours-2060-1-1-0-1.html
- Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2), 321-357. [https://doi.org/10.1016/0004-3702\(94\)00041-X](https://doi.org/10.1016/0004-3702(94)00041-X)
- Dunne, P. E. (2008). The Computational Complexity of Ideal Semantics I: Abstract Argumentation Frameworks. Dans *Proceedings of the 2008 conference on Computational Models of Argument: Proceedings of COMMA 2008* (p. 147-158). IOS Press.
- Durst, R. K. (1987). Cognitive and Linguistic Demands of Analytic Writing. *Research in the Teaching of English*, 21(4), 347-76.
- Dusmanu, M., Cabrio, E. et Villata, S. (2017). Argument Mining on Twitter: Arguments, Facts and Sources. Dans *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (p. 2317-2322). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D17-1245>
- Duthie, R., Budzynska, K. et Reed, C. (2016). Mining Ethos in Political Debate. Dans P. Baroni, T. F. Gordon, T. Scheffler et M. Stede (dir.), *Computational Models of Argument* (vol. 287, p. 299-310). IOS Press. <https://doi.org/10.3233/978-1-61499-686-6-299>

- Eckle-Kohler, J., Kluge, R. et Gurevych, I. (2015). On the Role of Discourse Markers for Discriminating Claims and Premises in Argumentative Discourse. Dans *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (p. 2236-2242). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D15-1267>
- Edwards, D. (1997). *Discourse and Cognition*. <https://doi.org/10.4135/9781446221785>
- Eger, S., Daxenberger, J. et Gurevych, I. (2017). Neural End-to-End Learning for Computational Argumentation Mining. Dans *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (p. 11-22). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P17-1002>
- Eger, S., Youssef, P. et Gurevych, I. (2018). Is it Time to Swish? Comparing Deep Learning Activation Functions Across NLP tasks. Dans *EMNLP 2018*. <https://arxiv.org/abs/1901.02671v1>
- Ehninger, D. (1954). The Logic of Argument. Dans D. Potter (dir.), *Argumentation and debate: principles and practices / prepared under the auspices of Tau Kappa Alpha ; edited by David Potter*. (p. 101-124). Dryden Press.
- Eisenberg, A. R. et Garvey, C. (1981). Children's use of verbal strategies in resolving conflicts. *Discourse Processes*, 4(2), 149-170. <https://doi.org/10.1080/01638538109544512>
- Elkan, C. et Noto, K. (2008). Learning classifiers from only positive and unlabeled data. Dans *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (p. 213-220). Association for Computing Machinery. <https://doi.org/10.1145/1401890.1401920>
- Engelbart, D. C. (1962, octobre). *Augmenting Human Intellect: A Conceptual Framework* (AD0289565) [Summary report]. Stanford Research Institute. <http://archive.org/details/1962-engelbart-AHI-framework>
- Etaiwi, W. et Awajan, A. (2023). SemanticGraph2Vec: Semantic graph embedding for text representation. *Array*, 17, 100276. <https://doi.org/10.1016/j.array.2023.100276>
- Falakmasir, M. H., Ashley, K. D., Schunn, C. D. et Litman, D. J. (2014). Identifying Thesis and Conclusion Statements in Student Essays to Scaffold Peer Review. Dans *12th International Conference on Intelligent Tutoring Systems - Volume 8474* (p. 254-259). Springer-Verlag. https://doi.org/10.1007/978-3-319-07221-0_31
- Fawei, B., Wyner, A., Pan, J. Z. et Kollingbaum, M. (2018). Using Legal Ontologies with Rules for Legal Textual Entailment. Dans U. Pagallo, M. Palmirani, P. Casanovas, G. Sartor et S. Villata (dir.), *AI Approaches to the Complexity of Legal Systems* (p. 317-324). Springer International Publishing. https://doi.org/10.1007/978-3-030-00178-0_21
- Fayyad, U., Piatetsky-Shapiro, G. et Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27-34. <https://doi.org/10.1145/240455.240464>
- Feldman, R. (1994). Good arguments. Dans F. F. Schmitt (dir.), *Socializing Epistemology: The Social Dimensions of Knowledge* (p. 159-188). Rowman & Littlefield.

- Feng, V. W. et Hirst, G. (2011). Classifying arguments by scheme. Dans *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (p. 987-996). Association for Computational Linguistics. <https://aclanthology.org/P11-1099>
- Finkelstein, L. (2009). *Pocket book of technical writing for engineering and scientists* (3rd ed., Int. ed). McGraw-Hill.
- Firth, J. R. (1964). *Papers in linguistics, 1934-1951*. London : Oxford University Press. <http://archive.org/details/papersinlinguist0000firt>
- Fishelov, D. (1991). Genre theory and family resemblance—revisited. *Poetics*, 20(2), 123-138. [https://doi.org/10.1016/0304-422X\(91\)90002-7](https://doi.org/10.1016/0304-422X(91)90002-7)
- Fix, E. et Hodges, J. J. L. (1951). *Discriminatory Analysis - Nonparametric Discrimination: Consistency Properties* (ADA800276). California Univ. Berkeley. <https://apps.dtic.mil/sti/citations/ADA800276>
- Føllesdal, D. (1979). Hermeneutics and the hypothetico-deductive method. *Dialectica*, 33(3-4), 319-336. <https://doi.org/10.1111/j.1746-8361.1979.tb00759.x>
- Forest, D. (2006). *Application de techniques de forage de textes de nature prédictive et exploratoire à des fins de gestion et d'analyse thématique de documents textuels non structurés* [Thèse ou essai doctoral accepté, Université du Québec à Montréal]. <https://archipel.uqam.ca/9730/>
- Forest, D. et Meunier, J.-G. (2000). La classification mathématique des textes : un outil d'assistance à la lecture et à l'analyse de textes philosophiques. *Actes des 5es Journées internationales d'Analyse statistique des Données Textuelles*, 1.
- Forest, D. et Meunier, J.-G. (2004). *Classification et catégorisation automatiques : application à l'analyse thématique des données textuelles*, 11.
- Foucault, M. (1969). *L'archéologie du savoir*. Gallimard. <http://www.aspresolver.com/aspresolver.asp?SOTH;S10022158>
- Fowler, A. (1982). *Kinds of Literature: An Introduction to the Theory of Genres and Modes*. Oxford University Press.
- Freeman, J. B. (1991). *Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure*. De Gruyter. <https://doi.org/10.1515/9783110875843>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5). <https://doi.org/10.1214/aos/1013203451>
- Funahashi, K.-I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2(3), 183-192. [https://doi.org/10.1016/0893-6080\(89\)90003-8](https://doi.org/10.1016/0893-6080(89)90003-8)
- Fung, G. P. C., Yu, J. X., Lu, H. et Yu, P. S. (2006). Text Classification without Negative Examples Revisit. *IEEE Transactions on Knowledge and Data Engineering*, 18(1), 6-20. <https://doi.org/10.1109/TKDE.2006.16>

- Gadamer, H. G. (1982). *L'art de comprendre. Ecrits 1, Herméneutique et tradition philosophique* (P. Fruchon et M. Simon, trad.). Aubier-Montaigne.
- Gadamer, H.-G. (2018). *Vérité et Méthode. Les grandes lignes d'une herméneutique philosophique. Édition intégrale*. POINTS.
- Galatanu, O. (2007). Sémantique des “possibles argumentatifs” et axiologisation discursive. Dans D. Bouchard, I. Evrard et E. Vocaj (dir.), *Représentation du sens linguistique II* (p. 313-325). De Boeck.
- Galatanu, O. (2018). *La sémantique des possibles argumentatifs. Génération et (re)construction discursive du sens linguistique* (vol. 45). Peter Lang. <https://www.peterlang.com/document/1113513>
- García Villalba, M. P. et Saint-Dizier, P. (2012a). A Framework to Extract Arguments in Opinion Texts. *International Journal of Cognitive Informatics and Natural Intelligence*, 6(3), 62-87.
- García Villalba, M. P. et Saint-Dizier, P. (2012b). Some Facets of Argument Mining for Opinion Analysis. *Computational Models of Argument*, 23-34. <https://doi.org/10.3233/978-1-61499-111-3-23>
- Genette, G. (1982). *Palimpsestes. La littérature au second degré* (Seuil). <https://www.seuil.com/ouvrage/palimpsestes-la-litterature-au-second-degre-gerard-genette/9782020061162>
- Geurts, P., Ernst, D. et Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3-42. <https://doi.org/10.1007/s10994-006-6226-1>
- Ghadessy, M. (1999). Textual Features and Contextual Factors for Register Identification. Dans M. Ghadessy (dir.), *Text and context in functional linguistics*. J. Benjamins.
- Ghasemi, A., Bahadoran, Z., Mirmiran, P., Hosseinpanah, F., Shiva, N. et Zadeh-Vakili, A. (2019). The Principles of Biomedical Scientific Writing: Discussion. *International Journal of Endocrinology and Metabolism*, 17(3), e95415. <https://doi.org/10.5812/ijem.95415>
- Gigerenzer, G. (2007). *Gut feelings: the intelligence of the unconscious*. Viking. <http://catdir.loc.gov/catdir/enhancements/fy0745/2006052810-b.html>
- Gini, C. (1921). Measurement of Inequality of Incomes. *The Economic Journal*, 31(121), 124-126. <https://doi.org/10.2307/2223319>
- Glock, H.-J. (2008). *What is Analytic Philosophy?* Cambridge University Press. <https://doi.org/10.1017/CBO9780511841125>
- Goldman, A. I. (1999). *Knowledge in a Social World*. Clarendon Press.
- Grabmair, M., Ashley, K. D., Chen, R., Sureshkumar, P., Wang, C., Nyberg, E. et Walker, V. R. (2015). Introducing LUIIMA: an experiment in legal conceptual retrieval of vaccine injury decisions using a UIMA type system and tools. Dans *Proceedings of the 15th International Conference on Artificial Intelligence and Law* (p. 69-78). Association for Computing Machinery. <https://doi.org/10.1145/2746090.2746096>

- Grasso, F., Rahwan, I., Reed, C. et Simari, G. R. (2010). Introducing Argument & Computation. *Argument & Computation*, 1(1), 1-5. <https://doi.org/10.1080/19462161003734804>
- Graves, H., Moghaddasi, S. et Hashim, A. (2013). Mathematics is the method: Exploring the macro-organizational structure of research articles in mathematics. *Discourse Studies*, 15(4), 421-438. <https://doi.org/10.1177/1461445613482430>
- Green, C. (1969, 1^{er} mars). *Application of Theorem Proving to Problem Solving*: Defense Technical Information Center. <https://doi.org/10.21236/ADA459656>
- Grennan, W. (1997). *Informal Logic: Issues and Techniques*. McGill-Queen's University Press.
- Grimes, J. E. (1975). The Thread of Discourse. Dans *The Thread of Discourse*. De Gruyter Mouton. <https://doi.org/10.1515/9783110886474>
- Grishman, R. (2015). Information Extraction. *IEEE Intelligent Systems*, 30(5), 8-15. <https://doi.org/10.1109/MIS.2015.68>
- Groncin, J. (1993). *L'Universalité de l'herméneutique*. Presses Universitaires de France - PUF.
- Groza, A. et Szabo, R. (2015). Enacting textual entailment and ontologies for automated essay grading in chemical domain. Dans *arXiv.org*. <https://arxiv.org/abs/1511.02669v1>
- Guespin, L. (1971). 1. Problématique des travaux sur le discours politique. *Langages*, 6(23), 3-24. <https://doi.org/10.3406/lgge.1971.2048>
- Gustavii, B. (2000). *How to Write and Illustrate a Scientific Paper* (Cambridge University Press). <https://www.cambridge.org/core/books/how-to-write-and-illustrate-a-scientific-paper/2703C27BA4D7E6244E6B3DDE4EC2FF86>
- Habermas, J. (1981). *Théorie de l'agir communicationnel. Tome 1. Rationalité de l'agir et rationalisation de la société* (J.-M. Ferry, trad.). Fayard. <https://www.fayard.fr/sciences-humaines/theorie-de-lagir-communicationnel-tome-1-9782213018935>
- Habermas, J. (1987). *Théorie de l'agir communicationnel. Tome 2. Pour une critique de la raison fonctionnaliste* (J. L. Schlegel, trad.). Fayard.
- Habernal, I. et Gurevych, I. (2017). Argumentation Mining in User-Generated Web Discourse. *Computational Linguistics*, 43(1), 125-179. https://doi.org/10.1162/COLI_a_00276
- Habernal, I., Wachsmuth, H., Gurevych, I. et Stein, B. (2018). SemEval-2018 Task 12: The Argument Reasoning Comprehension Task. Dans *Proceedings of the 12th International Workshop on Semantic Evaluation* (p. 763-772). Association for Computational Linguistics. <https://doi.org/10.18653/v1/S18-1121>
- Halliday, M. A. K. et Hasan, R. (1976). *Cohesion in English*. Longman.
- Halliday, M. A. K. et Matthiessen, C. M. I. (2004). *An Introduction to Functional Grammar* (3^e éd.). Routledge. <https://doi.org/10.4324/9780203783771>

- Hamblin, C. L. (1970). *Fallacies* (1st edition). Methuen.
- Harris, Z. S. (1951). *Methods in structural linguistics* (p. xv, 384). University of Chicago Press.
- Hastie, T., Tibshirani, R. et Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
<https://doi.org/10.1007/978-0-387-84858-7>
- Hastings, A. C. (1962). *A reformulation of the modes of reasoning in argumentation* [Northwestern University].
- Hatim, B. (1991). The pragmatics of argumentation in Arabic: The rise and fall of a text type. *Text - Interdisciplinary Journal for the Study of Discourse*, 11(2), 189-200.
<https://doi.org/10.1515/text.1.1991.11.2.189>
- Hayes, P. J. (1979). The logic of frames. Dans P. J. Hayes et D. Metzger (dir.), *Frame Conceptions and Text Understanding* (p. 46-61). De Gruyter. <https://doi.org/10.1515/9783110858778>
- He, F., Liu, T., Webb, G. I. et Tao, D. (2020, 2 mars). *Instance-Dependent PU Learning by Bayesian Optimal Relabeling*. arXiv. Récupéré le 23 octobre 2022 de <http://arxiv.org/abs/1808.02180>
- Henderson, V. P., Clemow, L., Massion, A. O., Hurley, T. G., Druker, S. et Hébert, J. R. (2012). The effects of mindfulness-based stress reduction on psychosocial outcomes and quality of life in early-stage breast cancer patients: a randomized trial. *Breast Cancer Research and Treatment*, 131(1), 99-109. <https://doi.org/10.1007/s10549-011-1738-1>
- Hinton, G. E., Ghahramani, Z. et Teh, Y. W. (1999). Learning to Parse Images. Dans *Advances in Neural Information Processing Systems* (vol. 12). MIT Press.
https://papers.nips.cc/paper_files/paper/1999/hash/5a142a55461d5fef016acfb927fee0bd-Abstract.html
- Hinton, G. et Salakhutdinov, R. (2011). Discovering Binary Codes for Documents by Learning Deep Generative Models. *Topics in Cognitive Science*, 3(1), 74-91. <https://doi.org/10.1111/j.1756-8765.2010.01109.x>
- Hockey, S. (2001). *Electronic Texts in the Humanities: Principles and Practice*. Oxford University Press.
- Hoerl, A. E. et Kennard, R. W. (1970a). Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics*, 12(1), 69-82. <https://doi.org/10.2307/1267352>
- Hoerl, A. E. et Kennard, R. W. (1970b). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55-67. <https://doi.org/10.2307/1267351>
- Hornik, K., Stinchcombe, M. et White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359-366.
- Houngbo, H. et Mercer, R. (2014). An automated method to build a corpus of rhetorically-classified sentences in biomedical texts. Dans *Proceedings of the First Workshop on Argumentation Mining* (p. 19-23). Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-2103>

- Hua, X. et Wang, L. (2017). Understanding and Detecting Supporting Arguments of Diverse Types. Dans *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (p. 203-208). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/P17-2032>
- Ibeke, E., Lin, C., Wyner, A. et Barawi, M. H. (2017). Extracting and Understanding Contrastive Opinion through Topic Relevant Sentences. Dans *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)* (p. 395-400). Asian Federation of Natural Language Processing. <https://aclanthology.org/I17-2067>
- Ienco, D., Pensa, R. G. et Meo, R. (2012). From Context to Distance: Learning Dissimilarity for Categorical Data Clustering. *ACM Transactions on Knowledge Discovery from Data*, 6(1), 1:1-1:25.
<https://doi.org/10.1145/2133360.2133361>
- Ignatow, G. et Mihalcea, R. (2016). *Text Mining: A Guidebook for the Social Sciences*. SAGE Publications.
- Jacob, E. K. (2004). *Classification and Categorization: A Difference that Makes a Difference*.
<https://hdl.handle.net/2142/1686>
- Javitt, J. C. et Schiffman, R. M. (2000). Clinical success and quality of life with brimonidine 0.2% or timolol 0.5% used twice daily in glaucoma or ocular hypertension: a randomized clinical trial. Brimonidine Outcomes Study Group I. *Journal of Glaucoma*, 9(3), 224-234.
<https://doi.org/10.1097/00061198-200006000-00005>
- Jenny, L. (1976). La stratégie de la forme. *Poétique*, (27), 257-81.
- Jia, X. et Wang, L. (2022). Attention enhanced capsule network for text classification by encoding syntactic dependency trees with graph convolutional neural network. *PeerJ Computer Science*, 8, e831. <https://doi.org/10.7717/peerj-cs.831>
- Jiang, S., Cormier, S., Angarita, R. et Rousseaux, F. (2023). Improving text mining in plant health domain with GAN and/or pre-trained language model. *Frontiers in Artificial Intelligence*, 6.
<https://www.frontiersin.org/articles/10.3389/frai.2023.1072329>
- Johnson, R. H. (2000). *Manifest Rationality: A Pragmatic Theory of Argument*. Routledge.
<https://doi.org/10.4324/9781410606174>
- Kaiser, K., Lareau, F., Malaterre, C. et Pulizzotto, D. (à paraître). L'analyse de corpus : démontrer par le texte. Dans J. Prud'homme et M. Kao (dir.), *Faire preuve. Comment nos sociétés distinguent le vrai du faux*. (PUM).
- Katzav, J. et Reed, C. A. (2004). On Argumentation Schemes and the Natural Classification of Arguments. *Argumentation*, 18(2), 239-259. <https://doi.org/10.1023/B:ARGU.0000024044.34360.82>
- Kearney, M. H. (2017). The Discussion Section Tells Us Where We Are. *Research in Nursing & Health*, 40(4), 289-291. <https://doi.org/10.1002/nur.21803>
- Kienpointner, M. (1992). How to Classify Arguments. Dans F. H. van Eemeren, R. Grootendorst, J. A. Blair et C. A. Willard (dir.), *Argumentation illuminated* (p. 178-188). SICSAT.

- Kiesel, J., Al-Khatib, K., Hagen, M. et Stein, B. (2015). A Shared Task on Argumentation Mining in Newspaper Editorials. Dans *Proceedings of the 2nd Workshop on Argumentation Mining* (p. 35-38). Association for Computational Linguistics. <https://doi.org/10.3115/v1/W15-0505>
- Kintsch, W. et van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394. <https://doi.org/10.1037/0033-295X.85.5.363>
- Kirschner, P. A., Buckingham-Shum, S. J. et Carr, C. S. (dir.). (2002). *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making* (Softcover reprint of the original 1st ed. 2003 edition). Springer.
- Kitcher, P. (1991). Persuasion. Dans M. Pera et W. R. Shea (dir.), *Persuading Science: The Art of Scientific Rhetoric* (p. 3-27). Science History Publications.
- Klobucar, A. et O'Neill, M. (2021). Reading and Collaboration: Developing Digital Reading Practices With Computer-Assisted Text Analysis Tools. Dans H. Rahman (dir.), *Advances in Human and Social Aspects of Technology* (p. 177-202). IGI Global. <https://doi.org/10.4018/978-1-7998-5849-2.ch008>
- Kneale, W. et Kneale, M. (1962). *The Development of Logic*. Oxford, England: Clarendon Press.
- Knorr, M. et Hitzler, P. (2014). Description Logics. Dans J. H. Siekmann (dir.), *Handbook of the History of Logic* (vol. 9, p. 659-678). North-Holland. <https://doi.org/10.1016/B978-0-444-51624-4.50015-0>
- Knott, A. et Dale, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18(1), 35-62. <https://doi.org/10.1080/01638539409544883>
- Kretchmer, N. (1972). Lactose and lactase. *Scientific American*, 227(4), 71-78.
- Kristeva, J. (1969). *Séméiôtikè : Recherches pour une sémanalyse* (Seuil).
- Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, 77(3), 319-337. <https://doi.org/10.1002/sce.3730770306>
- Kuroda, S.-Y. (1972). The Categorical and the Thetic Judgment: Evidence from Japanese Syntax. *Foundations of Language*, 9(2), 153-185.
- Kusupati, A., Bhatt, G., Rege, A., Wallingford, M., Sinha, A., Ramanujan, V., Howard-Snyder, W., Chen, K., Kakade, S., Jain, P. et Farhadi, A. (2024, 7 février). *Matryoshka Representation Learning*. arXiv. <https://doi.org/10.48550/arXiv.2205.13147>
- Labov, W. (1972). *Language in the Inner City: Studies in the Black English Vernacular*. University of Pennsylvania Press Inc.
- Labov, W. et Waletzky, J. (1967). Narrative analysis. Dans J. Helm (dir.), *Essays on the verbal and visual arts: proceedings of the 1966 annual spring meeting of the American Ethnological Society* (p. 12-44). Distributed by the University of Washington Press.
- Lacharité, N. (1989). *Introduction à la méthodologie de la pensée écrite*. PUQ. <https://www.puq.ca/catalogue/livres/introduction-methodologie-pensee-ecrite-438.html>

- Lambrech, K. (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511620607>
- Lamirel, J.-C., Lareau, F. et Malaterre, C. (2023a). *La méthode de modélisation thématique CFMf basée sur le clustering neuronal avec maximisation des traits : Comparaison avec LDA sur des études scientifiques*. Congrès SFC'2023, Strasbourg, 6-7 juillet 2023.
- Lamirel, J.-C., Lareau, F. et Malaterre, C. (2023b). *The CFMf Topic-Modeling Method Based on Neural Clustering with Feature Maximization: Comparison with LDA*. ISSI 2023 Conference, Bloomington, 2-5 juillet 2023.
- Lamirel, J.-C., Lareau, F. et Malaterre, C. (2024). CFMf Topic-Model: Comparison with LDA and Top2Vec. *Scientometrics*. <https://doi.org/10.1007/s11192-024-05017-z>
- Lareau, F. (2016). *Analyse philosophique en contexte numérique du concept de dualité chez Jonathan St B. T. Evans* [Mémoire accepté, Université du Québec à Montréal].
<https://archipel.uqam.ca/9125/>
- Lareau, F. (2022). Approche computationnelle de l'analyse conceptuelle : présentation opérationnelle et approfondissement méthodologique de la détection de concept dans des extraits textuels. *Philosophiques*, 49(2).
- Lareau, F. (2023a). *Fouille d'argument (argument mining) : La détection de thèses principales*. Colloque du CRIHN 2023, 11 au 13 octobre 2023, Université de Montréal, Montréal.
- Lareau, F. (2023b). *La détection de thèses principales en fouille d'argument (argument mining)*. Travaux en cours, Chaire de recherche du Canada en Épistémologie pratique, 25 octobre 2023, Université de Sherbrooke, Sherbrooke.
- Lareau, F., Chartrand, L., Chartier, J.-F. et Pulizzotto, D. (2018). Lecture et analyse conceptuelle de texte assistées par ordinateur (LACTAO) appliquées à des textes de haut niveau théorique : illustration de la méthode et résultats de l'analyse du concept de dualité chez Jonathan St B. T. Evans. *Applied Semiotics/Semiotique appliquée*, (26), 171-192.
- Lareau, F. et Malaterre, C. (2023-a). *Inferring social networks from unstructured text data: Code and datasets* (<https://doi.org/10.5281/zenodo.7967417>).
- Lareau, F. et Malaterre, C. (2023-b). *On the use of machine translation and topic-modeling to analyze non-parallel multilingual corpora: A case study in the history of philosophy of science* (<https://doi.org/10.5281/zenodo.6484582>).
- Lareau, F., Malaterre, C. et Simard, J.-C. (2023a). *50 ans de philosophie au Québec : analyse thématique de la revue Philosophiques (1974-2023)*. Atelier sur la philosophie québécoise et la recherche assistée par ordinateur, Département de philosophie de l'UQAM, 3 novembre 2023.
- Lareau, F., Malaterre, C. et Simard, J.-C. (2023b). *Perspectives philosophiques québécoises: analyse thématique computationnelle de Philosophiques (1974-2018)*. ACFAS, Colloque « Approches numériques en histoire de la culture au Québec », Montréal, 9 mai 2023.

- Lareau, F., Malaterre, C. et Simard, J.-C. (2024). *50 ans de philosophie québécoise: une analyse thématique de Philosophiques de 1974 à 2023*. Table ronde au Congrès 2024 de la SPQ - 50e anniversaire de la revue *Philosophiques* : histoire et perspectives, Cégep de Saint-Jean-sur-Richelieu, du lundi 3 au 6 juin 2024.
- Lareau, F., Malaterre, C. et Simard, J.-C. (s. d.). Cinquante ans de *Philosophiques*: une analyse thématique computationnelle. *Philosophiques*, 51(1).
- Lawrence, J. et Reed, C. (2016). Argument Mining Using Argumentation Scheme Structures: The 6th International Conference on Computational Models of Argument. Dans P. Baroni, T. F. Gordon, T. Scheffler et M. Stede (dir.), *Computational Models of Argument* (vol. 287, p. 379-390). IOS Press. <https://doi.org/10.3233/978-1-61499-686-6-379>
- Le, N. T., Meunier, J.-G., Chartrand, L., Pulizzotto, D., Lopez, J. A., Lareau, F. et Chartier, J.-F. (2016). Nouvelle méthode d'analyse syntactico-sémantique profonde dans la lecture et l'analyse de textes assistées par ordinateur (LATAO). *JADT 2016: 13ème Journées internationales d'Analyse statistique des Données Textuelles*, 2.
- Lebart, L. et Salem, A. (1994). *Statistique textuelle*. Dunod.
- Lee, C.-P. et Lin, C.-J. (2013). A Study on L2-Loss (Squared Hinge-Loss) Multiclass SVM. *Neural Computation*, 25(5), 1302-1323. https://doi.org/10.1162/NECO_a_00434
- Lee, W. S. et Liu, B. (2003). Learning with Positive and Unlabeled Examples Using Weighted Logistic Regression. Dans *Proceedings of the Twentieth International Conference on Machine Learning* (vol. 20, p. 448-455).
- Lemke, J. L. (1990). *Talking Science: Language, Learning, and Values*. Ablex Publishing Corporation, 355 Chestnut Street, Norwood, NJ 07648 (hardback: ISBN-0-89391-565-3; paperback: ISBN-0-89391-566-1). <https://eric.ed.gov/?id=ED362379>
- Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Rivista di Linguistica*, 20(1), 1-31.
- Levy, R., Bilu, Y., Hershovich, D., Aharoni, E. et Slonim, N. (2014). Context Dependent Claim Detection. Dans *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (p. 1489-1500). Dublin City University and Association for Computational Linguistics. <https://aclanthology.org/C14-1141>
- Levy, R., Gretz, S., Sznajder, B., Hummel, S., Aharonov, R. et Slonim, N. (2017). Unsupervised corpus-wide claim detection. Dans *Proceedings of the 4th Workshop on Argument Mining* (p. 79-84). Association for Computational Linguistics. <https://doi.org/10.18653/v1/W17-5110>
- Lewenstein, B. (2007). Why should we care about science books? *JCOM : Journal of Science Communication*, 6(1). <https://doi.org/10.22323/2.06010303>
- Lewis, D. M. (2000). *Some emergent discourse connectives in English* [<http://purl.org/dc/dcmitype/Text>, University of Oxford]. <https://ora.ox.ac.uk/objects/uuid:137436a0-a0e7-4764-a667-6312d899f909>

- Li, Q. et Zhou, W. (2020). Connecting the Dots Between Fact Verification and Fake News Detection. Dans *Proceedings of the 28th International Conference on Computational Linguistics* (p. 1820-1825). International Committee on Computational Linguistics. <https://doi.org/10.18653/v1/2020.coling-main.165>
- Li, R. V., Baoxin. (2017). *Convolutional Neural Networks in Visual Computing: A Concise Guide*. CRC Press. <https://doi.org/10.4324/9781315154282>
- Li, X. et Liu, B. (2003). Learning to classify texts using positive and unlabeled data. Dans *Proceedings of the 18th international joint conference on Artificial intelligence* (p. 587-592). Morgan Kaufmann Publishers Inc.
- Li, X., Liu, B. et Ng, S.-K. (2007). Learning to Identify Unexpected Instances in the Test Set. Dans *Proceedings of the 20th International Joint Conference on Artificial Intelligence* (vol. 7, p. 2802-2807).
- Li, X.-L. et Liu, B. (2005). Learning from Positive and Unlabeled Examples with Different Data Distributions. Dans J. Gama, R. Camacho, P. B. Brazdil, A. M. Jorge et L. Torgo (dir.), *Machine Learning: ECML 2005* (p. 218-229). Springer. https://doi.org/10.1007/11564096_24
- Ligibel, J. A., Meyerhardt, J., Pierce, J. P., Najita, J., Shockro, L., Campbell, N., Newman, V. A., Barbier, L., Hacker, E., Wood, M., Marshall, J., Paskett, E. et Shapiro, C. (2012). Impact of a Telephone-Based Physical Activity Intervention upon Exercise Behaviors and Fitness in Cancer Survivors Enrolled in a Cooperative Group Setting. *Breast Cancer Research and Treatment*, 132(1), 205-213. <https://doi.org/10.1007/s10549-011-1882-7>
- Lippi, M. et Torroni, P. (2015). Context-Independent Claim Detection for Argument Mining. Dans Q. Yang et M. J. Wooldridge (dir.), *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015* (p. 185-191). AAAI Press. <http://ijcai.org/Abstract/15/033>
- Lippi, M. et Torroni, P. (2016a). Argument Mining from Speech: Detecting Claims in Political Debates. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1). <https://doi.org/10.1609/aaai.v30i1.10384>
- Lippi, M. et Torroni, P. (2016b). Argumentation Mining: State of the Art and Emerging Trends. *ACM Transactions on Internet Technology*, 16(2), 10:1-10:25. <https://doi.org/10.1145/2850417>
- Lippi, M. et Torroni, P. (2016c). MARGOT: A web server for argumentation mining. *Expert Systems with Applications*, 65, 292-303. <https://doi.org/10.1016/j.eswa.2016.08.050>
- Liu, B. (2011). *Web Data Mining*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-19460-3>
- Liu, B., Dai, Y., Li, X., Lee, W. S. et Yu, P. S. (2003). Building text classifiers using positive and unlabeled examples. Dans *Third IEEE International Conference on Data Mining* (p. 179-186). IEEE Comput. Soc. <https://doi.org/10.1109/ICDM.2003.1250918>
- Liu, B., Lee, W. S., Yu, P. S. et Li, X. (2002). Partially Supervised Classification of Text Documents. Dans *Proceedings of the Nineteenth International Conference on Machine Learning* (vol. 2, p. 387-394).

- Liu, L. et Peng, T. (2014). Clustering-based Method for Positive and Unlabeled Text Categorization Enhanced by Improved TFIDF. *Journal of Information Science and Engineering*, 30, 1463-1481.
- Liu, S., Wang, X., Collins, C., Dou, W., Ouyang, F., El-Assady, M., Jiang, L. et Keim, D. A. (2019). Bridging Text Visualization and Mining: A Task-Driven Survey. *IEEE transactions on visualization and computer graphics*, 25(7), 2482-2504. <https://doi.org/10.1109/TVCG.2018.2834341>
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129-137. <https://doi.org/10.1109/TIT.1982.1056489>
- Loiseau, S. (2005). Thématique et sémantique contextuelle d'un concept philosophique. Dans G. Williams (dir.), *La linguistique de corpus* (p. 129-140). Presses Universitaires de Rennes - PUR.
- Lores, R. (2004). On RA Abstracts: From Rhetorical Structure to Thematic Organisation. *English for Specific Purposes*, 23(3), 280-302. <https://doi.org/10.1016/j.esp.2003.06.001>
- Lounsbury, M. et Carberry, E. J. (2005). From King to Court Jester? Weber's Fall from Grace in Organizational Theory. *Organization Studies*, 26(4), 501-525. <https://doi.org/10.1177/0170840605051486>
- Lu, Z., Du, P. et Nie, J.-Y. (2020, 12 avril). *VGCN-BERT: Augmenting BERT with Graph Embedding for Text Classification*. arXiv. <https://doi.org/10.48550/arXiv.2004.05707>
- Lucas, K. J. (2022). *The Structure of Philosophical Discourse* [thesis, Purdue University Graduate School]. <https://doi.org/10.25394/PGS.19597411.v1>
- Luhn, H. P. (1958). A Business Intelligence System. *IBM Journal of Research and Development*, 2(4), 314-319. <https://doi.org/10.1147/rd.24.0314>
- Lumer, C. (1988). The Disputation. A Special Type of Cooperative Argumentative Dialogue. *Argumentation*, 2(4), 441-464. <https://doi.org/10.1007/BF00128986>
- Lumer, C. (1991). Structure and function of argumentations An epistemological approach to determining criteria for the validity and adequacy of argumentations. Dans F. H. van Eemeren, R. Grootendorst, J. A. Blair et C. A. Willard (dir.), *Proceedings of the Second International Conference on Argumentation, organized by the International Society for the Study of Argumentation (ISSA) at the University of Amsterdam, June 19-22, 1990* (p. 98-107). SICSAT.
- Lumer, C. (2005). The Epistemological Theory of Argument--How and Why? *Informal Logic*, 25(3). <https://doi.org/10.22329/il.v25i3.1135>
- Lumer, C. (2010). Pragma-Dialectics and the Function of Argumentation. *Argumentation*, 24(1), 41-69. <https://doi.org/10.1007/s10503-008-9118-7>
- Luo, N., Yuan, F. et Zuo, W. (2009). An Integration of CoTraining and Affinity Propagation for PU Text Classification. Dans *2009 International Conference on Computer Engineering and Technology* (vol. 1, p. 150-154). <https://doi.org/10.1109/ICCET.2009.131>

- Macagno, F., Walton, D. et Reed, C. (2017). Argumentation Schemes.: History, Classifications, and Computational Applications. *IfCoLog Journal of Logics and their Applications*, 4(8), 2493-2556.
- Macciò, A., Madeddu, C., Gramignano, G., Mulas, C., Floris, C., Sanna, E., Cau, M. C., Panzone, F. et Mantovani, G. (2012). A randomized phase III clinical trial of a combined treatment for cachexia in patients with gynecological cancers: evaluating the impact on metabolic and inflammatory profiles and quality of life. *Gynecologic Oncology*, 124(3), 417-425.
<https://doi.org/10.1016/j.ygyno.2011.12.435>
- Malaterre, C., Chartier, J.-F. et Lareau, F. (2020a). *80 années de Philosophy of Science à la loupe du topic-modeling: une approche computationnelle de l'histoire de la philosophie des sciences*. École ATO 2020, 11 septembre 2020.
- Malaterre, C., Chartier, J.-F. et Lareau, F. (2020b). The recipes of Philosophy of Science: Characterizing the semantic structure of corpora by means of topic associative rules. *PLOS ONE*, 15(11), e0242353. <https://doi.org/10.1371/journal.pone.0242353>
- Malaterre, C. et Lareau, F. (2022). The early days of contemporary philosophy of science: novel insights from machine translation and topic-modeling of non-parallel multilingual corpora. *Synthese*, 200(3), 242. <https://doi.org/10.1007/s11229-022-03722-x>
- Malaterre, C. et Lareau, F. (2023a). Identifying Hidden Communities of Interest with Topic-based Networks: A Case Study of the Community of Philosophers of Science (1930-2017). Dans *Proceedings of the 56th Hawaii International Conference on System Sciences*.
<https://hdl.handle.net/10125/102936>
- Malaterre, C. et Lareau, F. (2023b). *Mapping the Contours of the Emerging Discipline of Astrobiology with Text-Mining Approaches*. Big Data & HPS 2023, Toronto, 18-19 mai 2023.
- Malaterre, C. et Lareau, F. (2023c). *Mining eight decades of philosophy and philosophers of science*. Colloque DR2 conference, Rome, 16-17 janvier 2023.
- Malaterre, C. et Lareau, F. (2023d). The Emergence of Astrobiology: A Topic-Modeling Perspective. *Astrobiology*, 23(5), 496-512. <https://doi.org/10.1089/ast.2022.0122>
- Malaterre, C. et Lareau, F. (2023e). *Visualizing Hidden Communities of Interest: A Preliminary Analysis of Topic-based Social Networks in Astrobiology*. ISSI 2023 Conference, Bloomington, 2-5 juillet 2023.
- Malaterre, C. et Lareau, F. (2024a). Inferring social networks from unstructured text data: A proof of concept detection of “hidden communities of interest”. *Data & Policy*.
<https://doi.org/10.1017/dap.2023.48>
- Malaterre, C. et Lareau, F. (2024b). Visualizing hidden communities of interest: A case-study analysis of topic-based social networks in astrobiology. *Scientometrics*. <https://doi.org/10.1007/s11192-024-05047-7>
- Malaterre, C., Lareau, F., Pulizzotto, D. et St-Onge, J. (2020c). *Huit revues sur huit décennies : une analyse computationnelle de la philosophie des sciences contemporaine*. TEC – Travaux en cours (UQAM), 2 décembre 2020.

- Malaterre, C., Lareau, F., Pulizzotto, D. et St-Onge, J. (2021). Eight journals over eight decades: a computational topic-modeling approach to contemporary philosophy of science. *Synthese*, 199(1), 2883-2923. <https://doi.org/10.1007/s11229-020-02915-6>
- Malaterre, C., Pulizzotto, D. et Lareau, F. (2019). Revisiting three decades of Biology and Philosophy: a computational topic-modeling perspective. *Biology & Philosophy*, 35(1), 5. <https://doi.org/10.1007/s10539-019-9729-4>
- Manevitz, L. M. et Yousef, M. (2001). One-Class SVMs for Document Classification. *Journal of Machine Learning Research*, 2, 139-154.
- Mann, W. C. et Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text - Interdisciplinary Journal for the Study of Discourse*, 8(3), 243-281. <https://doi.org/10.1515/text.1.1988.8.3.243>
- Manning, C. et Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. The MIT Press.
- Mantovani, G., Proto, E., Massa, E., Mulas, C., Madeddu, C., Mura, L., Mudu, M. C., Astara, G., Murgia, V., Gramignano, G., Ferreli, L., Camboni, P., Lusso, M. R., Mocci, M., Tore, G., Mura, M., Amichetti, M. et Maccio, A. (2002). Induction chemotherapy followed by concomitant chemoradiation therapy in advanced head and neck cancer: a phase II study for organ-sparing purposes evaluating feasibility, effectiveness and toxicity. *International Journal of Oncology*, 20(2), 419-427.
- Marandin, J.-M. (1988). A propos de la notion de thème de discours. Éléments d'analyse dans le récit. *Langue française*, 78(1), 67-87. <https://doi.org/10.3406/lfr.1988.4744>
- March, W. F. et Ochsner, K. I. (2000). The long-term safety and efficacy of brinzolamide 1.0% (azopt) in patients with primary open-angle glaucoma or ocular hypertension. The Brinzolamide Long-Term Therapy Study Group. *American Journal of Ophthalmology*, 129(2), 136-143. [https://doi.org/10.1016/s0002-9394\(99\)00343-8](https://doi.org/10.1016/s0002-9394(99)00343-8)
- Marchand, P. (1998). *L'Analyse du discours assistée par ordinateur : concepts, méthodes, outils*. Armand Colin. https://www.mediatheques.strasbourg.eu/Conservatoire/doc/IGUANA_2/609658/l-analyse-du-discours-assistee-par-ordinateur-concepts-methodes-outils-par-pascal-marchand
- Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. W.H. Freeman. <http://digitool.hbz-nrw.de:1801/webclient/DeliveryManager?pid=2948165>
- Mayer, T. (2020). *Argument Mining on Clinical Trials* [phdthesis, Université Côte d'Azur]. <https://theses.hal.science/tel-03209489>
- Mayer, T., Cabrio, E., Lippi, M., Torroni, P. et Villata, S. (2018). Argument Mining on Clinical Trials. Dans *Computational Models of Argument* (p. 137-148). IOS Press. <https://doi.org/10.3233/978-1-61499-906-5-137>

- Mayer, T., Cabrio, E. et Villata, S. (2020). Transformer-based Argument Mining for Healthcare Applications. Dans *ECAI 2020 - 24th European Conference on Artificial Intelligence*. <https://hal.science/hal-02879293>
- McNally, J. R. (1967). Rudolph Agricola's de inventione dialectica libri tres: A translation of selected chapters. *Communications Monographs*, world. <https://doi.org/10.1080/03637756709375551>
- Medawar, P. (1964, 1^{er} août). Is the scientific paper a fraud? *Saturday Review*, 42-43.
- Mellet, S. et Longrée, D. (2009). *New approaches in text linguistics*. John Benjamins. <http://books.google.com/books?id=QQhSAQAIAAJ>
- Menini, S., Cabrio, E., Tonelli, S. et Villata, S. (2018). Never Retreat, Never Retract: Argumentation Analysis for Political Speeches. Dans *AAAI 2018 - 32nd AAAI Conference on Artificial Intelligence* (p. 4889-4896). <https://hal.archives-ouvertes.fr/hal-01876442>
- Mercier, H. et Sperber, D. (2011a). Argumentation: Its adaptiveness and efficacy. *Behavioral and Brain Sciences*, 34(2), 94-111. <https://doi.org/10.1017/S0140525X10003031>
- Mercier, H. et Sperber, D. (2011b). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57-74. <https://doi.org/10.1017/S0140525X10000968>
- Mergenthaler, E. (1996). Computer-Assisted Content Analysis. Dans C. Züll, J. Harkness et J. H. P. Hoffmeyer-Zlotnik (dir.), *Text analysis and computers* (vol. 1, p. 3-32). Zentrum für Umfragen, Methoden und Analysen -ZUMA-.
- Meunier, J. G. (1976). *Système d'analyse des textes par ordinateur (SATO): rapport technique*. Université du Québec à Montréal, Département de philosophie.
- Meunier, J. G. (1997). La lecture et l'analyse de texte assistées par ordinateur. *Sciences Cognitives*, (22), 211-223.
- Meunier, J.-G. (2009). CARAT—Computer-Assisted Reading and Analysis of Texts: The Appropriation of a Technology. *Digital Studies / Le champ numérique*, 1(3). <https://doi.org/10.16995/dscn.263>
- Meunier, J.-G. (2012). Cognitive Representation : Computable Or Non Computable? Dans E. Aggazi (dir.), *Representation in Science*. International Academy of Philosophy of Science. <https://archipel.uqam.ca/6944/>
- Meunier, J.-G. (2017). Humanités numériques et modélisation scientifique. *Questions de communication*, (31), 19-48. <https://doi.org/10.4000/questionsdecommunication.11040>
- Meunier, J.-G. (2019). Le paradoxe des humanités numériques. *Quaderni. Communication, technologies, pouvoir*, (98), 19-31. <https://doi.org/10.4000/quaderni.1407>
- Meunier, J.-G. (2022). Le numérique, la lecture et l'analyse de textes philosophiques. *Philosophiques*, 49(2), 369-393. <https://doi.org/10.7202/1097458ar>

- Meunier, J.-G., Biskri, I. et Forest, D. (2005). Classification and Categorization in Computer Assisted Reading and Analysis of Texts. Dans C. Lefebvre et H. Cohen (dir.), *Handbook of Categorization in Cognitive Science* (p. 955-978). Elsevier. <https://archipel.uqam.ca/6936/>
- Meunier, J.-G., Biskri, I., Nault, G. et Nyongwa, M. (1997). Exploration de classifieurs connexionnistes pour l'analyse de textes assistée par ordinateur. Dans *Colloque RIAO97* (p. 661-664). <https://archipel.uqam.ca/6922/>
- Meunier, J.-G. et Forest, D. (2008). Computer assisted conceptual analysis. (CACAT): the concept of Mind in The Collected Papers of C.S. Peirce. (p. 74-80). Digital Humanities conference. <https://archipel.uqam.ca/6940/>
- Meunier, J.-G. et Nault, G. (1997). Approche connexionniste au problème de l'extraction de connaissances terminologiques à partir de textes (p. 62-76). Les Techniques d'intelligence artificielle appliquées aux Technologies de l'Information, Les Cahiers scientifiques ACFAS. <https://archipel.uqam.ca/6921/>
- Meunier, J.-G., Remaki, L. et Forest, D. (1999). Use of classifiers in computer-assisted reading and analysis of text (CARAT). Dans *Actes du colloque international CISST 1999*. <https://srb.sdl.edu.sa/esploro/outputs/conferenceProceeding/Use-of-classifiers-in-computer-assisted-reading/9913244808331>
- Meunier, J.-G., Rolland, S. et Daoust, F. (1976). A System for Text and Content Analysis. *Computers and the Humanities*, 10(5), 281-286.
- Meunier, J.-G., Ste-Marie, M. et Chartier, J.-F. (2011). Did Darwin really talk about Evolution? E. Agazzi et F. Minazzi (dir.), (p. 127-146). Mimesis. <https://archipel.uqam.ca/6943/>
- Mika, P., Tudorache, T., Bernstein, A., Welty, C., Knoblock, C., Vrandečić, D., Groth, P., Noy, N., Janowicz, K. et Goble, C. (dir.). (2014). *The Semantic Web – ISWC 2014: 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part II* (vol. 8797). Springer International Publishing. <https://doi.org/10.1007/978-3-319-11915-1>
- Mikolov, T., Chen, K., Corrado, G. et Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. Dans *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*. arXiv. <http://arxiv.org/abs/1301.3781>
- Miller, C. (2005). *Habermas et l'éthique de la discussion : une réponse au relativisme postmoderne* [Mémoire accepté, Université de Montréal]. <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/16504>
- Misak, C. (2018). The Pragmatist Theory of Truth. Dans M. Glanzberg (dir.), *The Oxford Handbook of Truth* (p. 0). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199557929.013.11>
- Mizrahi, M. et Dickinson, M. (2020). Argumentation in Philosophical Practice: An Empirical Study. *OSSA Conference Archive*. <https://scholar.uwindsor.ca/ossaarchive/OSSA12/Saturday/3>

- Mochales Palau, R. et Moens, M.-F. (2009). Argumentation mining: the detection, classification and structure of arguments in text. Dans *Proceedings of the 12th International Conference on Artificial Intelligence and Law* (p. 98-107). Association for Computing Machinery. <https://doi.org/10.1145/1568234.1568246>
- Mochales Palau, R. et Moens, M.-F. (2011). Argumentation Mining. *Artificial Intelligence and Law*, 19(1), 1-22. <https://doi.org/10.1007/s10506-010-9104-x>
- Moens, M.-F., Boiy, E., Palau, R. M. et Reed, C. (2007). Automatic detection of arguments in legal texts. Dans *Proceedings of the 11th international conference on Artificial intelligence and law - ICAIL '07* (p. 225). ACM Press. <https://doi.org/10.1145/1276318.1276362>
- Molder, H. te et Potter, J. (dir.). (2005). *Conversation and Cognition*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511489990>
- Moore, R. C. (1985). Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25(1), 75-94. [https://doi.org/10.1016/0004-3702\(85\)90042-6](https://doi.org/10.1016/0004-3702(85)90042-6)
- Mordelet, F. et Vert, J.-P. (2011). ProDiGe: Prioritization Of Disease Genes with multitask machine learning from positive and unlabeled examples. *BMC Bioinformatics*, 12(1), 389. <https://doi.org/10.1186/1471-2105-12-389>
- Moretti, F. (2000). Conjectures on World Literature. *New Left Review*, (1), 54-68.
- Moretti, F. et Piazza, A. (2007). *Graphs, Maps, Trees: Abstract Models for Literary History* (p. 119). Verso Books.
- Morgan, J. N. et Sonquist, J. A. (1963). Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American Statistical Association*, 58(302), 415-434. <https://doi.org/10.1080/01621459.1963.10500855>
- Mounier, P. (2018). IBM ou International Busa Machines ? De l'informatique aux humanités. Dans *Les humanités numériques : Une histoire critique* (p. 21-43). Éditions de la Maison des sciences de l'homme. <https://doi.org/10.4000/books.editionsmsmsh.12033>
- Muennighoff, N., Tazi, N., Magne, L. et Reimers, N. (2023, 5 février). *MTEB: Massive Text Embedding Benchmark*. arXiv. <https://doi.org/10.48550/arXiv.2210.07316>
- Naderi, N. et Hirst, G. (2016). Argumentation Mining in Parliamentary Discourse. Dans M. Baldoni, C. Baroglio, F. Bex, F. Grasso, N. Green, M.-R. Namazi-Rad, M. Numao et M. T. Suarez (dir.), *Principles and Practice of Multi-Agent Systems* (vol. 9935, p. 16-25). Springer International Publishing. https://doi.org/10.1007/978-3-319-46218-9_2
- Nanba, H., Kando, N. et Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. *Advances in Classification Research Online*, 117-134. <https://doi.org/10.7152/acro.v11i1.12774>
- Nanthini, K., Sivabalaselvamani, D., Chitra, K., Gokul, P., KavinKumar, S. et Kishore, S. (2023). A Survey on Data Augmentation Techniques. Dans *2023 7th International Conference on Computing*

- Methodologies and Communication (ICCMC)* (p. 913-920).
<https://doi.org/10.1109/ICCMC56507.2023.10084010>
- Neelakantan, A., Xu, T., Puri, R., Radford, A., Han, J. M., Tworek, J., Yuan, Q., Tezak, N., Kim, J. W., Hallacy, C., Heidecke, J., Shyam, P., Power, B., Nekoul, T. E., Sastry, G., Krueger, G., Schnurr, D., Such, F. P., Hsu, K., ... Weng, L. (2022, 24 janvier). *Text and Code Embeddings by Contrastive Pre-Training*. arXiv. <https://doi.org/10.48550/arXiv.2201.10005>
- Newman, S. E. et Marshall, C. C. (1991). *Pushing Toulmin Too Far: Learning From an Argument Representation Scheme* (SSL 92–45) [Technical Report]. Xerox Palo Alto Research Center.
- Nguyen, H. et Litman, D. (2018). Argument Mining for Improving the Automated Scoring of Persuasive Essays. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).
<https://doi.org/10.1609/aaai.v32i1.12046>
- Nguyen, M.-Q., Nguyen, P. H. P., Nguyen, T.-H., Meunier, J.-G. et O'Shaughnessy, D. (2007). Reinforcement Learning With Simulated User For Automatic Dialog Strategy Optimization. Dans *Fifth International Conference on Industrial Automation, Ecole de technologie superieure, Montreal, Canada*. <https://archipel.uqam.ca/2437/>
- Nicholas, N. (1994). *Problems in the Application of Rhetorical Structure Theory to Text Generation* [University of Melbourne].
- Niculae, V., Park, J. et Cardie, C. (2017). Argument Mining with Structured SVMs and RNNs. Dans *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (p. 985-995). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/P17-1091>
- Nigam, K., McCallum, A. K., Thrun, S. et Mitchell, T. (2000). Text Classification from Labeled and Unlabeled Documents using EM. *Machine Learning*, 39(2), 103-134.
<https://doi.org/10.1023/A:1007692713085>
- Nundy, S., Kakar, A. et Bhutta, Z. A. (2022a). How to Write an Abstract? Dans S. Nundy, A. Kakar et Z. A. Bhutta (dir.), *How to Practice Academic Medicine and Publish from Developing Countries? A Practical Guide* (p. 179-184). Springer Nature. https://doi.org/10.1007/978-981-16-5248-6_15
- Nundy, S., Kakar, A. et Bhutta, Z. A. (2022b). How to Write the Introduction to a Scientific Paper? Dans S. Nundy, A. Kakar et Z. A. Bhutta (dir.), *How to Practice Academic Medicine and Publish from Developing Countries? A Practical Guide* (p. 193-199). Springer Nature.
https://doi.org/10.1007/978-981-16-5248-6_17
- Oaksford, M. et Chater, N. (2007). *Bayesian Rationality: The probabilistic approach to human reasoning*. Oxford University Press.
- Obrist, A. J. (1983). *The Microcomputer and the Primary School*.
- OpenAI. (2022). *Introducing text and code embeddings*. <https://openai.com/index/introducing-text-and-code-embeddings/>

- Osoba, D., Tannock, I. F., Ernst, D. S. et Neville, A. J. (1999). Health-related quality of life in men with metastatic prostate cancer treated with prednisone alone or mitoxantrone and prednisone. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 17(6), 1654-1663. <https://doi.org/10.1200/JCO.1999.17.6.1654>
- Panaccio, C. (2019). *Récit et reconstruction: Fondements de la méthode en histoire de la philosophie*. Vrin.
- Park, J. et Cardie, C. (2014). Identifying Appropriate Support for Propositions in Online User Comments. Dans *Proceedings of the First Workshop on Argumentation Mining* (p. 29-38). Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-2105>
- Parret, H. (1987). Argumentation and narrativity. Dans F. Eemeren, R. Grootendorst et A. Blair (dir.), *Across the Lines of Disciplines* (p. 165-175). De Gruyter Mouton. <https://doi.org/10.1515/9783110867718>
- Pêcheux, M. (1969). *Analyse automatique du discours*. Dunod. <http://catalogue.bnf.fr/ark:/12148/cb35371732t>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A. et Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825-2830.
- Peirce, C. S. (1909). *Existential Graphs. MS 514*. Existential Graphs. <http://www.jfsowa.com/peirce/ms514.htm>
- Peldszus, A. (2014). Towards segment-based recognition of argumentation structure in short texts. Dans *Proceedings of the First Workshop on Argumentation Mining* (p. 88-97). Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-2112>
- Peldszus, A. et Stede, M. (2013a). From Argument Diagrams to Argumentation Mining in Texts: A Survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 7(1), 1-31.
- Peldszus, A. et Stede, M. (2013b). Ranking the annotators: An agreement study on argumentation structure. Dans *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse* (p. 196-204). Association for Computational Linguistics. <https://aclanthology.org/W13-2324>
- Peldszus, A. et Stede, M. (2015). Joint prediction in MST-style discourse parsing for argumentation mining. Dans *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (p. 938-948). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D15-1110>
- Pelletier, G. (2021). Palatium Reginae Eloquentiae [1641]. Exercitatio II. Lectio XV. Locus a comparatione maiorum, minorum, parium. *Exercices de rhétorique*, (16). <https://doi.org/10.4000/rhetorique.1095>

- Peng, T., Zuo, W. et He, F. (2008). SVM based adaptive learning method for text classification from positive and unlabeled documents. *Knowl. Inf. Syst.*, 16, 281-301. <https://doi.org/10.1007/s10115-007-0107-1>
- Pera, M. et Shea, W. R. (1991). *Persuading Science: The Art of Scientific Rhetoric*. Science History Publications.
- Perelman, C. et Olbrechts-Tyteca, L. (1958). *La nouvelle rhétorique. Traité de l'argumentation*. Presses Universitaires de France - PUF. 2 vol.
- Perier-Camby, J., Bertin, M., Atanassova, I. et Armetta, F. (2019). A preliminary study to compare deep learning with rule-based approaches for citation classification. Dans *8th International Workshop on Bibliometric-enhanced Information Retrieval (BIR) co-located with the 41st European Conference on Information Retrieval (ECIR 2019)* (vol. 2345, p. 125-131). <https://hal.science/hal-02098831>
- Pho, P. D. (2008). Research article abstracts in applied linguistics and educational technology: a study of linguistic realizations of rhetorical structure and authorial stance. *Discourse Studies*, 10(2), 231-250. <https://doi.org/10.1177/1461445607087010>
- Platt, J. C. (1998). Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. *Advances in Kernel Methods-Support Vector Learning*, 208.
- Poincaré, H. (1902). *La science et l'hypothèse*. Ernest Flammarion.
- Pontille, D. (2007). Matérialité des écrits scientifiques et travail de frontières: le cas du format IMRAD. Dans P. Hert et M. Paul-Cavallier (dir.), *Sciences et frontières : délimitations du savoir, objets et passages* (E.M.E., p. 229-253).
- Popping, R. (2000). *Computer-Assisted Text Analysis* (University of Groningen). <https://uk.sagepub.com/en-gb/eur/computer-assisted-text-analysis/book205979>
- Portolés, J. (1998). *Marcadores del discurso*. Ariel.
- Potter, A. (2019). The Rhetorical Structure of Attribution. Dans *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019* (p. 38-49). Association for Computational Linguistics. <https://doi.org/10.18653/v1/W19-2706>
- Potter, J. et Wetherell, M. (1987). *Discourse and social psychology: beyond attitudes and behaviour*. Sage Publications. <http://catdir.loc.gov/catdir/enhancements/fy0656/87060198-t.html>
- Prasad, R., Webber, B., Lee, A. et Joshi, A. (2019, 15 mars). *Penn Discourse Treebank Version 3.0*. Linguistic Data Consortium. <https://doi.org/10.35111/QEBF-GK47>
- Prince, M. (2004). Does Active Learning Work? A Review of the Research. *Journal of Engineering Education*, 93(3), 223-231. <https://doi.org/10.1002/j.2168-9830.2004.tb00809.x>

- Pulizzotto, D., Chartier, J.-F., Lareau, F., Meunier, J.-G. et Chartrand, L. (2018). Conceptual Analysis in a computer-assisted framework: mind in Peirce. *Umanistica Digitale*, (2).
<https://doi.org/10.6092/issn.2532-8816/7305>
- Pulizzotto, D., Lopez Gonzalez, J. A., Chartier, J.-F., Meunier, J.-G., Chartrand, L., Lareau, F. et Ngoc, T. L. (2016). Recherche de « périsegments » dans un contexte d'analyse conceptuelle assistée par ordinateur : le concept d'« esprit » chez Peirce (Search of "perisegments" in computer-assisted conceptual analysis : the concept of "mind" in Peirce). Dans *Actes de la conférence conjointe JEP-TALN-RECITAL 2016. volume 2 : TALN (Posters)* (p. 522-530). AFCP - ATALA.
<https://aclanthology.org/2016.jeptalnrecital-poster.27>
- Pustejovsky, J. et Stubbs, A. (2012). *Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications*. O'Reilly Media, Inc.
- Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge: MIT Press.
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J. et Manning, C. D. (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. Dans *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (p. 101-108). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-demos.14>
- Ragni, M., Kola, I. et Johnson-Laird, P. N. (2017). The Wason Selection Task: A Meta-Analysis. Dans R. Granger, U. Hahn et R. Sutton (dir.), *CogSci 2017 Proceedings of the 39th Annual Meeting of the Cognitive Science Society. London, UK, 26-29 July 2017*. Computational Foundations of Cognition.
<https://mindmodeling.org/cogsci2017/>
- Rahwan, I., Zablith, F. et Reed, C. (2007). Laying the foundations for a world wide argument web. *Artificial Intelligence*, 171(10-15), 897-921. <https://doi.org/10.1016/j.artint.2007.04.015>
- Raju, M. K., Subrahmanian, S. T. et Sivakumar, T. (2017). A Comparative Survey on Different Text Categorization. *International Journal of Computer Science and Engineering Communications*, 5(3), 1612-1618.
- Ram, A. (1991). A Theory of Questions and Question Asking. *The Journal of the Learning Sciences*, 1(3/4), 273-318.
- Ranzato, M. A. et Szummer, M. (2008). Semi-supervised learning of compact document representations with deep networks. Dans *Proceedings of the 25th international conference on Machine learning - ICML '08* (p. 792-799). ACM Press. <https://doi.org/10.1145/1390156.1390256>
- Rastier, F. (1972). *Idéologie et théorie des signes. Analyse structurale des Éléments d'idéologie d'Antoine-Louis-Claude Destutt de Tracy*. De Gruyter Mouton.
- Rastier, F. (2005). Discours et texte. *Texto !* http://www.revue-texto.net/Reperes/Themes/Rastier_Discours.html

- Rastier, F. (2018). Computer-Assisted Interpretation of Semiotic Corpora. Dans D. Compagno (dir.), *Quantitative Semiotic Analysis* (p. 123-139). Springer International Publishing. https://doi.org/10.1007/978-3-319-61593-6_6
- Reboul, O. (1991). *Introduction à la rhétorique: Théorie et pratique*. Presses Universitaires France - PUF.
- Redeker, G. (1991). Linguistic Markers of Discourse Structure. *Linguistics*, 29, 1139-72.
- Reed, C. (2006). Preliminary Results from an Argument Corpus. Dans E. M. B. and L. R. Miyares et D. G. teaches drama and postcolonial theory at Y. S. John (dir.), *Linguistics in the Twenty First Century* (p. 185-196). Cambridge Scholars Publishing.
- Řehůřek, R. et Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. Dans M. Valletta (dir.), *Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks* (p. 46-50). University of Malta. <https://is.muni.cz/publication/884893/en/Software-Framework-for-Topic-Modelling-with-Large-Corpora/Rehurek-Sojka>
- Reimers, N. et Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Dans *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. <http://arxiv.org/abs/1908.10084>
- Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 13(1-2), 81-132. [https://doi.org/10.1016/0004-3702\(80\)90014-4](https://doi.org/10.1016/0004-3702(80)90014-4)
- Rezaeenour, J., Ahmadi, M., Jelodar, H. et Shahrooei, R. (2023). Systematic review of content analysis algorithms based on deep neural networks. *Multimedia Tools and Applications*, 82(12), 17879-17903. <https://doi.org/10.1007/s11042-022-14043-z>
- Rialle, V., Meunier, J.-G., Biskri, I., Oussedik, S. et Nault, G. (1998). Application de l'algorithmique génétique à l'analyse terminologique. Dans *Acte du colloque international JADT 98*.
- Ribeiro, H. J. (2009). Perelman and Toulmin as philosophers : on the inalienable connection between philosophy, rhetoric and argumentation. Dans H. J. Ribeiro, *Rhetoric and argumentation in the beginning of the XXIst century: proceedings of the XXIst century* (p. 33-51). Imprensa da Universidade de Coimbra. https://doi.org/10.14195/978-989-26-0498-5_2
- Ricœur, P. (1969). *Le conflit des interprétations, essais d'herméneutique*. Seuil.
- Ricœur, P. (1986). *Du texte à l'action. Essais d'herméneutique II*. Seuil.
- Rigotti, E. et Morasso, S. G. (2010). Comparing the Argumentum Model of Topics to Other Contemporary Approaches to Argument Schemes: The Procedural and Material Components. *Argumentation*, 24(4), 489-512. <https://doi.org/10.1007/s10503-010-9190-7>
- Rinott, R., Dankin, L., Alzate Perez, C., Khapra, M. M., Aharoni, E. et Slonim, N. (2015). Show Me Your Evidence - an Automatic Method for Context Dependent Evidence Detection. Dans *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (p. 440-450). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D15-1050>

- Rizkallah, É. (2013). L'analyse textuelle des discours assistée par ordinateur et les logiciels textométriques : réflexions critiques et prospectives à partir d'une modélisation des procédés analytiques fondamentaux. *Cahiers de recherche sociologique*, (54), 141-160. <https://doi.org/10.7202/1025996ar>
- Robbins, H. et Monro, S. (1951). A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3), 400-407. <https://doi.org/10.1214/aoms/1177729586>
- Robin, L. (1923). *La pensée grecque et les origines de l'esprit scientifique*. La renaissance du livre. <https://www.bibliotheque.nat.tn/KHNU/doc/SYRACUSE/226859/la-pensee-grecque-et-les-origines-de-l-esprit-scientifique>
- Rocchio, J. J. (1971). Relevance feedback in information retrieval. Dans G. Salton (dir.), *The Smart retrieval system - experiments in automatic document processing* (p. 313-323). Prentice-Hall. <https://www.bibsonomy.org/bibtex/c18d843e34fe4f8bd1d2438227857225>
- Rompré, L., Biskri, I. et Meunier, J.-G. (2017). Using Association Rules Mining for Retrieving Genre-specific Music Files. Dans *Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference, FLAIRS-30, Marco-Island (Florida)*.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386-408. <https://doi.org/10.1037/h0042519>
- Roux-Faucard, G. (2006). Intertextualité et traduction. *Meta : journal des traducteurs / Meta: Translators' Journal*, 51(1), 98-118. <https://doi.org/10.7202/012996ar>
- Rumelhart, D. E., Hinton, G. et Williams, R. J. (1985, 1^{er} septembre). *Learning Internal Representations by Error Propagation* (ADA164453) [Technical Report]. University of San Diego. <https://apps.dtic.mil/sti/citations/ADA164453>
- Runeson, P., Alexandersson, M. et Nyholm, O. (2007). Detection of Duplicate Defect Reports Using Natural Language Processing. Dans *29th International Conference on Software Engineering (ICSE'07)* (p. 499-510). <https://doi.org/10.1109/ICSE.2007.32>
- Saint-Dizier, P. (2012). Processing natural language arguments with the <TextCoop> platform. *Argument & Computation*, 3(1), 49-82. <https://doi.org/10.1080/19462166.2012.663539>
- Sainte-Marie, M. B., Meunier, J.-G., Payette, N. et Chartier, J.-F. (2011). The concept of evolution in the Origin of Species: a computer-assisted analysis. *Literary and linguistic computing*, 26(3), 329-334.
- Sainte-Marie, M., Meunier, J.-G., Payette, N. et Chartier, J.-F. (2010). *Reading Darwin between the lines: a computer-assisted analysis of the concept of evolution in the Origin of species*.
- Salton, G. (1971). *The SMART retrieval system experiments in automatic document processing*. Prentice-Hall. <http://books.google.com/books?id=7-M8AAAIAAJ>
- Salton, G. et Lesk, M. E. (1965). The SMART automatic document retrieval systems—an illustration. *Communications of the ACM*, 8(6), 391-398. <https://doi.org/10.1145/364955.364990>

- Salton, G., Wong, A. et Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613-620. <https://doi.org/10.1145/361219.361220>
- Samraj, B. (2005). An exploration of a genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes*, 24(2), 141-156. <https://doi.org/10.1016/j.esp.2002.10.001>
- Schiffrin, D. (1987). *Discourse Markers*. Cambridge University Press.
- Schmidt-Schauß, M. et Smolka, G. (1991). Attributive concept descriptions with complements. *Artificial Intelligence*, 48(1), 1-26. [https://doi.org/10.1016/0004-3702\(91\)90078-X](https://doi.org/10.1016/0004-3702(91)90078-X)
- Schneider, J., Davis, B. et Wyner, A. (2012). Dimensions of argumentation in social media. Dans *Proceedings of the 18th international conference on Knowledge Engineering and Knowledge Management* (p. 21-25). Springer-Verlag. https://doi.org/10.1007/978-3-642-33876-2_4
- Schneider, J., Samp, K., Passant, A. et Decker, S. (2013). Arguments about deletion: how experience improves the acceptability of arguments in ad-hoc online task groups. Dans *Proceedings of the 2013 conference on Computer supported cooperative work* (p. 1069-1080). Association for Computing Machinery. <https://doi.org/10.1145/2441776.2441897>
- Schourup, L. C. (1985). *Common Discourse Particles in English Conversation*. Routledge. <https://doi.org/10.4324/9781315401584>
- Scriven, M. (1976). *Reasoning*. McGraw-Hill.
- Searle, J. R. (1999). The future of philosophy. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 354(1392), 2069-2080. <https://doi.org/10.1098/rstb.1999.0544>
- Seffah, A. et Meunier, J.-G. (1997). ALADIN, Un atelier orienté objet pour l'analyse et la lecture de Textes assistée par ordinateur. *S. Bolasco. L. Lebart, A. Salem Analisi Statistica dei Dati in Testuali, vol II, CISU*, 2, 289-292.
- Shilo, G. et Ragonis, N. (2023). Students' abilities to formulate exhaustive titles as a key to comprehending arguments. *International Journal of Applied Linguistics*, 00, 1-18. <https://doi.org/10.1111/ijal.12454>
- Siegel, H. et Biro, J. (1997). Epistemic Normativity, Argumentation, and Fallacies. *Argumentation*, 11(3), 277-292. <https://doi.org/10.1023/A:1007799325361>
- Sinclair, S. (2003). Computer-assisted reading: Reconceiving text Analysis : Reconceiving text analysis. *Literary and linguistic computing*, 18(2), 175-184.
- Skelton, J. R. et Edwards, S. J. L. (2000). The function of the discussion section in academic medical writing. *BMJ : British Medical Journal*, 320(7244), 1269-1270.
- Skoufaki, S. (2020). Rhetorical Structure Theory and coherence break identification. *Text and Talk*, 40, 99-124. <https://doi.org/10.1515/text-2019-2050>

- Small, H. (1982). Citation context analysis. Dans B. Dervin et J. Voigt (dir.), *Progress in Communication Sciences* (vol. 3, p. 287-310). Ablex Publishing Corporation.
<https://books.google.ca/books?id=4kViAAAAMAAJ>
- Smith, E. L. (1985). Text type and discourse framework. *Text - Interdisciplinary Journal for the Study of Discourse*, 5(3), 229-247. <https://doi.org/10.1515/text.1.1985.5.3.229>
- Solanki, S. K. et Patel, J. T. (2015). A Survey on Association Rule Mining. Dans *2015 Fifth International Conference on Advanced Computing & Communication Technologies* (p. 212-216).
<https://doi.org/10.1109/ACCT.2015.69>
- Sollaci, L. B. et Pereira, M. G. (2004). The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. *Journal of the Medical Library Association*, 92(3), 364-371.
- Song, K., Tan, X., Qin, T., Lu, J. et Liu, T.-Y. (2020, 2 novembre). *MPNet: Masked and Permuted Pre-training for Language Understanding*. arXiv. <https://doi.org/10.48550/arXiv.2004.09297>
- Sperber, D. et Mercier, H. (2012). Reasoning as a Social Competence. Dans H. Landemore et J. Elster (dir.), *Collective Wisdom: Principles and Mechanisms* (p. 368-392). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511846427.016>
- Stab, C. et Gurevych, I. (2014a). Annotating Argument Components and Relations in Persuasive Essays. Dans *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (p. 1501-1510). Dublin City University and Association for Computational Linguistics. <https://aclanthology.org/C14-1142>
- Stab, C. et Gurevych, I. (2014b). Identifying Argumentative Discourse Structures in Persuasive Essays. Dans *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (p. 46-56). Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1006>
- Stab, C. et Gurevych, I. (2017). Parsing Argumentation Structures in Persuasive Essays. *Computational Linguistics*, 43(3), 619-659. https://doi.org/10.1162/COLI_a_00295
- Stab, C., Miller, T., Schiller, B., Rai, P. et Gurevych, I. (2018). Cross-topic Argument Mining from Heterogeneous Sources. Dans *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (p. 3664-3674). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/D18-1402>
- Stede, M., Taboada, M. et Das, D. (2017). *Annotation Guidelines for Rhetorical Structure*. University of Potsdam and Simon Fraser University.
- Stein, N. et Bernas, R. (1999). The early emergence of argumentative knowledge and skill. Dans J. Andriessen et P. Coirier (dir.), *Foundations of argumentative text processing* (p. 97-116). Amsterdam University Press.
- Stein, N. L. et Albro, E. R. (2001). The Origins and Nature of Arguments: Studies in Conflict Understanding, Emotion, and Negotiation. *Discourse Processes*, 32(2-3), 113-133.
<https://doi.org/10.1080/0163853X.2001.9651594>

- Stein, N. L. et Miller, C. A. (1993). The Development of Memory and Reasoning Skill in Argumentative Contexts: Evaluating, Explaining, and Generating Evidence. Dans *Advances in Instructional Psychology*. Routledge.
- Stevens, H. (2018). Hans Peter Luhn and the birth of the hashing algorithm. *IEEE Spectrum*, 55(2), 44-49. <https://doi.org/10.1109/MSPEC.2018.8278136>
- Stevenson, C. L. (1944). *Ethics and language*. Yale University Press ; H. Milford, Oxford University Press.
- Sun, C., Lo, D., Khoo, S.-C. et Jiang, J. (2011). Towards more accurate retrieval of duplicate bug reports. Dans *2011 26th IEEE/ACM International Conference on Automated Software Engineering (ASE 2011)* (p. 253-262). <https://doi.org/10.1109/ASE.2011.6100061>
- Suppe, F. (1993). Credentialing Scientific Claims. *Perspectives on Science*, 1(2), 153-203. https://doi.org/10.1162/posc_a_00434
- Suppe, F. (1998). The Structure of a Scientific Paper. *Philosophy of Science*, 65(3), 381-405.
- Svedlund, J., Sullivan, M., Liedman, B., Lundell, L. et Sjödin, I. (1997). Quality of life after gastrectomy for gastric carcinoma: controlled study of reconstructive procedures. *World Journal of Surgery*, 21(4), 422-433. <https://doi.org/10.1007/pl00012265>
- Swales, J. M. (1981). *Aspects of Article Introductions*. Language Studies Unit, University of Aston in Birmingham. <https://press.umich.edu/Books/A/Aspects-of-Article-Introductions-Michigan-Classics-Ed2>
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press. <http://catdir.loc.gov/catdir/description/cam024/90002464.html>
- Swales, J. M. et Feak, C. B. (2004). *Academic writing for graduate students: essential tasks and skills* (2nd ed). University of Michigan Press.
- Szabo, R. et Groza, A. (2017). Analysing debates on climate change with textual entailment and ontologies. Dans *2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)* (p. 39-46). <https://doi.org/10.1109/ICCP.2017.8116981>
- Taboada, M. et Mann, W. C. (2006). Applications of Rhetorical Structure Theory. *Discourse Studies*, 8(4), 567-588. <https://doi.org/10.1177/1461445606064836>
- Takiguchi, S., Yamamoto, K., Hirao, M., Imamura, H., Fujita, J., Yano, M., Kobayashi, K., Kimura, Y., Kurokawa, Y., Mori, M., Doki, Y. et Osaka University Clinical Research Group for Gastroenterological Study. (2012). A comparison of postoperative quality of life and dysfunction after Billroth I and Roux-en-Y reconstruction following distal gastrectomy for gastric cancer: results from a multi-institutional RCT. *Gastric Cancer: Official Journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association*, 15(2), 198-205. <https://doi.org/10.1007/s10120-011-0098-1>
- Taylor, C. (1992). *Multiculturalism and « The Politics of Recognition »*. Princeton University Press.

- Teruel, M., Cardellino, C., Cardellino, F., Alemany, L. A. et Villata, S. (2018). Legal text processing within the MIREL project. Dans G. Rehm, V. Rodríguez-Doncel et J. Moreno-Schneider (dir.), *Proceedings of the LREC 2018 "Workshop on Language Resources and Technologies for the Legal Knowledge Graph"* (p. 42-49).
- Teufel, S., Carletta, J. et Moens, M. (1999). An annotation scheme for discourse-level argumentation in research articles. Dans *Ninth Conference of the European Chapter of the Association for Computational Linguistics* (p. 110-117). Association for Computational Linguistics. <https://aclanthology.org/E99-1015>
- Teufel, S., Siddharthan, A. et Batchelor, C. (2009). Towards Domain-Independent Argumentative Zoning: Evidence from Chemistry and Computational Linguistics. Dans *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (p. 1493-1502). Association for Computational Linguistics. <https://aclanthology.org/D09-1155>
- Thagard, P. (1988). *Computational philosophy of science*. MIT Press.
- Thelwall, M. (2019). The rhetorical structure of science? A multidisciplinary analysis of article headings. *Journal of Informetrics*, 13(2), 555-563. <https://doi.org/10.1016/j.joi.2019.03.002>
- Thomas, S. N. (1973). *Practical Reasoning in Natural Language*. Prentice-Hall.
- Thompson, G. (2004). *Introducing Functional Grammar* (2^e éd.). Routledge. <https://doi.org/10.4324/9780203785270>
- Thung, F., Lo, D. et Lawall, J. (2013). Automated library recommendation. Dans *2013 20th Working Conference on Reverse Engineering (WCRE)* (p. 182-191). <https://doi.org/10.1109/WCRE.2013.6671293>
- Thyer, B. A. (2008). The Importance of Journal Articles. Dans B. A. Thyer (dir.), *Preparing Research Articles* (p. 0). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195323375.003.0001>
- Tindale, C. (2004). *Rhetorical Argumentation: Principles of Theory and Practice*. Sage Publications. <https://doi.org/10.4135/9781452204482>
- Toulmin, S. E. (1958). *The Uses of Argument*. Cambridge University Press.
- Trautmann, D., Fromm, M., Tresp, V., Seidl, T. et Schütze, H. (2020). Relational and Fine-Grained Argument Mining. *Datenbank-Spektrum*, 20(2), 99-105. <https://doi.org/10.1007/s13222-020-00341-z>
- Trent, A. et Cho, J. (2014). Interpretation strategies. Dans P. Leavy (dir.), *The Oxford handbook of qualitative research* (p. 639-657). Oxford University Press.
- Trenta, A., Hunter, A. et Riedel, S. (2015, 17 septembre). *Extraction of evidence tables from abstracts of randomized clinical trials using a maximum entropy classifier and global constraints*. arXiv. <https://doi.org/10.48550/arXiv.1509.05209>

- Tseronis, A. (2011). From Connectives to Argumentative Markers: A Quest for Markers of Argumentative Moves and of Related Aspects of Argumentative Discourse. *Argumentation*, 25, 427-447. <https://doi.org/10.1007/s10503-011-9215-x>
- Turney, P. et Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, 37. <https://doi.org/10.1613/jair.2934>
- Tyrrell, C. J., Kaisary, A. V., Iversen, P., Anderson, J. B., Baert, L., Tammela, T., Chamberlain, M., Webster, A. et Blackledge, G. (1998). A randomised comparison of « Casodex » (bicalutamide) 150 mg monotherapy versus castration in the treatment of metastatic and locally advanced prostate cancer. *European Urology*, 33(5), 447-456. <https://doi.org/10.1159/000019634>
- Underwood, T. (2017). A Genealogy of Distant Reading. *Digital Humanities Quarterly*, 011(2).
- Valette, M. (2003). Conceptualisation and Evolution of concepts. The example of French Linguist Gustave Guillaume. Dans K. Flottum et F. Rastier (dir.), *Academic discourse - multidisciplinary approaches* (p. 55-74). Novus Press. <https://halshs.archives-ouvertes.fr/halshs-00150134>
- van Dijk, T. A. (1972). *Some aspects of text grammars. A study in theoretical linguistics and poetics*. Mouton. http://digitool.hbz-nrw.de:1801/webclient/DeliveryManager?pid=2863087&custom_att_2=simple_viewer
- van Eemeren, F. H. et Grootendorst, R. (1984). *Speech acts in argumentative discussions: a theoretical model for the analysis of discussions directed towards solving conflicts of opinion*. Foris Publications.
- van Eemeren, F. H. et Grootendorst, R. (2003). *A Systematic Theory of Argumentation: The pragma-dialectical approach*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511616389>
- van Eemeren, F. H., Grootendorst, R. et Snoeck Henkemans, A. F. (1996). *Fundamentals of argumentation theory: a handbook of historical backgrounds and contemporary developments*. L. Erlbaum. http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=007266826&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA
- van Gelder, T. (2001). How to improve critical thinking using educational technology. Dans G. Kennedy, M. Keppell, C. McNaught et T. Petrovic (dir.), *Proceedings of the 18th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education* (p. 539-548). Biomedical Multimedia Unit, The University of Melbourne.
- Vapnik, V. (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control*, 774-780.
- Vapnik, V. (1995). *The nature of statistical learning theory*. Springer.
- Vapnik, V., Golowich, S. et Smola, A. (1996). Support Vector Method for Function Approximation, Regression Estimation and Signal Processing. Dans *Advances in Neural Information Processing Systems* (vol. 9). MIT Press.

- <https://proceedings.neurips.cc/paper/1996/hash/4f284803bd0966cc24fa8683a34afc6e-Abstract.html>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. et Polosukhin, I. (2017). Attention is All you Need. Dans *Advances in Neural Information Processing Systems* (vol. 30). Curran Associates, Inc.
<https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>
- Virbel, J. (1994). Annotation Dynamique Et Lecture Expérimentale: Vers Une Nouvelle Glose ? *Littérature*, (96), 91-105.
- Wagner, F. (2006). Intertextualité et théorie. *Cahiers de Narratologie. Analyse et théorie narratives*, (13).
<https://doi.org/10.4000/narratologie.364>
- Walker, M., Tree, J. F., Anand, P., Abbott, R. et King, J. (2012). A Corpus for Research on Deliberation and Debate. Dans *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (p. 812-817). European Language Resources Association (ELRA).
http://www.lrec-conf.org/proceedings/lrec2012/pdf/1078_Paper.pdf
- Walton, D. (2012). Using Argumentation Schemes for Argument Extraction: A Bottom-Up Method. *International Journal of Cognitive Informatics and Natural Intelligence*, 6(3), 33-61.
<https://doi.org/10.4018/jcini.2012070103>
- Walton, D. et Macagno, F. (2016). A classification system for argumentation schemes. *Argument & Computation*, 6(3), 1-27. <https://doi.org/10.1080/19462166.2015.1123772>
- Walton, D. N. (1996). *Argumentation schemes for presumptive reasoning*. Lawrence Erlbaum Associates.
- Walton, D., Reed, C. et Macagno, F. (2008). *Argumentation Schemes*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511802034>
- Wang, C., Guo, Z., Li, J., Li, G. et Pan, P. (2022). A Text-based Deep Reinforcement Learning Framework Using Self-supervised Graph Representation for Interactive Recommendation. *ACM/IMS Transactions on Data Science*, 2(4), 44:1-44:25. <https://doi.org/10.1145/3522596>
- Wason, P. C. (1966). Reasoning. Dans B. Foss (dir.), *New Horizons in Psychology* (p. 135-151). Penguin Books.
- Wenger, E., McDermott, R. A. et Snyder, W. (2002). *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Harvard Business Press.
- Werlich, E. (1976). *A text grammar of English*. Quelle und Meyer.
- Whately, R. (1854). *Elements of Logic*. Morton & Griswold.
- Wigmore, J. H. (1913). *The problem of proof*, 8(2), 77-103.
- Windes, R. R. et Kruger, A. N. (1961). *Championship Debating*. J. Weston Walch.

- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., ... Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. Dans *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (p. 38-45). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>
- Woods, W. A. (1975). What's in a link: Foundations for semantic networks. Dans A. Collins et D. G. Bobrow (dir.), *Representation and Understanding: Studies in Cognitive Science* (p. 35-82). Academic Press.
- Wotzlaw, A. (2010). Towards Better Ontological Support for Recognizing Textual Entailment. Dans P. Cimiano et H. S. Pinto (dir.), *Knowledge Engineering and Management by the Masses* (p. 316-330). Springer. https://doi.org/10.1007/978-3-642-16438-5_22
- Wray, D. (1983). Computer-assisted learning in language and reading. *Reading*, 17(1), 31-36. <https://doi.org/10.1111/j.1467-9345.1983.tb00723.x>
- Wu, J. (2011). Improving the writing of research papers: IMRAD and beyond. *Landscape Ecology*, 26(10), 1345-1349. <https://doi.org/10.1007/s10980-011-9674-3>
- Xiao, Y., Liu, B., Yin, J., Cao, L., Zhang, C. et Hao, Z. (2011). Similarity-based approach for positive and unlabelled learning. Dans *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume Two* (p. 1577-1582). AAAI Press.
- Youga, J. (1989). *Elements of Audience Analysis*. Macmillan Coll Div.
- Yu, H., Han, J. et Chang, K. C. (2004). Pebl: web page classification without negative examples. *IEEE Transactions on Knowledge and Data Engineering*, 16(1), 70-81. <https://doi.org/10.1109/TKDE.2004.1264823>
- Yu, H., Han, J. et Chang, K. C.-C. (2002). PEBL: positive example based learning for Web page classification using SVM. Dans *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (p. 239-248). Association for Computing Machinery. <https://doi.org/10.1145/775047.775083>
- Yu, S. et Li, C. (2007). PE-PUC: A Graph Based PU-Learning Approach for Text Classification. Dans P. Perner (dir.), *Machine Learning and Data Mining in Pattern Recognition* (p. 574-584). Springer. https://doi.org/10.1007/978-3-540-73499-4_43
- Zafar, M. R. et Khan, N. (2021). Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability. *Machine Learning and Knowledge Extraction*, 3(3), 525-541. <https://doi.org/10.3390/make3030027>
- Zhang, B. et Zuo, W. (2009). Reliable Negative Extracting Based on kNN for Learning from Positive and Unlabeled Examples. *Journal of Computers*, 4(1), 94-101. <https://doi.org/10.4304/jcp.4.1.94-101>
- Zhao, H., Phung, D., Huynh, V., Jin, Y., Du, L. et Buntine, W. (2021, 28 février). *Topic Modelling Meets Deep Neural Networks: A Survey*. arXiv. <https://doi.org/10.48550/arXiv.2103.00498>

